

Міністерство освіти і науки України

Системні технології

System technologies

6 (155) 2024

Регіональний міжвузівський збірник наукових праць

Засновано у січні 1997 року.

У випуску:

- ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ
ТА ОРГАНІЗАЦІЯ СУЧАСНОГО ВИРОБНИЦТВА
- МАТЕМАТИЧНЕ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ
ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ
- СИСТЕМНІ ТЕХНОЛОГІЇ ОБРОБКИ ІНФОРМАЦІЇ
ТА КІБЕРБЕЗПЕКА

Системні технології. Регіональний міжвузівський збірник наукових праць. –
Випуск 6 (155). - Дніпро, 2024. – 275 с.
ISSN 1562-9945 (Print).
ISSN 2707-7977 (Online).

Редакційна колегія випуску:

Алпатов А.П. - д.т.н., проф. (відп. редактор)
Архипов О.Є. - д.т.н., проф.
Білозьоров В.Є. - д.ф.-м.н., проф.
Бабічев С.А. (Чеська Республіка) - д.т.н., доц.
Єрьомін О.О. - д.т.н., проф.

Прогресивні інформаційні
технології та організація
сучасного виробництва

Гече Ф.Е. - д.т.н., проф., (відп. редактор)
Гуда А.І. - д.т.н., проф.
Гнатушенко Вік.В. - д.т.н., проф.
Кіріченко Л.О. - д.т.н., проф.

Математичне
та програмне забезпечення
інтелектуальних систем

Гнатушенко В.В. - д.т.н., проф., (відп. редактор)
Гожий О.П. - д.т.н., проф.
Светличний Д.С. (Польща) - д.т.н., проф.
Хандецький В.С. - д.т.н., проф.

Системні технології обробки
інформації
та кібербезпека

Збірник друкується за рішенням Вченої Ради
Українського державного університету науки і технологій
від 31.01.2024 р., № 6

Адреса редакції: 49600, Дніпро, пр. Гагаріна, 4
Український державний університет науки і технологій,
ННІ «Інститут промислових та бізнес технологій»
кафедра Інформаційних технологій та систем.
Тел. +38(097)6854525
E-mail: st@nmetau.edu.ua
<https://journals.nmetau.edu.ua/index.php/st>

© Український державний університет науки і технологій,
ННІ «Інститут промислових та бізнес технологій»,
ІВК «Системні технології», 2024

Системні технології
ЗБІРНИК НАУКОВИХ ПРАЦЬ
Випуск 6 (155)

Головний редактор: к.т.н., доц. Т.В. Селівьорстова

Технічний редактор та секретар збірки: к.т.н., доц. К.Ю. Островська

Здано до набору 24.12.2024. Підписано до друку 27.12.2024.

Формат 60x84 1/16. Друк - різнограф. Папір типограф.

Умов. друк арк. – 19,64. Обл.–видавн. арк. –17,19.

Тираж 300 прим. Замовл. – 06/24

Український державний університет науки і технологій,
ННІ «Інститут промислових та бізнес технологій»,
кафедра Інформаційних технологій та систем: ІВК «Системні технології»
49600, Дніпро, а/с 493

<http://journals.nmetau.edu.ua/index.php/st>

Свідоцтво про державну реєстрацію друкованого засобу масової інформації:

Серія КВ № 8684 від 23 квітня 2004 рік

Редакційна колегія

Селівьорстова Тетяна Віталіївна
(*головний редактор*)

доцент, кандидат технічних наук

Алпатов Анатолій Петрович

Член-кореспондент НАН України,
професор, доктор технічних наук

Архипов Олександр Євгенійович

професор, доктор технічних наук

Бабічев Сергій Анатолійович

доцент, доктор технічних наук

Білозьоров Василь Євгенович

професор,

доктор фізико-математичних наук

Гече Федір Елемирович

професор, доктор технічних наук

Гуда Антон Ігорович

(*заст. головного редактора*)

професор, доктор технічних наук

Гнатушенко Вікторія Володимирівна

(*вчений секретар*)

професор, доктор технічних наук

Гнатушенко Володимир Володимирович

професор, доктор технічних наук

Гожий Олександр Петрович

професор, доктор технічних наук

Єрьомін Олександр Олегович

професор, доктор технічних наук

Кіріченко Людмила Олегівна

професор, доктор технічних наук

Світличний Дмитро Святозарович

професор, доктор технічних наук

Скалозуб Владислав Васильович

професор, доктор технічних наук

Хандецький Володимир Сергійович

професор, доктор технічних наук

Український державний університет науки і технологій, ННІ «Інститут промислових та бізнес технологій», Україна

Інститут технічної механіки

НАНУ і ДКАУ, Україна

Національний технічний університет України
«Київський політехнічний інститут» імені Ігоря Сікорського", Україна

Jan Evangelista Purkyně University

in Ústí nad Labem

Університет імені Яна Євангеліста Пуркіне, Усті над Лабем, Чеська Республіка

Дніпровський національний університет імені

Олеся Гончара, Україна

Ужгородський національний університет, Україна

Український державний університет науки і технологій, ННІ «Інститут промислових та бізнес технологій», Україна

Український державний університет науки і технологій, ННІ «Інститут промислових та бізнес технологій», Україна

Національний технічний університет

«Дніпровська політехніка», Україна

Чорноморський національний університет імені

П.Могили, Україна

Український державний університет науки і технологій, ННІ «Інститут промислових та бізнес технологій», Україна

Харківський національний університет

радіоелектроніки, Україна

Akademia Górniczo-Hutnicza

Краківська гірничо-металургійна академія ім. С. Сташіца, Польща

Український державний університет науки і технологій, ННІ «Дніпровський інститут інфраструктури і транспорту» Україна

Дніпровський національний університет імені

Олеся Гончара, Україна

**БАГАТОКРИТЕРІАЛЬНЕ МОДЕЛЮВАННЯ
В УМОВАХ НЕОДНОРІДНОЇ НЕВИЗНАЧЕНОСТІ ДАНИХ
У РОБОТАХ СКАЛОЗУБА В.В.**

Анотація. Виконано узагальнення підходів до вирішення завдань багатокритеріального моделювання в умовах неоднорідної невизначеності даних, у тому числі отриманих при моніторингу з нерівномірними та нечіткими інтервалами. Усі розглянуті підходи були створені під керівництвом та за безпосередньої участі проф. Скалозуба В. В. У роботі виділено основні підходи до моделювання на основі багатопараметричних конструктивно-продукційних, сепарабельних та реляційно-сепарабельних моделей. Отримані результати можуть бути використані для подальшого розвитку запропонованих ідей та підходів.

Ключові слова: умови невизначеності, нерівномірний та нечіткий інтервал вибірки, упорядкування мультипослідовностей елементів, сепарабельна модель, конструктивно-продукційне моделювання, конструктивна модель, інформаційні технології.

Постановка проблеми. Проблеми багатокритеріального моделювання і управління в умовах неоднорідної невизначеності даних є актуальними для транспортної, економічної, медичної галузей тощо. Їх актуальність зумовлена наявністю численних наборів даних, отриманих в ході спостережень за реабілітацією пацієнтів, спостережень за станом віддалених об'єктів, даними щодо технічного стану транспортних засобів, які мають нечітку, розмиту структуру. Разом з тим, вони можуть бути використані в задачах аналізу та прогнозування відповідних процесів. Це можливо за умови побудови відповідних моделей структур даних та процесів.

Серед завдань, що потребують побудови моделей представлення нечіткості та неоднорідності даних, можна виділити такі, як моделювання та аналіз даних процесів моніторингу, які представлені часовими рядами і відрізняються змінним або нечітким інтервалом спостережень; планування потоків замовлень систем обслуговування за неточно визначених даних; побудова моделей процесів упорядкування мультипослідовностей елементів із нечіткими параметрами тощо.

Мета дослідження. Метою цієї роботи є систематизація підходів до вирішення вказаних завдань, запропонованих у роботах проф. Скалозуба Владислава Васильовича, а також узагальнення наукового доробку Скалозуба В. В. для подальшого розвитку запропонованих ідей та підходів.

Моделювання процесів упорядкування неоднорідних мультипослідовностей з використанням конструктивних моделей. Задачі упорядкування розглядаються в логістичних, технологічних, інформаційних, залізничних та інших процесах.

У [1] запропоновано моделі процесів оптимального розформування-формування багатогрупових поїздів. В основу покладено конструктивно-продукційне моделювання (КПМ) [2], що базується на використанні апарату формальних мов та граматики.

Основна ідея КПМ полягає у представленні предмету моделювання як трійки: носій, множина операцій на елементах носія та інформаційне забезпечення конструювання (ІЗК). Носій є неоднорідним, динамічним і представляється онтологічними засобами [3, 4]. ІЗК визначає особливості конструювання, зокрема алгоритми виконання операцій сигнатури.

У [1] розроблено багат шарову конструктивну модель з упорядкування наборів неоднорідних послідовностей замовлень, яка враховує складність операцій формування. Особливістю моделі є наявність у її складі додаткових структур конструювання, що забезпечують можливості задавання складності операцій формування, а також додаткового аналізу властивостей об'єктів, які формуються під час реалізації розв'язків.

Досліджувані в [1] завдання на практиці поширені в логістичних, технологічних, інформаційних та інших процесах.

Оскільки в оцінках складності операцій формування присутня нечіткість та замовлення можуть мати індивідуальні особливості, у [5] конструктивні багат шарові моделі набули подальшого розвитку. Запропоновано комплексні моделі процесів упорядкування мультипослідовностей елементів із нечіткими параметрами. Конструктивно-продукційна система, покладена в основу моделей, формує послідовності виконання алгоритмів, необхідних для упорядкування елементів. В залежності від визначення алгоритмів, утворені продукції набувають індивідуальності.

Запропоновані моделі містять шари, які забезпечують урахування нечітких оцінок складності операцій упорядкування, класифікацію нечітких параметрів вихідних елементів, формування та аналіз індивідуальних нечітких моделей процесів надходження замовлень до сервісних систем.

Багат шарові конструктивні моделі процесів упорядкування мультипослідовностей замовлень з урахуванням складності використаних операцій покладено в основу інтелектуальних процедур упорядкування послідовностей елементів (замовлень) [6]. У кожній з моделей процеси формування порядку визначаються через параметри зони обслуговування, що представляються як формальні структури, призначені для формування вихідних послідовностей з первинних. Аналіз вмісту зон дозволяє встановити всі властивості процесу конструювання рішення для формування упорядкованих послідовностей замовлень. Також за даними зон виконується перевірка завершеності знайденого рішення.

При формуванні рішення для побудови упорядкованих послідовностей вибір оператора може здійснюватися шляхом повного перебору доступних [1] або з використанням інтелектуальних процедур на основі моделі асоціативної пам'яті Хемінга [6].

З метою підвищення ефективності процесу упорядкування запропоновано спеціалізовані моделі вхідних (первинних) множин елементів для упорядкування, а також алгоритми їх оброблення [7]. Моделі були застосовані для завдань ефективного сортування мультипослідовностей замовлень, класифікації при інтервальній невизначеності даних. В основу моделей покладено структури бінарних дерев та алгоритми їх обробки, які підвищують ефективність процесу упорядкування.

Завдання упорядкування послідовності елементів за умов нечіткості вхідних даних, в тому числі тих, що мають природомовний характер, вирішується на основі продукційних моделей керування за шаблонами [8]. Для забезпечення адекватності формальних структур керування продукційної системи за неоднорідних, неповних даних передбачено етап моделювання первинних даних.

Моделювання даних моніторингу процесів з нерівномірними та нечіткими інтервалами вибірки. Нерівномірність та нечіткість інтервалів отримання даних під час моніторингу ускладнюють процеси їх аналізу, а також побудови прогнозів на їхній основі.

У [9] запропоновано сепарабельну модель (СПМ) для дослідження процесів клінічного моніторингу стану хворих на діабет з метою визначення оцінок величин нового періоду до виникнення станів/подій. У досліджені розглянуті періоди моделі моніторингу – надзвичайно нерівномірні та особливі для різних хворих.

СПМ процесів моніторингу в цілому визначає значення нового моменту виникнення чергової події спостережуваного процесу та відповідні кожному моменту характеристики процесу. Особливість цієї моделі полягає у виокремленні нерівномірних (нечітких) інтервалів вибірок у окрему складову моделі, що розглядається як один з компонентів даних процесу моніторингу [10].

Для підвищення ефективності моделювання і аналізу процесів з часовими рядами з невизначеними або нечіткими інтервалами запропоновано реляційно-сепарабельна модель (PCM) і комбінований квантильний алгоритм [11]. У склад PCM включено нечітку реляційну модель, що описується системою нечітких реляційних відношень першого та другого порядку, отриманих на основі вихідної послідовності даних. У відношеннях враховано такі параметри модельованих процесів, як контрольований показник (у [11] – рівень цукру у крові) та показник часу. Комбінація показників дозволяє врахувати кілька попередніх етапів розвитку процесів. У PCM, порівняно з СПМ, для процесів, представлених часовими рядами з невизначеними або нечіткими інтервалами, враховано додаткові об'єктивні властивості, пов'язані зі змінністю інтервалів, у формі нечітких відношень.

Невизначеність у завданнях багатокритеріального моделювання та управління. Опрацювання наукового доробку Скалозуба В. В. дозволяє визначити такі основні особливості вирішення поставлених завдань:

- неоднорідність та нечіткість даних, що утворюють мультипослідовність елементів;
- нерівномірність та нечіткість інтервалів спостережень під час моніторингу процесів.

Для їх врахування колективами авторів, що співпрацювали зі Скалозубом В.В., запропоновано підходи на основі моделей:

- багат шарових конструктивно-продукційних;
- сепарабельних та реляційно-сепарабельних.

Використання цих підходів дозволило формалізувати завдання обробки нечітких даних та інтервалів, а також розробити ряд інтелектуальних процедур, що враховують вимоги процесів різних предметних областей.

Висновки. В роботі виконано узагальнення підходів до вирішення проблем багатокритеріального моделювання і управління в умовах неоднорідної невизначеності даних. Виділено основні особливості та підходи до їх вирішення. Особливо слід відзначити, що у розглянутих роботах предметними областями для прикладів є транспортно-логістична та медична галузь, що вказують на широту сфери застосування наукового доробку Скалозуба В. В. та його співавторів.

Перспективним є розвиток виявлених підходів, у тому числі шляхом їх поєднання, а також доповнення конструктивно-продукційних моделей формалізованим описом конструктивних елементів, у тому числі вхідних даних, онтологічними засобами. Використання формалізованих способів представлення онтологій, наприклад, OWL, дозволить автоматизувати обробку відповідних даних.

Післямова. Владислав Васильович Скалозуб (1949 – 2024) – доктор технічних наук, професор кафедри «Комп’ютерні інформаційні технології» (КІТ) ННІ «Дніпровський інститут інфраструктури і транспорту» Українського державного університету науки і технологій.

Сфера наукових інтересів проф. Скалозуба В. В. охоплювала теоретичні і прикладні питання багатокритеріального моделювання і управління в умовах неоднорідної невизначеності даних; розвиток інтелектуальних технологій залізничного транспорту; проблематику управління парками складних систем; автоматизація управління на основі принципів сталого розвитку; проблеми управління неоднорідними потоками у мережах; моделювання процесів управління у автоматизованих системах залізничного транспорту України тощо.

Загальний науковий доробок Скалозуба В. В. складає декілька монографій та сотні наукових статей та доповідей. Під його керівництвом захищено незліченну кількість кваліфікаційних та наукових робіт, зокрема на здобуття ступеня кандидата технічних наук.

На сьогодні багатогранний науковий доробок Скалозуба В. В. сприяє розвитку широкого кола досліджень, зокрема пов’язаних з нечіткою логікою, невизначеними даними, що притаманні різним предметним областям, зокрема інтелектуальним транспортним системам.

ЛІТЕРАТУРА

1. Скалозуб В. В., Ільман В. М., Білий Б. Б. Конструктивні багат шарові моделі для впорядкування послідовностей з урахуванням складності операцій формування. Наука та прогрес транспорту. № 4 (88), 2020. С. 61-76. <https://doi.org/10.15802/stp2020/213232>

2. Shynkarenko V. I., Ilman V. M. Constructive-synthesizing structures and their grammatical interpretations. I. Generalized formal constructive-synthesizing structure. Cybernetics and Systems Analysis N. 50.5, 2014. P. 655-662. DOI:10.1007/s10559-014-9655-z
3. Skalozub V., Ilman V., Shynkarenko V. Development of ontological support of constructive-synthesizing modeling of information systems. Eastern-European Journal of Enterprise Technologies, v. 6/4 (90), 2017. P. 58-69. <https://doi.org/10.15587/1729-4061.2017.119497>
4. Skalozub V., Ilman V., Shynkarenko V. Ontological support formation for constructive-synthesizing modeling of information systems development processes. Eastern-European Journal of Enterprise Technologies, v. 5/4 (95), 2018. P. 55-63. 15. <https://doi.org/10.15587/1729-4061.2018.143968>
5. Скалозуб В. В., Горячкін В. М., Мурашов О. В. Комплексні моделі впорядкування мультипослідовностей із нечіткими параметрами. Наука і прогрес транспорту, № 2 (92), 2021. С. 50-64. DOI: 10.15802/stp2021/237291.
6. Скалозуб В. В., Горячкін В. М., Терлецький І. А. Інтелектуальні процедури упорядкування послідовностей замовлень неоднорідними операторами формування. Транспортні системи та технології перевезень, № 22, 2021. С. 67-79. <https://doi.org/10.15802/tstt2021/247885>
7. Скалозуб В. В., Горячкін В. М., Скалозуб М. В. Терлецький І. А. Дослідження алгоритмів оптимального формування транспортно-логістичних процесів реального часу гетерогенними операторами. Транспортні системи та технології перевезень, № 23, 2022. С. 65-73. <https://doi.org/10.15802/tstt2022/261657>
8. Скалозуб В. В., Горячкін В. М., Терлецький А. І. Інтелектуальна технологія оптимізації керування потоками замовлень сервісних систем із неточно визначеними і природномовними даними. Наука і прогрес транспорту, № 2 (102), 2023. С. 54-70. <https://doi.org/10.15802/stp2023/288077>
9. Скалозуб В. В., Мурашов О. В. Моделювання даних процесів моніторингу при нерівномірних і нечітких інтервалах спостережень. Системні технології, № 4 (135), 2021. С. 135-148. DOI 10.34185/1562-9945-4-135-2021-14
10. Скалозуб В. В., Мурашов О. В. Моделювання процесів моніторингу при нерівномірних і нечітких інтервалах спостережень. Інформаційні технології у металургії та машинобудуванні, 2021. С. 193-194.
11. Скалозуб В. В., Горячкін В. М., Мурашов О. В. Реляційно-сепарабельні моделі процесів моніторингу при перемінних і нечітких інтервалах спостережень. Системні технології, № 4 (147), 2023. С. 3 – 19. DOI 10.34185/1562-9945-4-147-2023-01

REFERENCES

1. Skalozub V. V., Ilman V. M., Bily B. V. Constructive multi-layer models for ordering a set of sequences, taking into account the complexity operations of formation. Science and Transport Progress. № 4 (88), 2020. P. 61-76. <https://doi.org/10.15802/stp2020/213232>
2. Shynkarenko V. I., Ilman V. M. Constructive-synthesizing structures and their grammatical interpretations. I. Generalized formal constructive-synthesizing structure. Cybernetics and Systems Analysis N. 50.5, 2014. P. 655-662. DOI:10.1007/s10559-014-9655-z

3. Skalozub V., Iman V., Shynkarenko V. Development of ontological support of constructive-synthesizing modeling of information systems. Eastern-European Journal of Enterprise Technologies, v. 6/4 (90), 2017. P. 58-69. <https://doi.org/10.15587/1729-4061.2017.119497>
4. Skalozub V., Iman V., Shynkarenko V. Ontological support formation for constructive-synthesizing modeling of information systems development processes. Eastern-European Journal of Enterprise Technologies, v. 5/4 (95), 2018. P. 55-63. 15. <https://doi.org/10.15587/1729-4061.2018.143968>
5. Skalozub V. V., Horiachkin V. M., Murachov O. V. Complex Models of Ordering Multi-Sequences with Fuzzy Parameters. Science and Transport Progress, № 2 (92), 2021. P. 50-64. DOI: 10.15802/stp2021/237291.
6. Skalozub V., Horiachkin V., Terletskii I. Intellectual procedures for ordering sequence orders by inhomogeneous forming operators. Transport systems and transportation technologies, № 22, 2021. P. 67-79. <https://doi.org/10.15802/tstt2021/247885>
7. Skalozub V., Horiachkin V., Skalozub M., Terletskii I. Investigation of the algorithms for the transport and logistics processes optimal formation by heterogeneous operators in real time. Transport systems and transportation technologies, № 23, 2022. P. 65-73. <https://doi.org/10.15802/tstt2022/261657>
8. Skalozub V., Horiachkin V., Terletskii I. Intelligent Technology for Optimizing the Management of Order Flows of Service Systems With Imprecisely Defined and Natural Language Data. Science and Transport Progress, № 2 (102), 2023. P. 54-70. <https://doi.org/10.15802/stp2023/288077>
9. Skalozub V., Murashov O. Modeling of monitoring processes with uneven and fuzzy observation intervals. System technologies, № 4 (135), 2021. P. 135-148. DOI 10.34185/1562-9945-4-135-2021-14
10. Skalozub V., Musachov O. Modeling of monitoring processes with uneven and unclear observation intervals. International scientific and technical conference Information technologies in metallurgy and machine building, 2021. P. 193-194.
11. Skalozub V., Horiachkin V. M., Murashov O. Relational-separable models of monitoring processes at variable and unclear observation intervals. System technologies, № 4 (147), 2023. P. 3 – 19. DOI 10.34185/1562-9945-4-147-2023-01

Received 23.08.2024.
Accepted 28.08.2024.

***Multi-criterion modeling in the conditions of non-homogeneous data
uncertainty in the V. Skalozub papers***

The problems of solving multi-criteria modeling tasks in conditions of heterogeneous data uncertainty are characteristic and relevant for many areas, in particular, transport and logistics, medical, and economic. This paper summarizes the approaches to solving these areas tasks, including cases for data obtained during monitoring with uneven and unclear intervals. All the considered approaches were created under the guidance and with the direct participation of prof. Skalozub V. V. The author's team, with his participation, illustrated examples of the use of models in the tasks of organizing multi-symbol sequences in transport (for-

mation-dismantling of trains, order processing) and in the analysis and forecasting of states/events based on data on diabetes diseases.

In this paper, the main approaches to modeling in conditions of heterogeneous data uncertainty based on multi-layer constructive-synthesizing, separable and relational-separable models are highlighted. Multi-layer constructive-synthesizing became new branch of develop constructive-synthesizing modeling. Their practical value currently lies in the possibility of obtaining a new form of implementation of the specified technological processes of railway transport. In the future, they can be applied to tasks that require step-by-step data processing with a clear demarcation of the relevant operations.

The obtained results can be used for further development of the proposed ideas and approaches.

It is promising to develop the identified approaches, including by combining them, as well as supplementing constructive-production models with a formalized description of constructive elements, including input data, by ontological means.

The paper contains information about the figure and scientific work of V. Skalozub.

Keywords: conditions of uncertainty, uneven and fuzzy sampling interval, arrangement of multisequences of elements, separable model, constructive-synthesizing modeling, constructive model, information technologies.

Куроп'ятник Олена Сергіївна – кандидат технічних наук, доцент, доцент кафедри «Комп'ютерні інформаційні технології», Український державний університет науки і технологій.

Kuropiatnyk Olena – candidate of engineering sciences, associate professor, associate professor of Computer Information Technology Department at Ukrainian State University of Science and Technologies.

С.В. Плаксін, А.М. Муха, Д.В. Устименко,

В.І. Доскоч, Т.А. Сандул, С.М. Куліков

ОБҐРУНТУВАННЯ ЗАСТОСУВАННЯ РОЗПОДІЛЕНОЇ МЕРЕЖІ ФОТОЕЛЕКТРИЧНИХ ПЕРЕТВОРЮВАЧІВ ДЛЯ ЖИВЛЕННЯ ЛІНІЙНОГО ДВИГУНА МАГНІТОЛЕВІТАЦІЙНОГО ТРАНСПОРТУ

Анотація. Сучасний високошвидкісний транспорт є основою сталого економічного та соціального розвитку держави, суспільства з дотримання вимог екологічності. Обґрунтована концепція енергозабезпечення лінійного двигуна магнітолевітаційного транспорту від розподіленої мережі фотоелектричних перетворювачів. Запропонований базовий силовий елемент шляхової енергетичної установки у вигляді завершеного вузла до складу якого входять сонячна панель, накопичувач та інвертор, що працює на навантаження у вигляді «короткої» шляхової котушки. Застосування «короткої» шляхової котушки дозволяє зменшити втрати електричної енергії, оскільки тягове зусилля формується лише в зоні взаємодії із рухомою одиницею, до незадіяних ділянок шляхової структури енергія не передається. За рахунок зменшення довжини робочої ділянки суттєво підвищуються інші енергетичні показники системи у цілому, зокрема коефіцієнт потужності та коефіцієнт корисної дії. Зменшення довжини робочих ділянок в умовах високих швидкостей рухомого складу призведе до підвищення вимог надійності та швидкодії шляхових вимикачів. Шляхові вимикачі створюються на базі силових напівпровідникових ключів, які керуються сучасними мікроконтролерами. Ефективність запропонованої структури залежить від сонячної активності в регіоні розташування транспортної артерії. Оціночні показники сонячної активності дозволяють стверджувати, що вагові показники рухомого складу, що має рухатися по колії з «короткою» секції, відповідають наявним ресурсам.

Ключові слова: лінійний двигун, фотоелектрична система, енергопостачання, магнітолевітаційний поїзд, транспорт, накопичувач.

Постановка проблеми

Транспортна система будь-якої країни є надважливим елементом її економіки. Серед вимог, що пред'являють до транспорту особливо значимими є швидкість доставки пасажирів та вантажів, а також його енергоефективність. Високошвидкісний наземний транспорт на раціональних енергосистемах є пріоритетним напрямом розвитку науки і техніки, що визначено 6-ю Рамочною Програмою Євросоюзу [1].

До сучасних високошвидкісних наземних транспортних систем, в першу чергу, відносяться рейкові опорні системи типу TGV (Alstom, Франція); монорельсовий поїзд

з використанням магнітної левітації Transrapid (Siemens та Thyssenkrupp Transrapid GmbH, Німеччина); JR-Maglev система швидкісних поїздів на магнітній підвісці (Japan Railway Technical Research Institute та Japan Railways, Японія). Враховуючи значимість транспортних систем для людства та зважаючи на останні досягнення науки і техніки інтерес до магнітолевітаційного транспорту буде тільки зростати [2, 3].

Класичні магнітолевітаційні системи дозволяють розвивати швидкості руху, які наближаються до швидкості повітряних перевезень але при цьому мають не найкращі енергетичні показники. Використання альтернативних джерел енергії, наприклад сонячної, здатне суттєво покращити енергетичні показники транспорту на магнітній левітації але потребує перегляду підходів до побудови його шляхової структури та способів керування лінійним двигуном.

Основний фокус при створенні нової шляхової структури повинен бути на: беззатратній доставці електричної енергії споживачу; співрозмірність питомих характеристик генерації та питомих характеристик навантаження; гарантованість надходження енергії та її екологічну чистоту.

Мета дослідження

Основною задачею представленого матеріалу є обґрунтування на концептуальному рівні застосування системи енергетичного забезпечення лінійного тягового електродвигуна шляхової структури магнітолевітаційного транспорту від розподіленої мережі фотоелектричних перетворювачів.

Викладення основного матеріалу дослідження

В основі обґрунтування можливості застосування розподіленої мережі фотоелектричних перетворювачів для живлення лінійного двигуна магнітолевітаційного транспорту є оцінка енергозатрат самого транспорту та енергетичних можливостей пропонуваної енергетичної системи.

Одним з показників, що характеризує ефективність лінійного тягового електропривода високошвидкісного наземного транспорту [4] є показник питомих енергозатрат $W_{\text{пит}}$ (Вт·год/т·км), що залежить від прийнятого закону руху, відстані між зупинками рухомого складу, величини опору руху та енергетичних показників тягового привода.

Дослідження енергетики руху поїзда виконано для поїзда з електродинамічним підвісом, що нараховує 10 вагонів та має загальну масу 400 тон. Довжина перегонів коливається від 100 до 300 км, а швидкість руху становить 500 км/год.

В результаті встановлено, що затрати питомої енергії на рух поїзда не перевищують 123 Вт·год/т·км і в середньому мають значення 110 Вт·год/т·км для перегону довжиною 100 км та 93 Вт·год/т·км для перегону довжиною 300 км. Потужність, що розвивається двигунами за умови руху поїзда з постійною швидкістю 500 км/год складає 14,66 МВт, потужність, що споживається з мережі –18,9 МВт.

Класична система енергопостачання тягового привода високошвидкісного наземного магнітолевітаційного транспорту має велику лінійну протяжність секцій статорної обмотки синхронного тягового електродвигуна, і їх, потрібно розглядати як частину системи енергопостачання. Довжина секцій обмотки статора, її параметри визначають основні параметричні та вартісні показники системи енергопостачання в цілому.

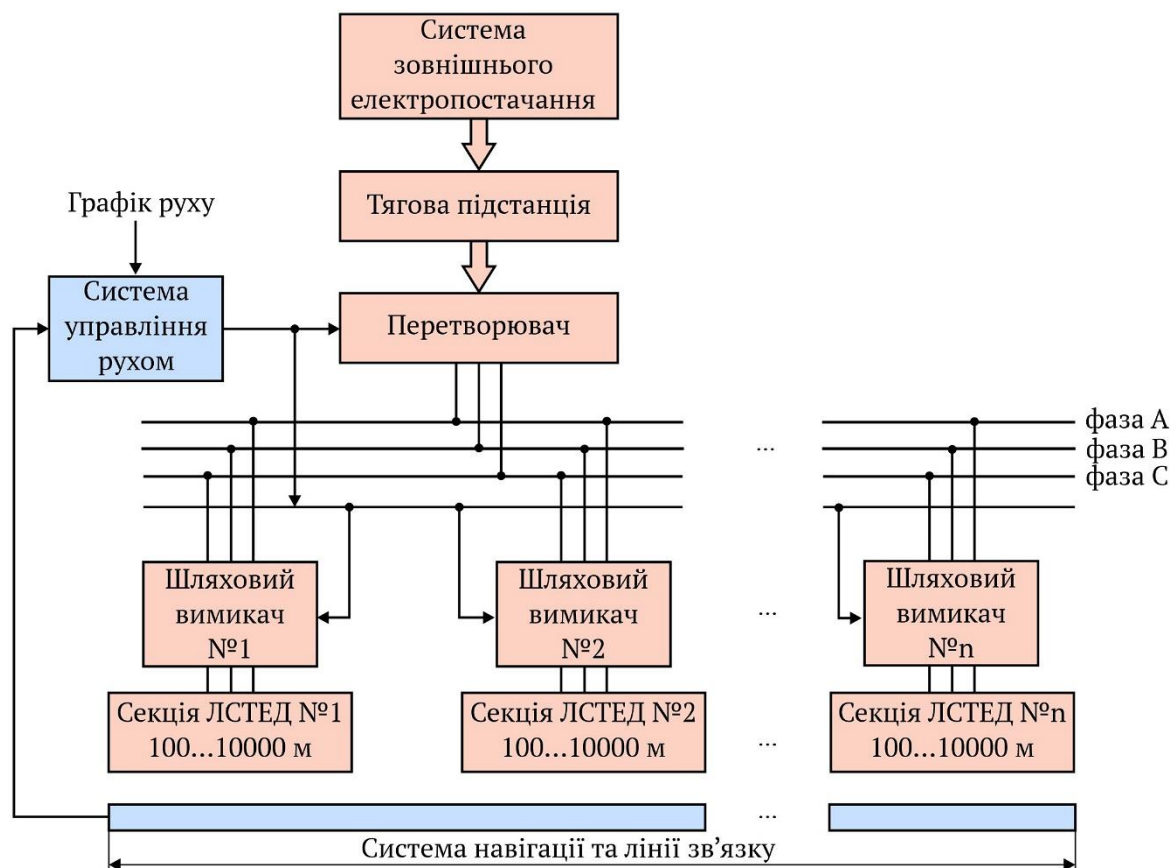


Рисунок 1 – Структура енергозабезпечення лінійного синхронного тягового електродвигуна (ЛСТЕД) високошвидкісного наземного магнітолевітаційного транспорту від системи зовнішнього джерела енергії

Узагальнена структурна схема енергозабезпечення лінійного синхронного тягового електродвигуна високошвидкісного наземного магнітолевітаційного транспорту від системи зовнішнього джерела енергії наведена на рис. 1 (силові елементи виділені червоним кольором). Якірна обмотка лінійного синхронного тягового електродвигуна (шляхові котушки секцій 1, 2...n) отримують живлення через відповідні шляхові вимикачі від системи трифазних фідерів (фази А, В та С), що з'єднані з перетворювачем. Напруга і частота на кожному фідері регулюються у відповідності з тяговим режимом. Вмикання на живлення як і вимикання кожної секції ЛСТЕД відбувається по мірі просування поїзда шляхом системою навігації та зв'язку, яка контролюється системою управління рухом. В структурі зображеній на рис. 1 довжина фідерної зони коливається в діапазоні 10...40 км. Величина струму в секції ЛСТЕД та його корегування, оптимальне значення кута навантаження, що дає високі значення ККД та коефіцієнта потужності, а також установка необхідної частоти напруги живлення в процесі розгону чи гальмування досягається керуванням перетворювачами. Поточна об'єктивна інформація про кут навантаження в систему керування перетворювачами надходить з системи навігації чи по каналам зв'язку з поїздом. Аналогічним чином влаштоване енерго-

постачання магнітолевітаційних транспортних систем Японії, Китаю, Німеччини, що випробовуються на їх полігонах.

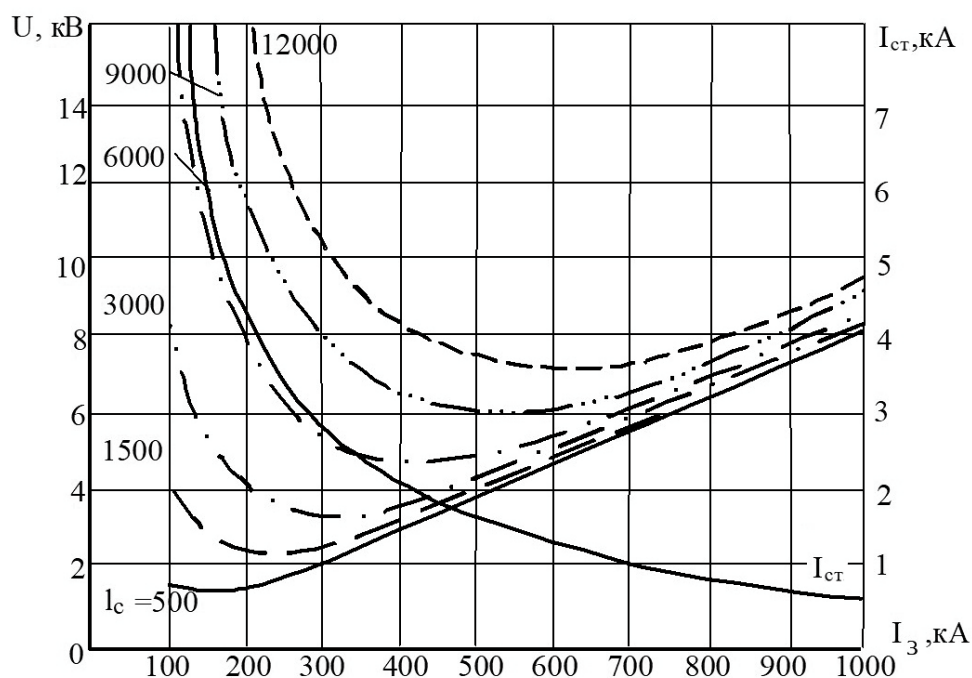


Рисунок 2 – Залежність напруги U і струму статора $I_{ст}$ ЛСТЕД від струму збудження I_3 в надпровідниковому контурі за різних довжин секцій статорної обмотки l_c

На рис. 2 наведені залежності напруги U і струму статора $I_{ст}$ ЛСТЕД від струму збудження I_3 в надпровідниковому контурі за різних довжин секцій статорної обмотки l_c для випадку усталеного руху з швидкістю 139 м/с та за частоти струму живлення 50 Гц (кількість надпровідникових контурів збудження 90, повітряний зазор ЛСТЕД 0,2 м, кут між векторами магнітних полів збудження та статорної обмотки 110° , сила тяги двигуна 105,5 кН). Для усіх довжин секцій залежність струму статорної обмотки від струму збудження має однаковий характер. Значення напруги живлення статора зростає із ростом довжини секції статорної обмотки від 500 м до 12000 м, тобто з ростом її індуктивного опору.

Умови за яких отримані залежності енергетичних показників ($\cos\phi$ та ККД) від струму збудження (рис. 3) аналогічні умовам, що й на рис. 2. Таким чином, для кожної довжини секції статорної обмотки лінійного двигуна можна знайти значення струму збудження за якого забезпечується компенсація реактивної складової енергії, що споживається.

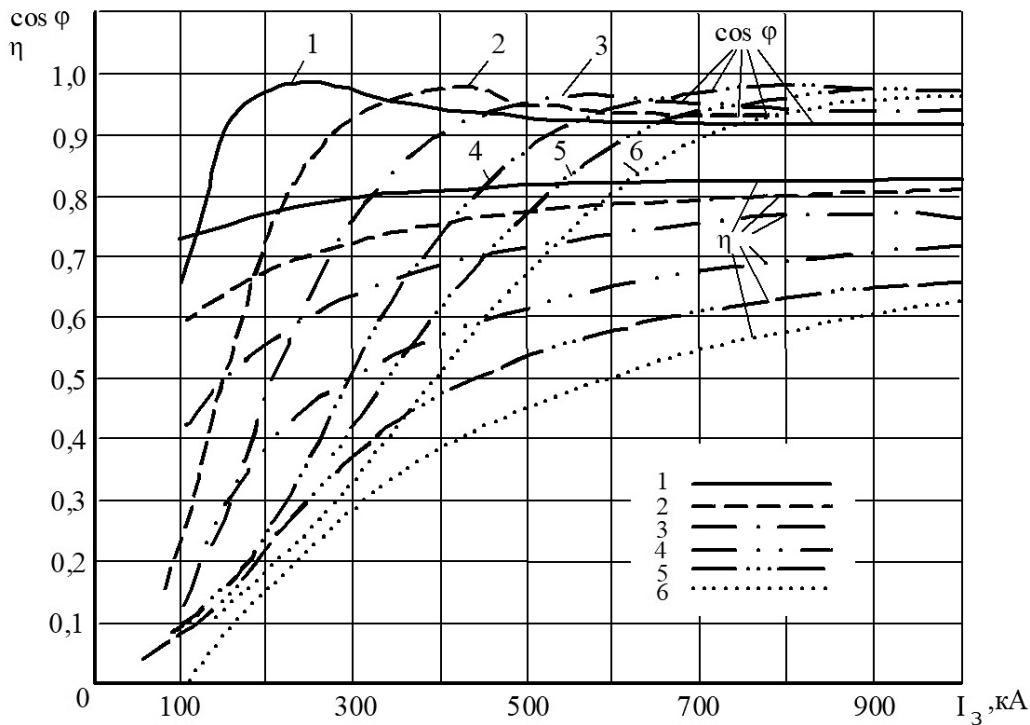


Рисунок 3 – Енергетичні показники ЛСТЕД в залежності від струму збудження двигуна (довжина секції $l_c=500$ м – крива 1, 1500 м – 2, 3000 м – 3, 6000 м – 4, 9000 м – 5 і 12000 м – 6)

Довжина ділянки, що отримує живлення, суттєво впливає на величини сумарної активної та повної потужностей споживані двигуном, середнє значення ККД і коефіцієнта потужності. Із зменшення довжини секції статорної обмотки тягового лінійного двигуна підвищуються його енергетичні показники (коефіцієнт потужності та ККД) але це потребує підвищення вимог до надійності та швидкодії шляхових вимикачів. На рис. 4 наведені залежності коефіцієнта потужності ($\cos\phi$) і ККД (η) від часу попереднього увімкнення наступної ділянки t_b за умови $t_b=t_0$ (t_0 – час вимикання ділянки, що виходить з роботи) та довжини ділянки 2000 м. Зона А (рис. 4) якій відповідає $t_b=t_0 < 1$ с., не може бути прийнятою в якості робочої не дивлячись на високі енергетичні показники, оскільки швидкодія шляхових перемикачів, що залежить від виду вимикачів які застосовуються і системи керування ними, не дозволяє досягнути таких значень часу вмикання і вимикання.

Зона С (рис. 4) якій відповідає $t_b=t_0 > 9$ с., також не може бути прийнятою в якості робочої, оскільки в цьому випадку одночасно будуть отримувати живлення три ділянки, а також ця зона характеризується найнижчими енергетичними показниками. Таким чином, для забезпечення надійної роботи вимикачів і досягнення найкращих енергетичних показників двигуна час вмикання і вимикання ділянок живлення повинно бути в зоні В (діапазон часу 1...9 с.). Для зони В середній енергетичний фактор (добуток коефіцієнта потужності і ККД) зменшується від 0,69 ($t_b=t_0=2$ с.) до 0,49 ($t_b=t_0=8$ с.)

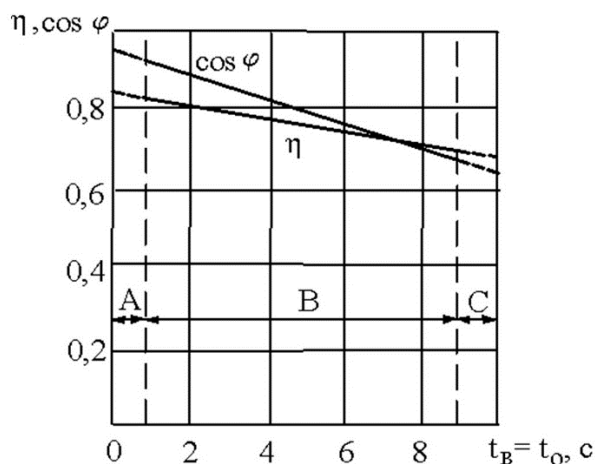


Рисунок 4 – Залежність енергетичних показників ЛСТЕД від часу увімкнення/вимкнення ділянок живлення ($l_c=2000$ м)

Для покращення техніко-економічних показників лінійного тягового електродвигуна час попереднього вмикання наступної і вимкання попередньої ділянки потрібно приймати мінімально можливим за умови збереження надійної роботи шляхових вимикачів.

Аналіз графічних залежностей представлених на рис. 2, 3 та 4 однозначно вказує на те, що ефективність енергопостачання зростає із зменшення довжини секції обмотки статора лінійного тягового електродвигуна і обмежується можливостями фізично відтворити розміри шляхової котушки $2\tau/3$, де τ – величина полюсного ділення.

Крім того, оптимізація розміру шляхових котушок з урахуванням динамічної стійкості і енергоефективності призводить до зменшення фізичних розмірів шляхових котушок до 0,5...2,0 м [5].

Але саме по собі використання коротких секцій ЛСТЕД за традиційного способу енергопостачання не вирішує задачі зниження втрат, так як фідери залишаються багатокілометровими, а потужність, що по ним пропускається, значною. Тільки подача живлення на окрему шляхову котушку від автономного джерела енергії з відповідною структурою системи керування дозволить вирішити проблему запасів реактивної потужності та знизити втрати, що обумовленні «холостою» роботою сусідніх секцій при проходженні поїздом стиків секцій. Розв'язати подібну задачу можна, якщо використати в якості джерела енергії розподілену фотоелектричну енергосистему.

Для оцінки інтенсивності сонячної радіації введено поняття сонячної постійної, що рівна енергії випромінення Сонця яка падає за одиницю часу на одиницю площі поверхні перпендикулярній потоку випромінення, в космічному просторі на середній відстані Землі від Сонця. Стандартне значення сонячної постійної дорівнює 1353 Вт/м^2 . Через особливості орбітального руху Землі густина потоку біля земної поверхні залежить від пори року і часу доби.

Аналіз статистичних даних надходження сонячної енергії в Україні показав, що середньорічна кількість сумарної сонячної радіації на квадратний метр площі поверхні коливається від $1070 \text{ кВт}\cdot\text{год/м}^2$ на півночі країни до $1400 \text{ кВт}\cdot\text{год/м}^2$ на її півдні [6].

Для різних регіонів характерна різна інтенсивність сонячної радіації на протязі року, так в пустелі взимку сумарний час надходження сонячної радіації майже в 2 рази менший у порівнянні з літнім періодом. Для території України це співвідношення знаходиться в межах 4...7 [6]. Для практичних розрахунків з урахуванням орієнтації і ККД сонячних фотоелектричних панелей можна користуватись цифрою 50 кВт·год/м² на рік.

Оціночні розрахунки дозволяють говорити, що 1 м² фотоелектричної панелі на протязі року може забезпечити енергією шляхове полотно для переміщення вантажу в 1 тону на відстань, що рівна довжині шляхової котушки $2\tau/3$ м, близько $625000 \cdot 1/\tau$ раз. Якщо $\tau=2,5$ м, то число переміщень складе приблизно 240000 разів, що для вантажу 1 тонна відповідає пропускній здатності шляху 1,5...2,6 хв. Оскільки шляхова котушка є основним елементом лінійного тягового двигуна, то структура системи енергопостачання ЛСТЕД буде модифікована до виду наведеному на рис. 5.

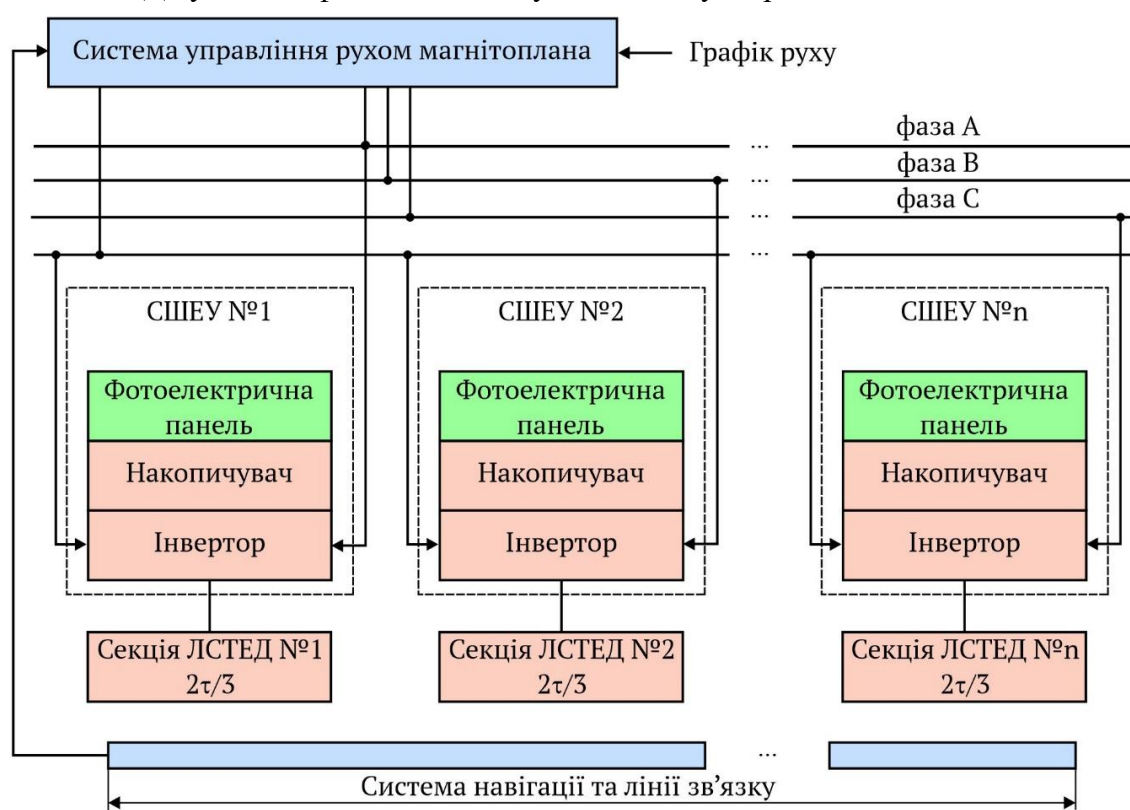


Рисунок 5 – Структурна схема енергопостачання ЛСТЕД від сонячних панелей (силові елементи виділені червоним кольором)

В структурі запропонованій на рис. 5 відпадає необхідність в передачі значних об'ємів енергії на великі відстані, що призводить до зменшення загальносистемних втрат. Основним силовим елементом такої системи є сонячна шляхова енергетична установка (СШЕУ) яка складається з фотоелектричної панелі, накопичувача та інвертора. Однотипні, малоенергоємні СШЕУ розміщуються вздовж магістралі забезпечуючи тим самим більш високий рівень надійності функціонування системи, кращу електромагнітну сумісність її елементів та безпеку експлуатації.

Враховуючи нерівномірність генерації електричної енергії фотоелектричними установками на протязі доби і року, для забезпечення незалежності системи енергопостачання до СШЕУ пред'являються наступні вимоги:

- накопичувач та інвертор повинні працювати в циклі розряду-заряду так аби до чергового розряду накопичена питома енергія була за величиною не меншою аніж це необхідно для транспортування одиниці ваги поїзда на одиницю довжини шляху;
- накопичувач повинен забезпечувати накопичення достатнього об'єму енергії в світлу пору доби аби нормально функціонувати вночі;
- накопичувач повинен мати певний запас енергії аби забезпечити безперебійну роботу транспортної системи протягом декількох діб за несприятливих погодних умов;
- накопичувач повинен мати здатність в літній період акумулювати стільки енергії аби взимку віддавати її поступово на навантаження гарантуючи цим незалежне енергопостачання транспортній системі на протязі року.

Вище викладеним вимогам може задовольняти накопичувач, що складається з акумулятора, іоністора, електролізера, ємності для зберігання водню та паливного елемента. Акумулятор в змозі забезпечити енергію для роботи СШЕУ годинами чи навіть декілька діб, іоністор маючи високі показники заряду/розряду оберігає акумулятор від значних струмів заряду/розряду. Інші три елементи покликані забезпечити роботу енергосистеми під навантаженням на протязі року.

Головною вимогою ж до сонячних панелей є вимога високого ККД.

Висновки

Виконаний аналіз балансу енергетичних потреб магнітолевітаційного поїзда і можливостей фотоелектричної системи енергозабезпечення показує, що цілком реально використовувати розподілену фотоелектричну систему в якості системи енергопостачання високошвидкісного наземного транспорту. Так, в умовах реальної забезпеченості сонячною енергією в Україні, електричної енергії, що генерується 1 м² фотоелектричних панелей достатньо щоб приблизно 1 тис. разів транспортувати магнітолевітаційний поїзд вагою 400 тонн на відстань 1 м.

Особливістю такого підходу до побудови системи енергозабезпечення є її глибока інтегрованість в систему управління магнітолевітаційним транспортом, підвищена надійність та покращена електромагнітна сумісність її елементів, екологічність.

ЛІТЕРАТУРА

1. 6 FPE Instruments, (Implementing the Prior by Thematic Areas of the Sixth Frame Program // European Commission. Community Research. – 2002. – № 2. – P. 4
2. Fritz E., Kluhspies J., Kircher R., Witt M., Blow L. Energy Consumption of Track-Based High-Speed Transportation Systems: Maglev Technologies in Comparison with Steel-Wheel Rail. *Transportation Systems and Technology*. Vol. 4(3). P. 134-155. <https://doi.org/10.17816/transsyst201843s1134-155>
3. Wenk M., Klühspies J., Blow L., Kircher R., Fritz E., Witt M., Hekler M. Maglev: Science Experiment or the Future of Transport? Practical Investigation of Future Perspectives and Limitations of Maglev Technologies in Comparison with Steel-Wheel-Rail. Germany: The International Maglev Board, 2018. 44 p.

4. Системи управління і енергозабезпечення магнітолевітаційного транспорту / Дзензерський В. О., Плаксін С. В., Погоріла Л. М., Толдаєв В. Г., Шкіль Ю. В. Київ: Наукова думка, 2014. 276 с.
5. Муха А.М., Плаксін С.В., Погоріла Л.М., Устименко Д.В. і Шкіль Ю.В. Комбінована система синхронізованого керування рухом і підвісом магнітоплана // Наука та прогрес транспорту. 1(97) (Жов 2022), 23–31. doi: <https://doi.org/10.15802/stp2022/265332>
6. Мхітарян Н.М. Геліоенергетика. Системи, технології, використання. – Київ: Наукова думка, 2002. – 320 с.

REFERENCES

1. 6 FPE Instruments, (Implementing the Prior by Thematic Areas of the Sixth Frame Program // European Commission. Community Research. – 2002. – № 2. – P. 4
2. Fritz E., Kluhspies J., Kircher R., Witt M., Blow L. Energy Consumption of Track-Based High-Speed Transportation Systems: Maglev Technologies in Comparison with Steel-Wheel Rail. Transportation Systems and Technology. Vol. 4(3). P. 134-155. <https://doi.org/10.17816/transsyst201843s1134-155>
3. Wenk M., Klühspies J., Blow L., Kircher R., Fritz E., Witt M., Hekler M. Maglev: Science Experiment or the Future of Transport? Practical Investigation of Future Perspectives and Limitations of Maglev Technologies in Comparison with Steel-Wheel-Rail. Germany: The International Maglev Board, 2018. 44 p.
4. Systemy upravlinnia i enerhozabezpechennia mahnitolevitatsiinoho transportu / Dzenzerskyi V. O., Plaksin S. V., Pohorila L. M., Toldaiev V. H., Shkil Yu. V. Kyiv: Naukova dumka, 2014. 276 s.
5. Mukha, A. M., Plaksin, S. V., Pohorila, L. M., Ustymenko, D. V., & Shkil, Y. V. (2022). Combined System of Synchronized Simultaneous Control of Magnetic Plane Movement and Suspension. Science and Transport Progress, 1(97), 23-31. DOI: <https://doi.org/10.15802/stp2022/265332>
6. Mkhitarian N.M. Helioenerhetyka. Systemy, tekhnolohii, vykorystannia. – Kyiv: Naukova dumka, 2002. – 320 s

Received 23.08.2024.
Accepted 28.08.2024.

Justification of the application of a distributed network of photoelectric converters to power a linear motor of magnetolevitation transport

Abstract. Modern high-speed transport is the basis of sustainable economic and social development of the state, society in compliance with environmental requirements. The concept of power supply of a linear motor of magneto-levitation transport from a distributed network of photovoltaic converters is substantiated. The basic power element of a track power plant is proposed in the form of a completed unit consisting of a solar panel, a storage device and an inverter, which operates on a load in the form of a "short" track coil. The use of a "short" track coil allows reducing electrical energy losses, since the traction force is formed only in the zone of interaction with the rolling stock, energy is not transmitted to unused sections of the track structure. By reducing the length of the working section, other energy indicators of the system as a whole are significantly suspended, in particular the power factor and efficien-

cy. Reducing the length of the working sections in conditions of high speeds of rolling stock will lead to increased requirements for reliability and speed of operation of track switches. The track switches are created on the basis of power semiconductor switches, which are controlled by modern microcontrollers. The effectiveness of the proposed structure depends on the solar activity in the region of the transport artery. The estimated solar activity indicators allow us to state that the weight indicators of the rolling stock that must move on the track with a "short" section correspond to the available resources.

Keywords: linear motor, photovoltaic system, power supply, magnetic levitation train, transport, storage device.

Плаксін Сергій Вікторович – професор, д.ф.-м.н., завідувачий відділом «Системи керування транспортними засобами», Інститут транспортних систем та технологій НАН України, Дніпро, Україна.

Муха Андрій Миколайович – професор, д.т.н., завідувачий кафедрою «Електротехніка та електромеханіка», Український державний університет науки і технологій, Дніпро, Україна.

Устименко Дмитро Володимирович – доцент, к.т.н., доцент кафедри «Електротехніка та електромеханіка», Український державний університет науки і технологій, Дніпро, Україна.

Доскоч Володимир Ігорович – аспірант кафедри «Електротехніка та електромеханіка», Український державний університет науки і технологій, Дніпро, Україна.

Сандул Тарас Анатолійович – аспірант кафедри «Електротехніка та електромеханіка», Український державний університет науки і технологій, Дніпро, Україна.

Куліков Сергій Миколайович – аспірант кафедри «Електротехніка та електромеханіка», Український державний університет науки і технологій, Дніпро, Україна.

Plaksin Serge Viktorovich – professor, doctor of physical and mathematical sciences, head of the department of vehicle control systems, institute of transport systems and technologies of the National Academy of Sciences of Ukraine, Dnipro, Ukraine.

Mukha Andrew Mykolayovych – Professor, doctor of technical sciences, head of the department of electrical engineering and electromechanics, Ukrainian State University of Science and Technology, Dnipro, Ukraine.

Ustymenko Dmytro Volodymyrovych – associate professor, candidate of technical sciences, associate professor of the department of electrical engineering and electromechanics, Ukrainian State University of Science and Technology, Dnipro, Ukraine.

Doskoch Volodymyr Igorovich – postgraduate student of the department of electrical engineering and electromechanics, Ukrainian State University of Science and Technology, Dnipro, Ukraine.

Sandul Taras Anatolievich – postgraduate student of the department of electrical engineering and electromechanics, Ukrainian State University of Science and Technology, Dnipro, Ukraine.

Kulikov Serhii Mykolayovych – postgraduate student of the department of electrical engineering and electromechanics, Ukrainian State University of Science and Technology, Dnipro, Ukraine.

ПОБУДОВА ДЕРЕВА ШТЕЙНЕРА ЗА ДОПОМОГОЮ МЕТОДА КЛАСТЕРИЗАЦІЇ

Анотація. В даній роботі розглядається метод побудови дерева Штейнера для оптимізації мережесих структур у розподілених комп'ютерних системах. Основна мета роботи полягає у дослідженні та впровадженні вдосконаленого алгоритму для знаходження точок Штейнера за допомогою методу кластеризації. Основна ідея методу полягає у використанні специфічного підходу до визначення точок Штейнера, що оптимізують під'єднання заданих точок у просторі. Метою цього підходу є зменшення обчислювальної складності, зберігаючи при цьому адекватну точність у побудові дерева Штейнера. Через спрощений підхід до кластеризації та визначення точок Штейнера, цей метод має потенціал значно оптимізувати процес вирішення поставленого завдання, особливо в сценаріях з великою кількістю точок. Для визначення його ефективності проведено дослідження на графах з чотирма, п'яти та шести вершинами розташованими на площині випадковим чином. Тестування проводилось за допомогою спеціального програмного забезпечення, написаного мовою Python. Загалом, дослідження показало, що метод кластеризації є ефективним інструментом для визначення точок Штейнера, що дозволяє знизити обчислювальну складність та забезпечити адекватну точність у побудові дерева Штейнера. Подальші дослідження в цьому напрямку можуть сприяти вдосконаленню методів оптимізації мережесих структур, що є важливим для широкого спектру практичних застосувань.

Ключові слова: дерево Штейнера, мінімальне остовне дерево, метод оптимізації, кластеризація графа.

Постановка проблеми. Задача Штейнера (або задача дерева Штейнера) - це комбінаторна оптимізаційна задача в графах, яка полягає в пошуку найкоротшого можливого дерева (графа) для з'єднання заданого набору вершин, які називаються терміналами. Відрізняючись від більш традиційних мінімальних остовних дерев, дерево Штейнера може включати додаткові точки, які не є частиною первісного набору, але які допомагають скоротити загальну довжину дерева. Додаткові точки відомі як точки Штейнера.

Ця задача є комбінаторною оптимізаційною задачею NP-складності і може бути дуже складною для розв'язання, особливо на великих графах. Тому існують різні алгоритми, які намагаються наблизити оптимальне рішення або знайти розв'язок в прийнятний час.

Метод Штейнера має безліч застосувань у різних галузях, включаючи телекомунікації (для з'єднання мереж з мінімальною довжиною кабелів), маршрутизацію мереж, виробництво печатних плат, дизайн логістичних систем і багато інших областей.

Найпростіший приклад задачі Штейнера, коли кількість точок дорівнює трьом. Рішення задачі Штейнера для трьох точок було знайдене ще в XVII столітті. П. Ферма, Є. Торричеллі та Б. Кавальєрі в свій час довели, що дерево Штейнера для трикутника будується за допомогою однієї додаткової точки (при умові що всі кути трикутника менші 120°), котра називається точкою Ферма-Торричеллі-Штейнера. Якщо один з кутів трикутника більше або дорівнює 120° , то дерево Штейнера складається зі сторін цього кута [1].

На Рис. 1 показаний процес побудови точки Штейнера S для трьох точок A, B, C . Для пошуку точки Штейнера треба на будь-якій стороні початкового трикутника добудувати назвні рівносторонній трикутник $A'BC$. Точка S перетин кола, описаного навколо трикутника $A'BC$, та відрізка AA' буде являтися точкою Штейнера. Також цю точку можна знайти, якщо добудувати рівносторонні трикутники к будь-яким іншим сторонам [2].

Рішення задачі Штейнера для трьох точок є основою для побудови найкоротшої сітки Штейнера для великої кількості точок. Відомо, що для будь-якої кінцевої системи точок на площині існує кінцева множина сіток Штейнера. Для знаходження найкоротшого шляху серед множини точок треба знайти всі сітки Штейнера, а потім вибрати найкоротшу. Однак точне рішення задачі потребує розгляду великої кількості варіантів, так що навіть найкращі алгоритми, що виконуються на самих швидкодіючих комп'ютерах, не в змозі надати рішення для великої множини заданих точок за реально прийнятний час. Більше того, задача Штейнера належить до класу задач, для яких, на думку багатьох сучасних дослідників, ефективні алгоритми ніколи не будуть знайдені. З огляду на затребуваність задачі для практичних додатків, актуальною є побудова нових алгоритмів, у тому числі дають наближене рішення задачі Штейнера [3].

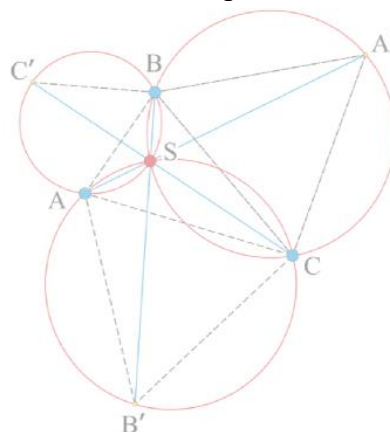


Рисунок 1 - Побудова точки Штейнера S для трикутника ABC геометричним методом[1]

Мета. В цій роботі передбачається дослідити вдосконалений алгоритм знаходження точок Штейнера для побудови мереж в розподілених комп'ютерних системах за

допомогою розробленого програмного забезпечення. Для програмної реалізації обрано мову Python.

Методика. Попри значну актуальність та важливість побудови дерева Штейнера, варто зазначити, що ця задача належить до надскладних проблем. Це означає, що не існує відомого алгоритму, який здатен знайти точне оптимальне рішення для довільного вхідного набору даних у поліноміальний час. В результаті, більшість існуючих алгоритмів зосереджуються на наближених або евристичних рішеннях, які можуть не гарантувати оптимальності, але забезпечують практично прийнятні результати у розумний час.

Однак, навіть ці методи часто стикаються з обмеженнями у вигляді високих обчислювальних витрат та недостатньої точності, особливо при роботі з великими або складними графами. Таким чином, розробка альтернативних алгоритмів, які здатні ефективніше вирішувати цю задачу, є критично важливою для подальшого прогресу в областях, де використовується дерево Штейнера.

Одним з таких методів може бути метод кластеризації точок на площині. Основна ідея методу полягає у використанні специфічного підходу до визначення точок Штейнера, що оптимізують під'єднання заданих точок у просторі. Метою цього підходу є зменшення обчислювальної складності, зберігаючи при цьому адекватну точність у побудові дерева Штейнера. Через спрощений підхід до кластеризації та визначення точок Штейнера, цей метод має потенціал значно оптимізувати процес вирішення поставленого завдання, особливо в сценаріях з великою кількістю точок. Для визначення його ефективності проведемо дослідження на графах з чотирма, п'яти та шести вершинами розташованими на площині випадковим чином.

Програма Steiner Point Calculator для знаходження точки Штейнера для трикутника. Експериментальна програма для обчислення та візуалізації точки Штейнера виконана мовою Python з використанням стандартних бібліотек: Tkinter для створення графічного інтерфейсу користувача (GUI), Matplotlib для малювання графіків та NumPy для математичних обчислень. Основна частина програми, функція обчислення точки Штейнера, здійснюється за допомогою методу оптимізації Nelder-Mead, який є частиною бібліотеки `scipy.optimize`. Цей метод дозволяє знайти точку всередині трикутника, яка мінімізує загальну відстань до трьох вершин.

Інтерфейс користувача програми складається з полів для вводу координат трьох точок, що формують трикутник, та вікно для графічного відображення. Результат представлений як текстове повідомлення з координатами точки та візуалізацією на графіку, де зображено вхідні точки, їхнє з'єднання та розташування точки Штейнера. Інтерфейс модуля Steiner Point Calculator представлений на рис. 2.

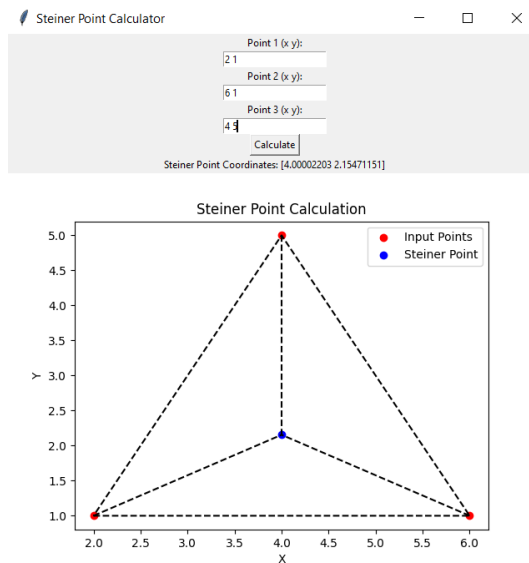


Рисунок 2 - Інтерфейс модуля Steiner Point Calculator

Виконання коду відбувається наступним чином:

Імпорт необхідних бібліотек.

Опис функцій:

- `calculate_steiner_point(p1, p2, p3)`: Розраховує точку Штейнера для заданих точок $p1$, $p2$, $p3$. Використовує метод оптимізації Nelder-Mead з `scipy.optimize` для мінімізації суми відстаней від точки до вершин трикутника.

- `distance_sum(point, p1, p2, p3)`: Допоміжна функція всередині `calculate_steiner_point`, яка обчислює суму відстаней від заданої точки до трьох точок трикутника.

- `on_calculate()`: Зчитує координати з текстових полів, обчислює точку Штейнера, викликає функцію `plot_graph` для візуалізації результатів та виводить координати точки Штейнера.

- `plot_graph(p1, p2, p3, steiner_point)`: Малює графік з вхідними точками та точкою Штейнера.

- Програма виконує функцію `on_calculate`, яка перетворює введені дані в числовий формат, перевіряє їх валідність та викликає `calculate_steiner_point`.

- `calculate_steiner_point` використовує метод оптимізації для знаходження точки Штейнера.

- Після знаходження точки Штейнера, `plot_graph` візуалізує вхідні точки та точку Штейнера на графіку.

- Координати точки Штейнера відображаються на інтерфейсі.

Фрагмент коду програми представлено на Рис. 3.

```
def calculate_steiner_point(p1, p2, p3):  
    def distance_sum(point, p1, p2, p3):  
        return (  
            np.linalg.norm(np.array(point) - np.array(p1)) +  
            np.linalg.norm(np.array(point) - np.array(p2)) +  
            np.linalg.norm(np.array(point) - np.array(p3))  
        )  
    initial_point = [(p1[0] + p2[0] + p3[0]) / 3, (p1[1] + p2[1] + p3[1]) / 3]  
    result = minimize(distance_sum, initial_point, args=(p1, p2, p3), method='Nelder-Mead')  
    if result.success:  
        return result.x  
    else:  
        raise ValueError("Optimization failed to find the Steiner point")
```

Рисунок 3 - Фрагмент коду програми Steiner Point Calculator»

Для чотирьох точок

Вибір Точок: Візьмемо чотири точки, розміщених випадково на площині з такими координатами (Рис. 4):

A (2, 3); B(4, 9); C(13, 7); D(10, 1)

Довжина ребер рахується за формулою (1):

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

AB=6,32; BC= 9,22; CD=6,71; DA=8,25.

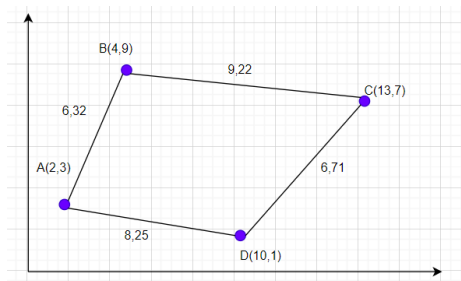


Рисунок 4 - Початкові точки графа

Кластеризація: Спочатку ми поділимо площу на 2 трикутника (Рис. 5). Далі знаходимо точки Штейнера (ТШ) для кожного з цих трикутників за допомогою програми Steiner Point Calculator.

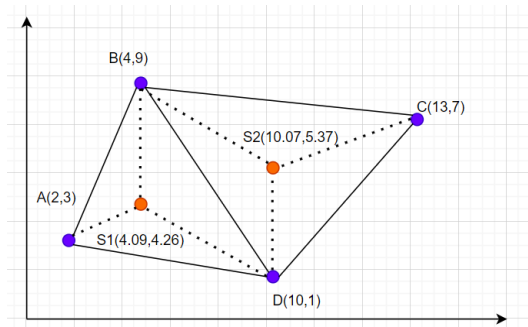


Рисунок 5 - Розділення графа на кластери

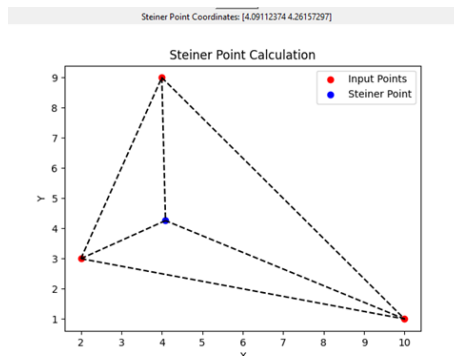


Рисунок 6 - Знаходження ТШ для трикутника

Координати нових точок: S1 (4.09, 4.26); S2(10.07, 5.37); Об'єднаємо точки в дерево Штейнера (ДШ): по дві найближчі до отриманих та з'єднаємо їх.

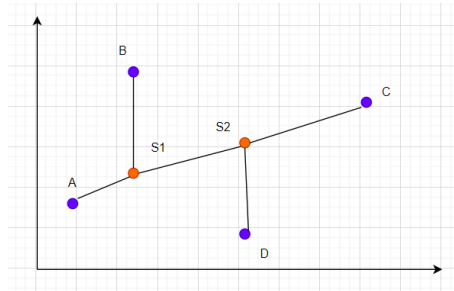


Рисунок 7 - ДШ для 4-х точок прямим з'єднанням

Для порівняння знайдемо суму всіх ребер в отриманому дереві Штейнера та мінімальним остовним деревом (МОД) для чотирьох початкових точок.

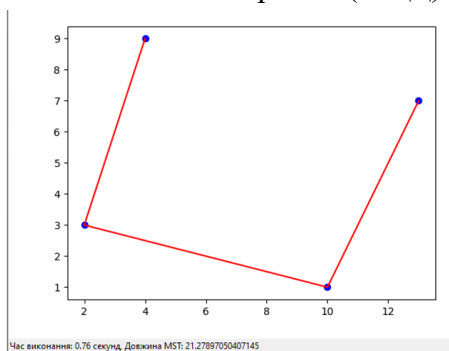


Рисунок 8 - МОД для 4 точок за допомогою EMST

Довжина МОД 21.28

Довжина ребер дерева Штейнера за формулою (1):

$A-S1 = 2,44$; $B-S1 = 4,74$; $C-S2 = 3,35$; $D-S2 = 4,37$; $S1-S2 = 6,08$;

Загальна сума довжин ребер після з'єднання на графі: 20,98

Загальна сума довжин ребер ДШ за допомогою EMST: 20,98

Загальна сума довжин ребер МОД за допомогою EMST: 21,28

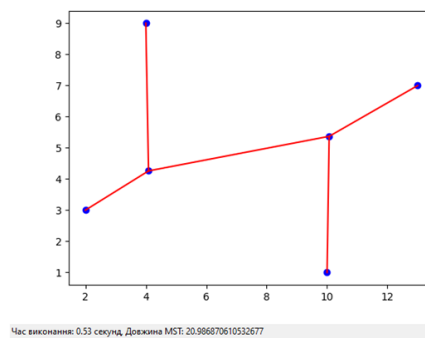


Рисунок 9 - ДШ для 4-х точок за допомогою EMST

Для п'яти точок

Візьмемо п'ять точок, розміщених випадково на площині з такими координатами (Рис. 10):

A (1, 3); B(4, 8); C(9, 8); D(10, 3); E(6, 1)

Довжина ребер: AB=5,83; BC=5,00; CD=5,10; DE=4,47; EA=5,39.

Кластеризація: Спочатку ми поділимо площу на три трикутника (Рис. 11).

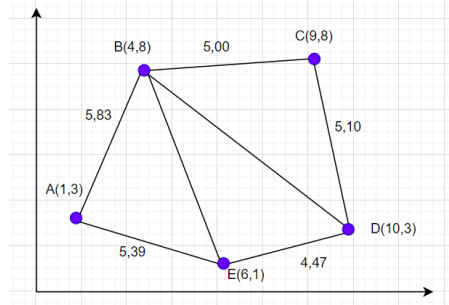


Рисунок 10 - Початкові точки графа з п'ятьох вершин

Знаходження Точок Штейнера для Кластерів: Для кожного з цих кластерів з трьох точок ми визначимо відповідну точку Штейнера за допомогою програми Steiner Point Calculator (Рис. 11).

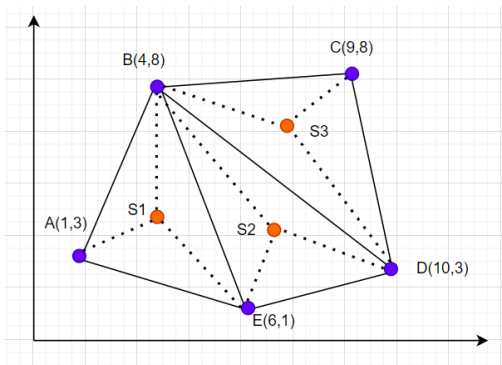


Рисунок 11 - Розділення графа з п'ятьма точками на кластери

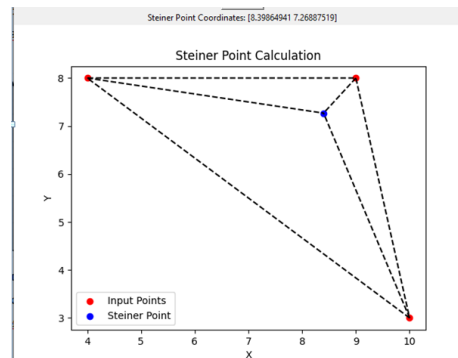


Рисунок 12 - Знаходження ТШ для трикутника

S1 (3.06, 3.66); S2(7.07, 2.93); S3(8.40, 7.27)

Об'єднання отриманих точок S1, S2, S3 в новий трикутник та знаходження точки Штейнера для нового трикутника S4(6.54, 3.78)

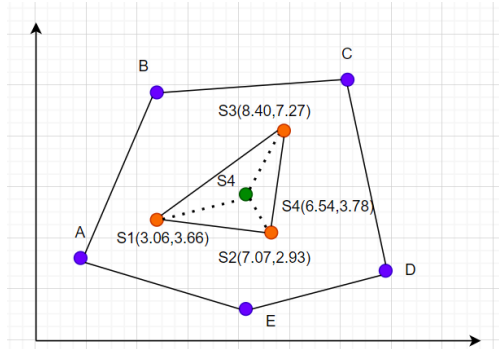


Рисунок 13 - Результуюча ТШ для графа з п'яти точок

Нова точка в отриманому трикутнику і буде розрахунковою точкою Штейнера для початкових п'яти вершин (Рис. 14).

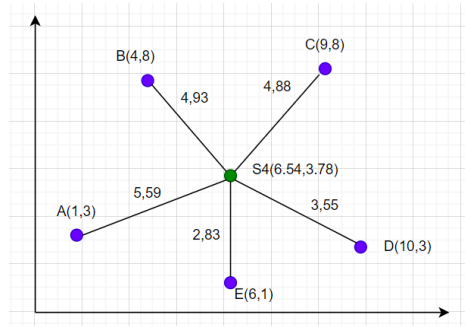


Рисунок 14 - ДШ для 5-и точок прямим з'єднанням

Для порівняння знайдемо суму всіх ребер в отриманому дереві Штейнера та мінімальним остовним деревом для п'ятих початкових точок за допомогою програми EMST.

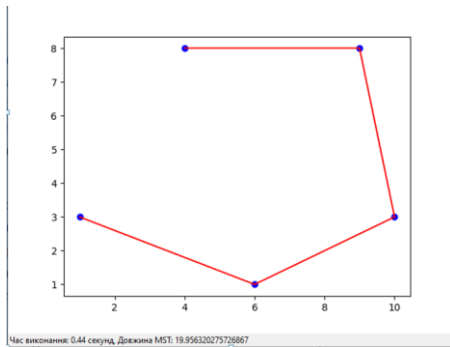


Рисунок 15 - МОД для п'яти точок за допомогою EMST

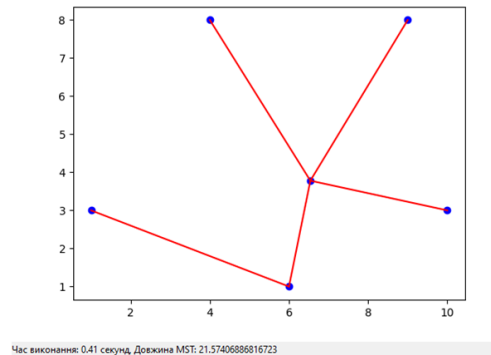


Рисунок 16 - ДШ для п'яти точок за допомогою EMST

Довжина ребер ДШ прямим з'єднанням(Рис. 14): 21.78.

($A-S4=5.59$; $B-S4=4.93$; $C-S4=4.88$; $D-S4=3.55$; $E-S4=2.83$).

Довжина МОД за допомогою EMST(Рис 15): 19.9.

Довжина ДШ за допомогою програми EMST(Рис.16): 21,5.

Для шести точок

Вибір Точок: Візьмемо шість точок, розмішених випадково на площині з такими координатами:

A (1, 2); B(2, 7); C(6, 5); D(8, 3.5);E(5, 1);F(3, 3).

потім поділимо площу в середині графа на трикутники (Рис 17).

Вага ребер графа дорівнює відстані між його точками:

A-B=5.10; B-C=4.47; C-D=2.50; D-E=3.91; E-A=4.12;

A-F=2.24; B-F=4.12; C-F=3.61; D-F=5.03; E-F=2.83.

Тепер за допомогою програми Steiner Point Calculator знайдемо точки Штейнера для кожного трикутника (Рис. 18):

$$S1=(A; B; F) = (2.53; 3.19)$$

$$S2=(F; B; C) = (3.73; 4.72)$$

$$S3=(F; C; D) = (5.99; 4.69);$$

$$S4=(F; D; E) = (5.01; 1.81);$$

$$S5=(F; E; A) = (2.95; 2.71);$$

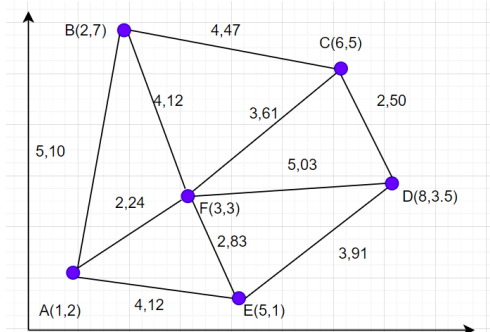


Рисунок 17 - Початкові точки графа з шести вершин

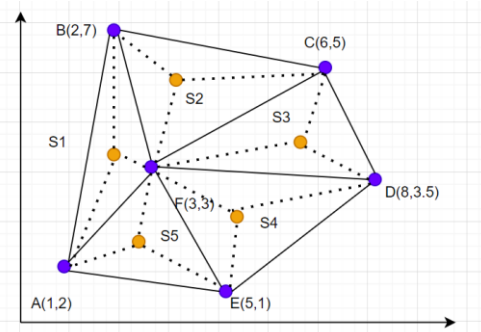


Рисунок 18 - Розділення графа з шести точками на кластери

Далі розділяємо площу графа з п'яти точок на три трикутника та знаходимо для кожного точку Штейнера H_1, H_2, H_3 (Рис. 19).

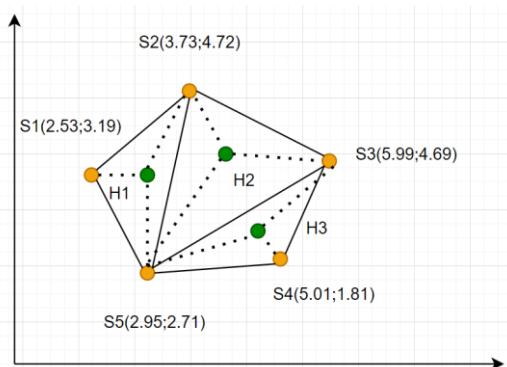


Рисунок 19 - ТШ для кластерів другої ітерації

$$H1(S1; S2; S5)=(2.71; 3.18)$$

$$H2(S2; S3; S5)=(3.85; 4.54)$$

$$H3(S3; S4; S5)=(4.61; 2.61)$$

Точка Штейнера (Рис. 21) для трикутника $H_1-H_2-H_3$ буде передбачуваною точкою Штейнера (T) для шести початкових вершин графа $T(3.63; 3.48)$.

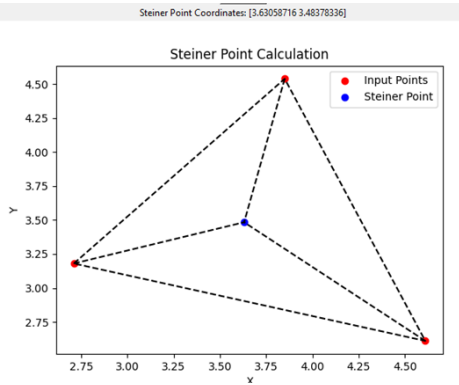


Рисунок 20 - Результуюча ТШ для графа з шести точок

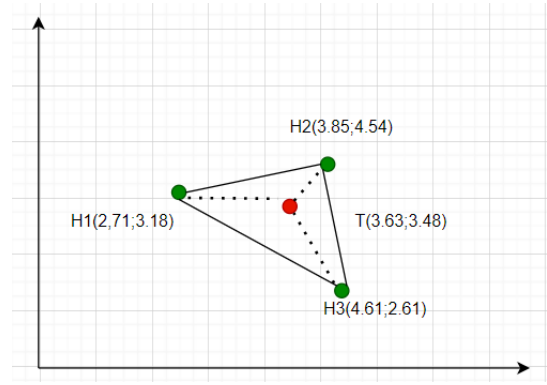


Рисунок 21 - Результуюча ТШ для графа з шести вершин

Для порівняння з МОД знайдемо суму всіх ребер отриманого дерева Штейнера за формулою:

$$A-T = 3.02; B-T = 3.88; C-T = 2.82; D-T = 4.37; E-T = 2.83; F-T = 0.79.$$

Сума дорівнює 17,71

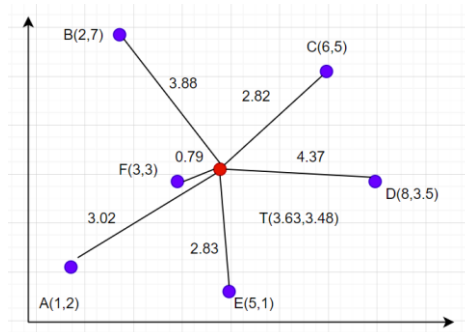
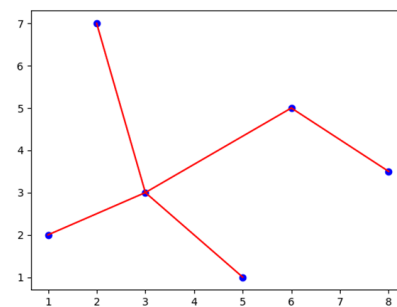


Рисунок 22 - ДШ для графа з шести точок прямим з'єднанням

Знайдемо МОД початкових шести точок за допомогою програми EMST



Час виконання: 0.33 секунд, Довжина MST: 15.293152003327629

Рисунок 23 - МОД для графа з шести точок за допомогою EMS

Довжина МОД дорівнює 20,23.

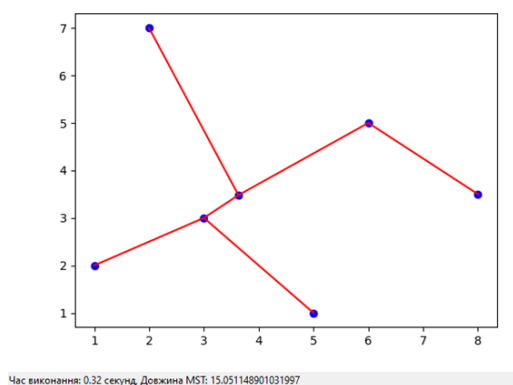


Рисунок 24 - ДШ для графа з шести точок за допомогою EMST

Довжина ребер ДШ прямим з'єднанням(Рис. 22): 17,71.

Довжина МОД за допомогою EMST(Рис. 23): 15,29.

Довжина ДШ за допомогою програми EMST(Рис. 24): 15,05.

Результати дослідження: При розробці вдосконаленого алгоритму знаходження точок Штейнера були проведені дослідження мінімальних остовних дерев та дерев Штейнера побудованих різними методами на графах з різною кількістю початкових вершин.

Чотири точки:

За методом прямого з'єднання та EMST отримано однакову довжину ребер (20,98), свідчачи про однакову ефективність обох методів у цьому конкретному випадку.

Метод МОД за допомогою EMST показав дещо вищу довжину ребер (21,28), що може вказувати на його меншу ефективність для невеликої кількості точок.

П'ять точок:

Пряме з'єднання дало найбільшу довжину ребер (21.78), що свідчить про його відносно меншу ефективність у цьому сценарії.

Метод EMST з деревом Штейнера показав кращі результати (21,5) порівняно з прямим з'єднанням.

Найкращі результати показав метод МОД за допомогою EMST (19.9), вказуючи на його вищу ефективність для цієї кількості точок.

Шість точок:

Пряме з'єднання знову показало найбільшу довжину ребер (17,71).

Метод EMST з деревом Штейнера виявився найефективнішим, демонструючи найменшу довжину ребер (15,05).

Метод МОД за допомогою EMST показав трохи вищу довжину ребер (15,29) порівняно з методом EMST з деревом Штейнера, але все ще кращі результати, ніж пряме з'єднання.

1. Результати вказують на необхідність адаптації методів залежно від конкретних сценаріїв та кількості точок для досягнення оптимальних результатів.

2. У випадку з чотирма точками, методи прямого з'єднання ДШ та ДШ за допомогою програми EMST показали ідентичні результати, тоді як МОД за допомогою EMST був менш ефективним.

3. Для п'яти та шести точок метод МОД за допомогою EMST виявився більш ефективним, ніж інші методи, особливо порівняно з прямим з'єднанням.

Висновки:

1. Дослідження ефективності кластеризації в визначенні точок Штейнера показало, що використання цього методу може бути ефективним у зниженні обчислювальної складності, особливо у сценаріях з великою кількістю точок. Цей метод демонструє адекватну точність у побудові дерева Штейнера, що робить його корисним для оптимізації процесу розв'язання задач.

2. Детальне дослідження, виконане на графах з різною кількістю точок, вказує на те, що різні методи знаходження оптимальних рішень виявляються більш або менш ефективними залежно від конкретної кількості точок. Цей аналіз підкреслює необхідність адаптації алгоритму до специфіки задачі та підтверджує його універсальність і гнучкість.

3. Розроблений експериментальний зразок програми реалізації алгоритму дозволяє знаходити точку Штейнера для графа з трьох точок.

4. Порівняльний аналіз різних методів (пряме з'єднання, ДШ за допомогою програми EMST, МОД за допомогою EMST) показав, що ефективність кожного методу значно відрізняється залежно від кількості точок. Наприклад, виявилось, що для п'яти точок метод МОД за допомогою EMST є найбільш ефективним, тоді як для чотирьох точок методи прямого з'єднання та EMST показали ідентичні результати.

Базуючись на отриманих результатах, важливо провести подальші дослідження з використанням більшої кількості точок і різних конфігурацій графів. Це допоможе більш детально зрозуміти масштаби застосовності розробленого алгоритму та його ефективність у широкому спектрі задач.

Отримані результати можуть бути корисними в практичних застосуваннях, де необхідна оптимізація мережевих структур, наприклад, у телекомунікаціях або при плануванні транспортних мереж. Адаптація алгоритму до специфічних умов цих застосувань може значно підвищити їх ефективність.

ЛІТЕРАТУРА

1. Melzak, Z. A. Companion to Concrete Mathematics. John Wiley & Sons, Inc., 1973.
2. Winter, Pawel. Steiner Problem in Networks: A Survey. Networks, 17(2), 129–167.
3. The Shortest-Path Problem
URL: https://mathweb.ucsd.edu/~ronspubs/89_01_shortest_network.pdf
4. Ольшевський А.І. Алгоритми побудови маршруту при груповому розсиланні мережевих пакетів даних дистанційного навчання на базі ДонНТУ. Штучний інтелект, № 4, 2007, 483-490.
5. Wirsansky, E. Hands-On Genetic Algorithms with Python. Packt Publishing, 2020, 346 p.
6. Литвиненко В.А., Ховансков С.А., Максюта Д.Ю. Адаптивний алгоритм побудови дерева Штейнера, КПУ, 2016, 9 с.

7. Библиотека алгоритмов обчислювальної геометрії

URL: <https://www.cgal.org/>.

8. Kruskal-based approximation algorithm for the multi-level Steiner tree problem. URL:

<https://arxiv.org/pdf/2002.06421v2.pdf>

REFERENCES

1. Melzak, Z. A. Companion to Concrete Mathematics. John Wiley & Sons, Inc., 1973.

2. Winter, Pawel. Steiner Problem in Networks: A Survey. Networks, 17(2), 129–167.

3. The Shortest-Network Problem.

URL: https://mathweb.ucsd.edu/~ronspubs/89_01_shortest_network.pdf

4. Olshevskiy, A.I. Algorithms for Routing in Group Distribution of Network Data Packets for Distance Learning Based on DonNTU. Artificial Intelligence, No. 4, 2007, 483-490.

5. Wirsansky, E. Hands-On Genetic Algorithms with Python. Packt Publishing, 2020, 346 p.

6. Lytvynenko, V.A., Khovanskov, S.A., Maksyuta, D.Yu. Adaptive Algorithm for Building a Steiner Tree, KPU, 2016, 9 p.

7. Library of Computational Geometry Algorithms. URL: <https://www.cgal.org/>

8. Kruskal-based Approximation Algorithm for the Multi-Level Steiner Tree Problem. URL:

<https://arxiv.org/pdf/2002.06421v2.pdf>

Received 01.08.2024.

Accepted 04.09.2024.

Construction of a Steiner Tree Using the Clustering Method

This paper examines the method of constructing a Steiner tree for optimizing network structures in distributed computer systems. The primary goal of the work is to investigate and implement an advanced algorithm for finding Steiner points using the clustering method. The main idea of the method is to use a specific approach to determining Steiner points that optimize the connection of given points in space. The objective of this approach is to reduce computational complexity while maintaining adequate accuracy in constructing the Steiner tree. Due to the simplified approach to clustering and determining Steiner points, this method has the potential to significantly optimize the problem-solving process, especially in scenarios with a large number of points. To determine its effectiveness, studies were conducted on graphs with four, five, and six vertices randomly located on a plane. Testing was carried out using special software written in Python. Overall, the research showed that the clustering method is an effective tool for determining Steiner points, allowing for reduced computational complexity and providing adequate accuracy in constructing the Steiner tree. Further research in this direction may contribute to the improvement of network structure optimization methods, which is important for a wide range of practical applications.

Keywords: Steiner tree, minimum spanning tree, optimization method, graph clustering.

Глушков Олег Володимирович - аспірант кафедри електронних обчислювальних машин Українського державного університету науки і технологій (м. Дніпро).

Hlushkov Oleh - Postgraduate Student, Department of Electronic Computing Machines, Ukrainian State University of Science and Technology, Dnipro, Ukraine.

ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ ТОЧНОСТІ ПРОКАТКИ НА СТАНІ 800

Анотація. Метою роботи є експериментальне дослідження процесу прокатки на стані 800 квадратних профілів для визначення впливу параметрів прокатки на розміри і довжину розкату. При дослідженні визначили коливання маси злитка, розмірів розкату, довжини розкату, отримали рівняння регресії і коефіцієнти кореляції залежностей розмірів та довжини розкату від маси злитка, змінення міжвалкового зазору і затримки часу перед чистою кліткою. Встановлено, що коливання маси злитку відбувається в значних межах – 638 кг, що спричинено коливаннями маси плавки. Коливання маси злитка в межах одного піддона спричинено розмірами виливниць і знаходиться в межах 154 кг. Коливання розмірів розкату здійснюється в межах: товщина -1,0 мм, ширина – 0,5 мм, довжина – 5,3 м. Зміна міжвалкового зазору може бути використана для корекції довжини готового розкату в невеликому діапазоні довжин, якщо коливання маси відбуваються в межах точності розливу на одному піддоні. Необхідно додатково вживати заходів щодо стабілізації маси злитка. Вплив затримок у лінії стану, які виникають при прокатці і призводять до нестабільності температурного режиму прокатки, для даного типу профілів не позначається суттєво на розмірах поперечного перерізу і довжині розкату.

Ключові слова: маса злитку, товщина та ширина розкату, довжина розкату, струм двигуна, міжвалковий зазор.

Постановка проблеми і ціль дослідження. В реальних умовах прокатки розміри прокату та заготовки не залишаються постійними. Спостерігається одночасний вплив кількох факторів на довжину розкату, зокрема, коливань маси злитка та маси заготовки через неточний розкрій розкату, змінення температури розкату та поперечних розмірів готового прокату через знос валків. Це, як правило, не дозволяє отримати однозначні залежності між розмірами прокату і параметрами прокатки [1].

Дослідження коливань розмірів прокату здійснювалось для прокатних профілів, які є готовим продуктом, тому що коливання розмірів поперечного перерізу і довжини розкату призводять до появи штанг немірної довжини при порізці на пилах [2]. Розкрій розкату не завжди дозволяє отримати масу заготовки, з якої на наступному переділі будуть отримані тільки штанги мірної довжини. Тому підвищення точності маси квадратних або круглих заготовок дозволяє зменшити коливання довжини розкату і появу штанг немірної довжини і запобігти появі коротких заготовок.

Метою роботи є експериментальне дослідження впливу технологічних параметрів та параметрів обладнання стана 800 на довжину прокату з ранжуванням їх за ступенем впливу.

Матеріали і результати досліджень. У промисловому експерименті отримали дані про поперечні розміри прокатних профілів та довжину розкатів під час прокатки квадратного профілю 125 мм на правій та квадратного профілю 150 мм на лівій нитці стана 800.

Під час проведення досліджень на правому боці стана було прокатане дві плавки. При прокатці першої плавки були виконані затримки при передачі розкатів з передчистої кліті в чистову кліть і зміна міжвалкового зазору чистової кліті, другу плавку прокатали в режимі пасивного експерименту.

При проведенні експериментальних досліджень реєстрували струм та частоту обертання двигуна головного приводу, здійснювали відбір проб, що вирізалися на пилках гарячого різання з середини розкату, після чого обмірювалися вручну за допомогою штангенциркуля. Довжину розкату контролювали за положенням заднього кінця розкату відносно нанесеної на огородження відповідного рольганга розмітки. Положення заднього кінця фіксували за допомогою цифрового фотоапарата з подальшою роздруковкою фотографій та визначенням довжини розкату. На рис.1 наведено фотографію положення заднього кінця розкату.



Рисунок 1 - Визначення положення заднього кінця розкату

Струм та частоту обертання двигуна головного приводу вимірювали тахогенератором та датчиком струму, що були підключені до розташованого в машинній залі стана 800 комп'ютера за допомогою модуля введення/виведення аналогових сигналів PCL-818L. Експериментальні осцилограми струму та частоти обертання двигуна наведені на рис. 2.

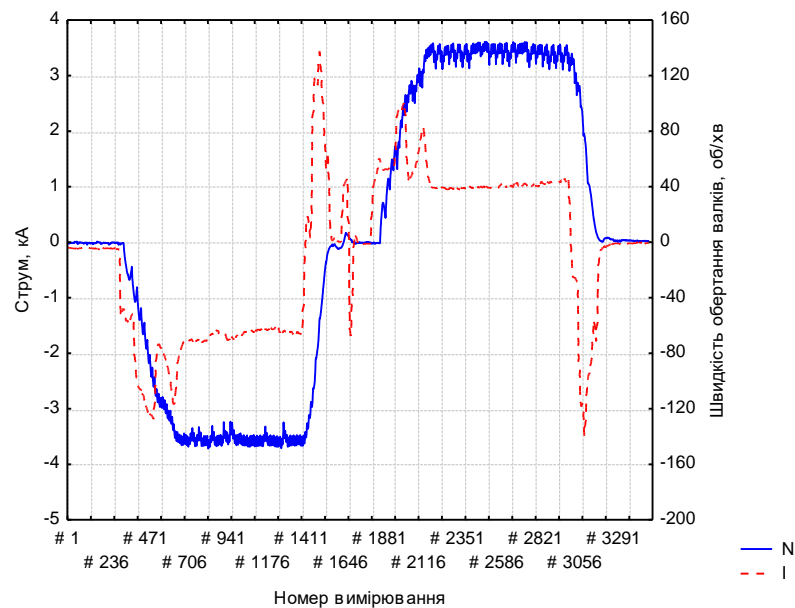


Рисунок 2 - Осцилограми струму I та частоти N двигуна чистової прокатної кліти у передчистовому та чистовому проходах

На рис.3 наведені реалізації розмірів розкатів, маси розкатів і струмів в передчистовому і чистовому проходах чистової кліти стана. По отриманим результатам розмірів поперечного перерізу розкатів і довжини розкатів були визначені маси розкатів, які відрізняються від маси злитків на величину передньої і задньої обрізей блюмса на парогідравлічному ножі, які є нормативною величиною і мають незначні коливання.

Визначені за результатами пасивного експерименту значення статистичних параметрів розміру прокату наведені у таблиці 1. Отримані дані свідчать, що маса розкатів (злитків) у плавці коливається у широких межах, що обумовлено процесом розливання сталі у виливниці в конвертерному цеху, адже перший піддон наливається на задану масу, а другий заповнюється залишком сталі у ковші. Наприклад, у двох плавках маса перших піддонів становить 7175 кг, а других – 6682 кг. Коливання маси розкатів (злитків) на плавці становить 830 кг, у піддоні – 154 кг.

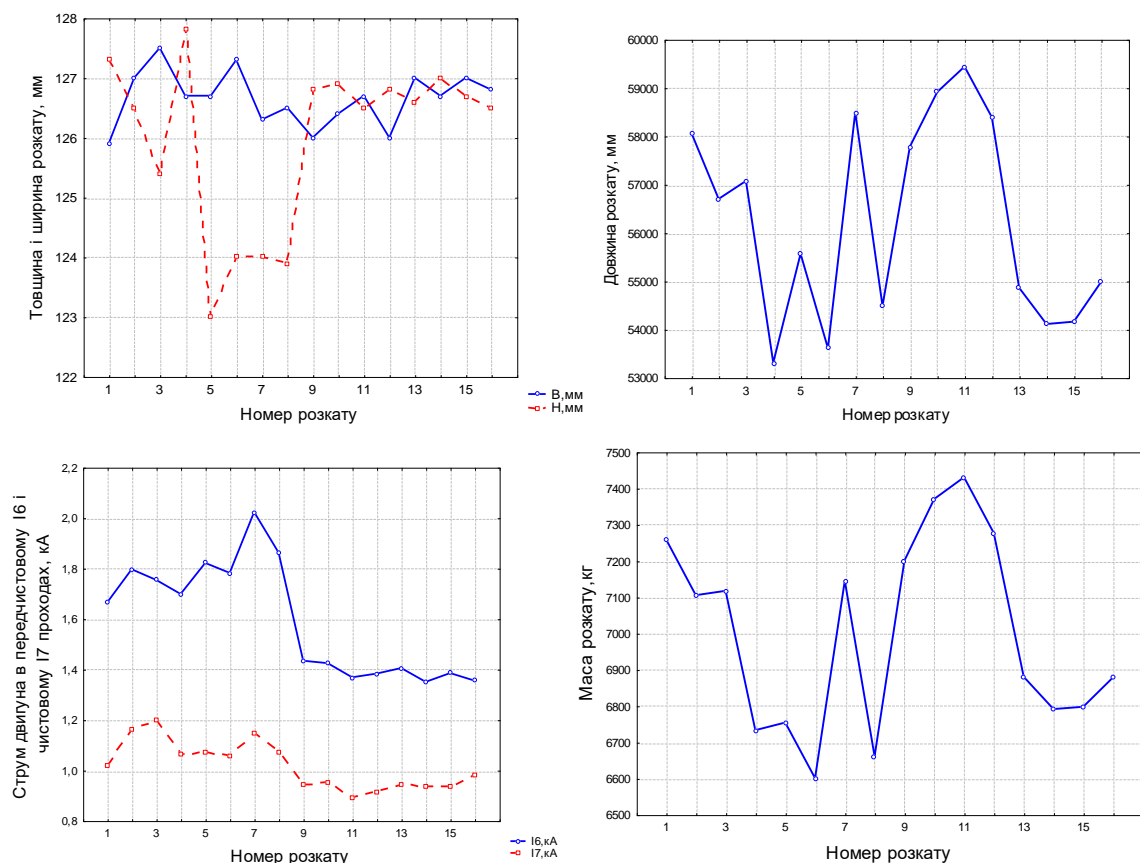


Рисунок 3 - Реалізації розмірів розкатів, маси розкатів та струму двигуна

Таблиця 1

Результати статистичної обробки експериментальних даних для прокатки квадрата 125 у чистовій клітці стана 800 (друга плавка)

Параметри	Кількість вимірювань	Середнє значення	Мінімальне значення	Максимальне значення	Діапазон змінення	Дисперсія	Середнє квадр. відхилення
Товщина, мм	8	126,58	126,00	127,00	1,000	0,162	0,403
Ширина, мм	8	126,72	126,50	127,00	0,500	0,033	0,183
Площа поперечного перерізу, мм ²	8	16040	15976	16090	114	2112	45,95
Довжина розкату, м	8	56,57	54,11	59,43	5,325	5,110	2,260
Маса розкату, кг		7078	6791	7430	638	71966	268
Струм двигуна у передчисто-вому проході, кА	8	1,39	1,35	1,44	0,090	0,00109	0,033
Струм двигуна у чистовому проході, кА	8	0,94	0,89	0,98	0,090	0,00073	0,027

Крім цього, прокатка здійснюється не по поддонах, а в різнобій. Тобто злитки в плавці можуть прокатуватися з обох поддонів одночасно, що створює додаткові труднощі при порізці розкатів і що відбувалось при проведенні експериментальних досліджень і це вплинуло на обробку отриманих даних.

Значні коливання маси розкату (злитка) очікувано визначають їхній превалюючий вплив на довжину розкату. Вплив маси злитка на довжину прокату описується рівнянням регресії (1) та коефіцієнтом кореляції 0,99 (табл. 2), з якого випливає тісний зв'язок між довжиною розкату та масою розкату (злитка). На рис.4 наведена залежність довжини розкату від маси розкату.

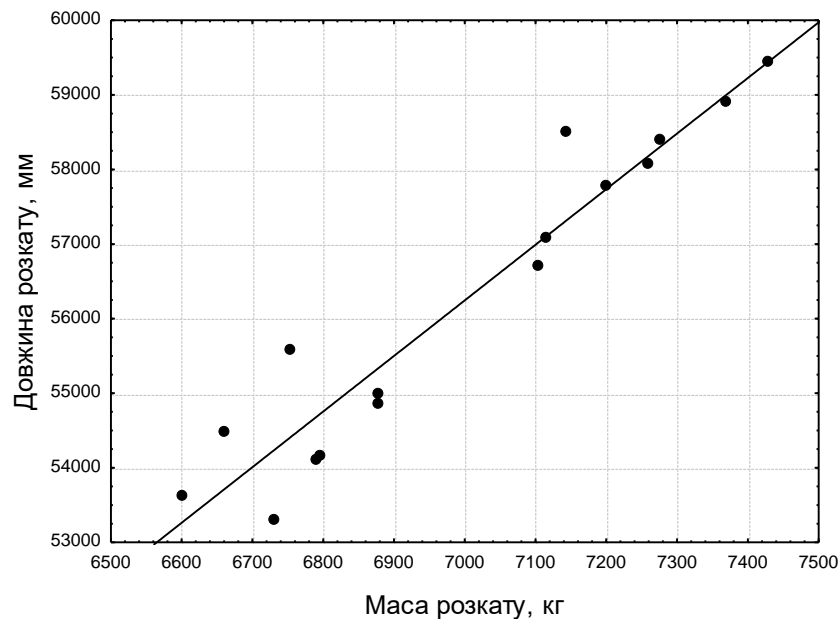


Рисунок 4 - Залежність довжини розкату від маси розкату

Суттєві коливання маси злитка визначили порядок обробки отриманих експериментальних даних. З отриманого масива даних обирали розкати з близькою масою розкатів (злитків), які знаходяться в заданому діапазоні коливань, що відбуваються за рахунок погрішності наливу в вилівниці на поддоні, розгаром вилівниці, і який складає за результатами експериментальних даних ± 210 кг. Для цих розкатів визначали вплив інших параметрів.

Змінення міжвалкового зазору в чистовий кліті виконували для визначення впливу його змінення на довжину розкатів в межах допуску на розміри поперечного перерізу. На рис.5 наведені залежності розмірів розкатів від змінення міжвалкового зазору, які свідчать, що збільшення міжвалкового зазору призводять до збільшення товщини розкату і зменшенню довжини розкатів. Ширина розкатів при цьому суттєво не змінюється і не впливає на довжину розкату. Рівняння регресії і коефіцієнти кореляції наведені в таблиці 2. Мала кількість експериментальних даних компенсується збільшеною величиною змінення міжвалкового зазору, що підтверджує достовірність отриманих даних.

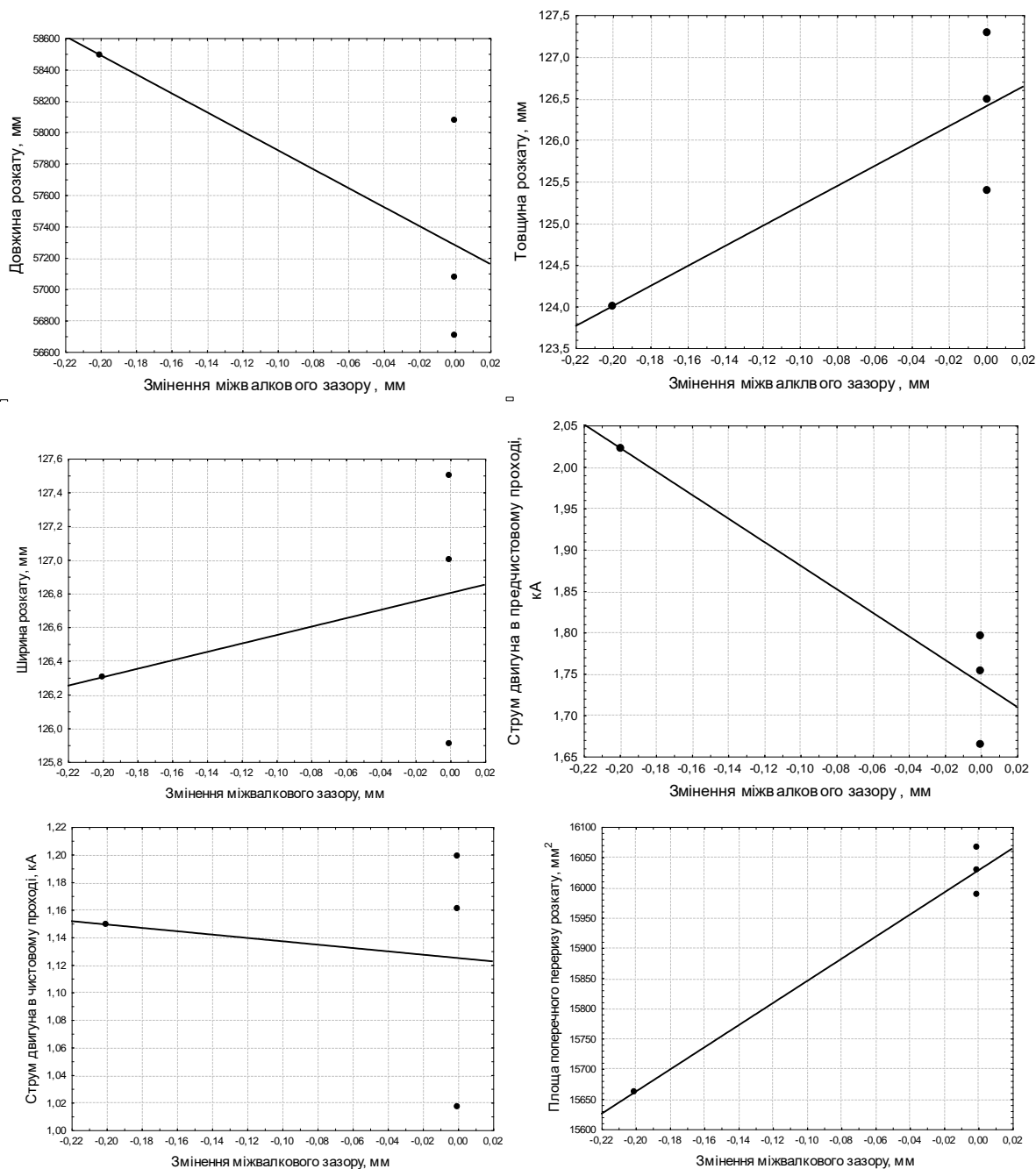


Рисунок 5 - Залежності розмірів розкату, струму двигуна від змінення міжвалкового зазора в чистовий кліті

Теоретичні значення впливу товщини прокату на довжину практично збігаються з експериментальними. Здійснений за даними таблиці 1 розрахунок зміни довжини розкату при зміні товщини прокату на 1 мм та незмінній ширині прокату показав, що зміна довжини розкату становить 450 мм, а зміна довжини розкату за експериментальними даними становила - 331-351 мм.

При проведенні експериментальних досліджень впливу часової затримки на розміри розкатів затримку подачі розкатів здійснювали перед прокаткою в чистовий кліті.

Реальна величина затримок визначалася за осцилографами струму. Залежності розмірів розкатів і струму двигуна від затримки часу наведені на рис.6.

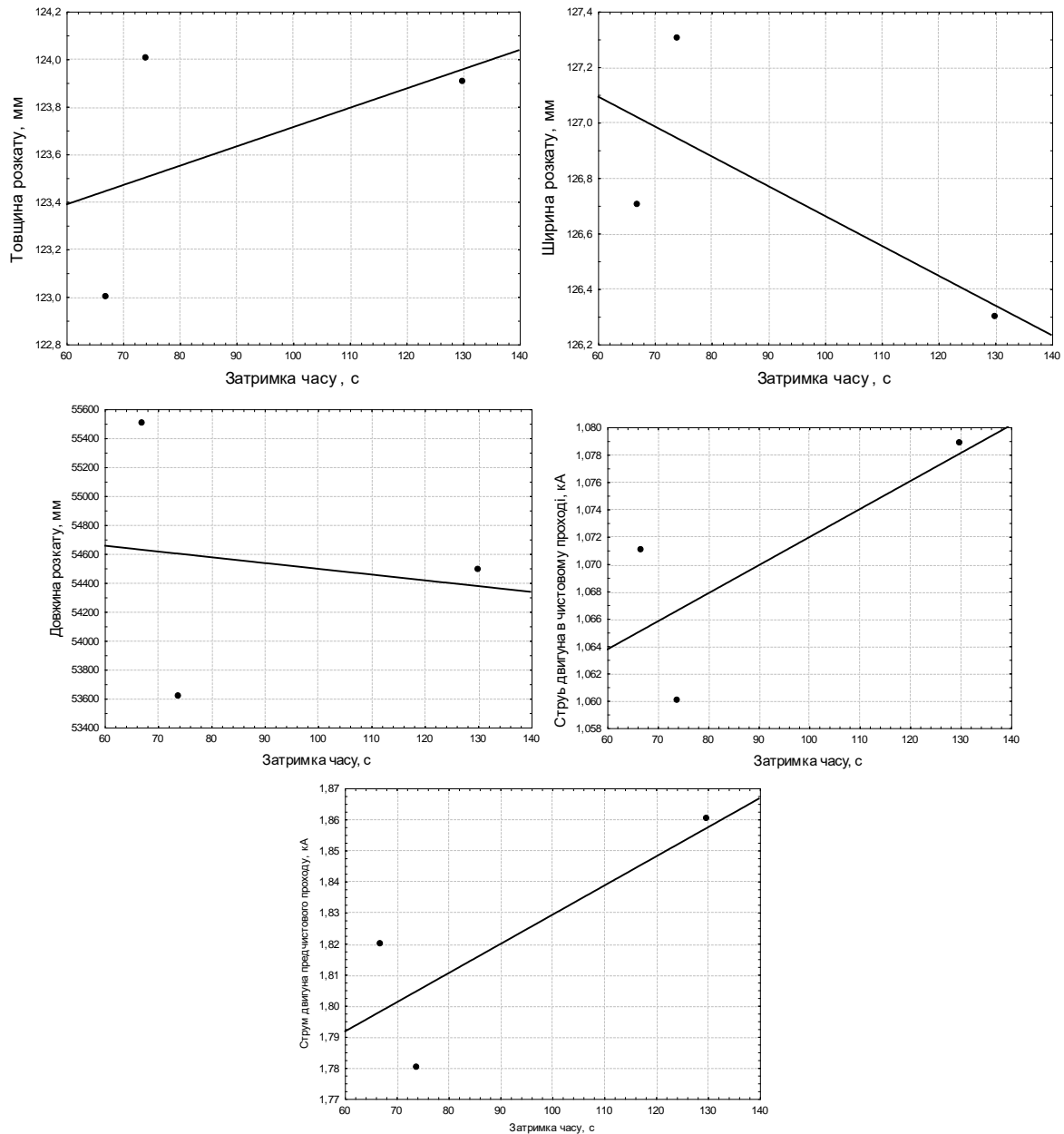


Рисунок 6 - Залежності розмірів розкатів і струму двигуна від затримки часу

Рівняння регресії та коефіцієнти кореляції між та розмірами прокату, величиною затримки та іншими параметрами прокатки наведені у таблиці 2. З них випливає, що затримка не суттєво впливає на розміри прокату. Це пояснюється підвищеним тепловмістом квадратних заготовок з перетином від 100x100 мм та вище, середньомасова температура яких залишається достатньо високою навіть при суттєвому зниженні температури їх поверхні [2].

Результати експерименту в заготівельній кліті, що отримані впродовж прокатки квадратного профілю 150 з заготовок двох плавков, наведені у таблиці 3.

Таблиця 2

Рівняння регресії для прокатки квадрата 125 в чистовий кліті стана 800

Рівняння регресії	Коефіцієнт кореляції
Вплив маси розкату M на довжину розкату L	
$L = 4033 + 7,46M$	0,96
Вплив змінення міжвалкового зазору dS на розміри прокату та струм прокатки в передчистовому I_{np} та чистовому $I_{ч}$ проходах	
$H = 126,63 + 1,32dS$	0,89
$B = 126,67 + 0,13dS$	0,16
$L = 57480 - 505dS$	-0,66
$I_{np} = 1,77 - 0,124dS$	-0,98
$I_{ч} = 1,124 - 0,0128dS$	-0,16
Вплив затримки часу T (с) перед подачею до чистової кліті на розміри прокату та струм прокатки в передчистовому I_{np} та чистовому $I_{ч}$ проходах	
$H = 122,9 + 0,081T$	0,51
$B = 127,73 - 0,011T$	-0,74
$L = 54893 - 3,98T$	-0,14
$I_{np} = 1,73 + 0,0009T$	0,81
$I_{ч} = 1,051 - 0,0002T$	0,74

Таблиця 3

Результати статистичної обробки експериментальних даних для прокатки квадрата 150 у заготівельній кліті

Параметри	Кількість вимірювань	Середнє значення	Мінімальне значення	Максимальне значення	Діапазон змінення	Дисперсія	Середнє квадр. відхилення
Товщина, мм	16	149,19	148,50	149,6	1,100	0,078	0,279
Ширина, мм	16	149,89	149,00	150,5	1,500	0,14	0,373
Площа поперечного перерізу, мм ²	16	22362	22260	22454	194,75	3193	56,51
Довжина розкату, м	23	40,23	38,40	41,50	3,10	5,43	0,73
Струм двигуна в першому проході, кА	23	4,66	2,59	5,67	3,08	0,679	0,7
Струм двигуна в другому проході, кА	23	3,52	2,05	4,21	2,16	0,167	0,2
Струм двигуна в третьому проході, кА	23	1,55	0,83	3,97	3,14	0,620	0,6
Маса розкату, кг	16	6831	6505	7012	507	21267	145,83

Висновки. Проведені на стані 800 експериментальні дослідження дозволили отримати реальні значення діапазону коливань розмірів поперечного перерізу квадрата на правій стороні та в заготівельній кліті. Коливання довжини розкату значною мірою обумовлені коливаннями маси злитка.

Зміна міжвалкового зазору може бути використана для корекції довжини готового розкату в обмеженому діапазоні. Проте ця корекція не зможе повністю вирішити проблему стабілізації довжини розкату з метою підвищення виходу мірної продукції. Необхідно вживати додаткові заходи для стабілізації маси злитка в межах точності розливу на одному піддоні.

Часові затримки в лінії стана, що виникають під час прокатки квадратних профілів 125 та 150 і призводять до нестабільності температурного режиму прокатки та зміни енергосилових параметрів прокатки, не мають суттєвого впливу на розміри поперечного перерізу і довжину розкатів.

ЛІТЕРАТУРА

1. Зінченко М.Д. Експериментальне дослідження процесу прокатки на блюмінгу та стані 800 метзаводу ім.Петровського. Металургійна та гірничорудна промисловість. 2000. №4. С. 40-45.
2. Експериментальне дослідження впливу температурного режиму на розміри фасонних профілів прокату / А.Н. Чернишов, Н.І. Біда, Г.С. Щербина та ін. Металургійна та гірничорудна промисловість. 1977. №4. С.16-19.

REFERENCE

1. Zinchenko M.D. An experimental study of the blooming and 800 stand rolling process metal plant named after Petrovsky. Metallurgical and mining industry. 2000. No. 4. P. 40-45.
2. An experimental study of the effect of the temperature regime on the dimensions of the shaped rental profiles. A.N. Chernyshev, N.I. Bida, G.S. Shcherbina and others. - Metallurgical and mining industry. 1977. No. 4. P.16-19.

Received 01.09.2024.
Accepted 04.09.2024.

Experimental study of rolling accuracy on stand 800

The study of fluctuations in the dimensions of rolled products was carried out for rolled profiles, which are a finished product, because fluctuations in the dimensions of the cross section and the length of the roll lead to the appearance of rods of excessive length when cutting on hot saws. Cutting the rolls does not always allow obtaining a mass of the blank, from which only rods of a measured length will be obtained in the next division. Therefore, increasing the accuracy of mass of square or round blanks allows to reduce fluctuations in the roll length and the appearance of rods of off-gage length and prevent the appearance of short blanks.

The purpose of the work is an experimental study of the process of rolling on the state of 800 square profiles to determine the influence of rolling parameters on the dimensions of the cross-section and the length of the roll. During the study, fluctuations in ingot mass, rolling dimensions, and rolling length were determined, regression equations and correlation coefficients were obtained for the dependence of rolling dimensions and length on ingot mass,

changes in interroll gap and time delay before the finishing stand. It was established that the mass of the ingot fluctuates within significant limits - 638 kg, which is caused by fluctuations in the mass of the melt. Fluctuations in the mass of the ingot within one molding bed are caused by the dimensions of the ingot molds, the pouring height and is within 154 kg. Fluctuations in the dimensions of the roll are carried out within the following limits: thickness - 1.0 mm, width - 0.5 mm, length - 5.3 m. Changing the inter-roll gap can be used to correct the length of the finished roll in a small range of lengths, if the mass fluctuations occur within accuracy of filling on one molding bed. It is necessary to take additional measures to stabilize the weight of the ingot. The influence of delays in the state line, which occur during rolling and lead to instability of the rolling temperature regime, for this type of profiles does not significantly affect the dimensions of the cross section and the length of the roll.

Зінченко Михайло Дмитрович - доцент кафедри автоматизації виробничих процесів, Український державний університет науки і технологій.

Потап Олег Юхимович - професор кафедри автоматизації виробничих процесів, Український державний університет науки і технологій.

Бурчак Андрій Анатолійович - старший викладач кафедри автоматизації виробничих процесів, Український державний університет науки і технологій.

Zinchenko Mykhailo Dmytrovych - Associate Professor of the Department of Automation of Production Processes, Ukrainian State University of Science and Technology.

Potap Oleh Yukhymovich - Professor of the Department of Automation of Production Processes, Ukrainian State University of Science and Technology.

Burchak Andrii Anatoliyovych - Senior Lecturer of the Department of Automation of Production Processes, Ukrainian State University of Science and Technology.

П.В. Герасимчук, Т.А. Ліхоузова, Ю.О. Олійник

ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ АНАЛІЗУ ПОВІДОМЛЕНЬ В TWITTER

Анотація. В сучасному світі соціальні мережі стали невід'ємною частиною життя багатьох людей. Вони є зручним та доступним засобом повсякденного спілкування. Однак з розвитком соціальних мереж з'явилися й проблеми, однією з яких є кібербулінг. В роботі виконано попередню обробку текстових повідомлень, розглянуто різні методи бінарної класифікації текстів та застосовано їх для визначення кібербулінгу в повідомленнях Twitter. Проведено дослідження ефективності кожного з методів та виконано їх порівняльний аналіз. Програмне забезпечення реалізоване мовою Python із застосуванням бібліотек spaCy, gensim, scikit-learn, pandas, numpy, matplotlib.

Ключові слова: класифікація текстів, обробка природної мови, соціальні мережі, кібербулінг, бінарна класифікація, оцінка моделі, word2vec, Bag-of-Words, TF-IDF.

Булінг в інтернеті, зокрема в соціальних мережах, є серйозною загрозою для ментального здоров'я користувачів. Агресивні, образливі або принизливі повідомлення можуть викликати стрес, тривогу, депресію або інші психічні розлади. Через це виявлення та запобігання кібербулінгу є пріоритетним завданням для організацій, що розробляють комунікаційні платформи.

Метою дослідження є пошук моделей для класифікації повідомлень як таких, що містять ознаки булінгу або ні (бінарна класифікація).

Матеріали та методи. Автоматичне визначення булінгу у твітах за допомогою методів машинного навчання є важливим кроком у боротьбі з цією проблемою. Ефективні методи класифікації допоможуть швидко ідентифікувати образливі повідомлення й оперативно вжити відповідні заходи для їх усунення. Це дозволить захистити користувачів від негативного впливу й створити безпечне та комфортне середовище для спілкування [1, 2].

Для того, щоб ефективно протидіяти булінгу в мережі, необхідно розробити моделі для бінарної класифікації текстових повідомлень. Вхідними даними для цієї задачі є самі повідомлення, вихідними - позитивний чи негативний висновок щодо наявності ознак кібербулінгу в конкретному повідомленні.

Для роботи було обрано датасет [3]. Кількість унікальних записів у датасеті – 46017, 39747 твітів з ознаками булінгу і 7945 нейтральних повідомлень. Програмне забезпечення реалізоване мовою Python [4] із застосуванням бібліотек spaCy [5], gensim [6], scikit-learn [7], pandas, numpy, matplotlib.

Першим етапом попередньої обробки даних є очищення текстів від html-тегів, гіперпосилань, імен користувачів, хештегів, небуквених символів та приведення літер до нижнього регістру. Для цього було використано регулярні вирази. Після цього було проведено токенізацію, лематизацію та очищення від стоп слів за допомогою засобів бібліотеки spaCy.

Для застосування класифікаційних моделей було утворено 3 набори вхідних даних: матрицю Bag-of-Words, TF-IDF матрицю та матрицю word2vec. Кожен із наборів було поділено на навчальний та тестовий у співвідношенні 80% та 20% відповідно. Матриця Bag-of-Words містить кількість входжень кожного слова у кожен із документів, TF-IDF матриця – TF-IDF метрики слів у кожному документі, а матриця word2vec – векторне представлення кожного із документів.

Для класифікації текстових повідомлень було використано наступні методи:

- логістична регресія (LogisticRegression бібліотеки scikit-learn);
- метод k найближчих сусідів (KNearestNeighbors бібліотеки scikit-learn);
- випадковий ліс (RandomForestClassifier бібліотеки scikit-learn);
- метод опорних векторів (SVC бібліотеки scikit-learn);
- наївний баєсівський класифікатор (MultinomialNB бібліотеки scikit-learn).

Результати та основний матеріал дослідження. Для оцінки ефективності класифікації текстів моделями було використано показники влучності (precision), повноти (recall), f1-метрики (f1-score) та точності (accuracy), а також матриці невідповідності.

Логістична регресія (рисунки 1-3).

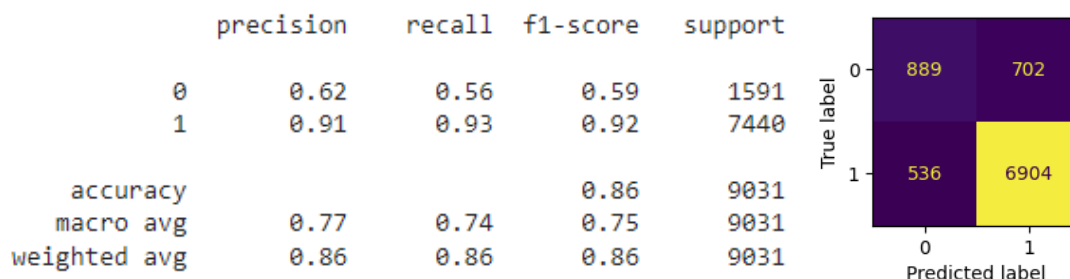


Рисунок 1 – LogisticRegression, вхідні дані – матриця Bag-of-words

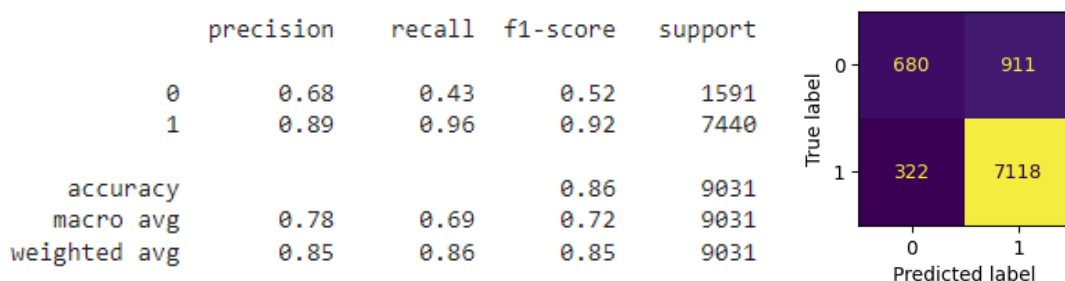


Рисунок 2 – LogisticRegression, вхідні дані – матриця TF-IDF

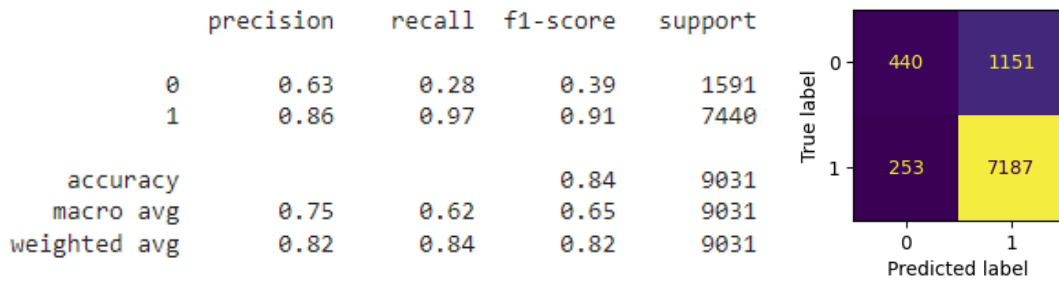


Рисунок 3 – LogisticRegression, вхідні дані – матриця word2vec

Метод *k* найближчих сусідів (рисунки 4-6). Найвища точність досягнута при *k* = 55.

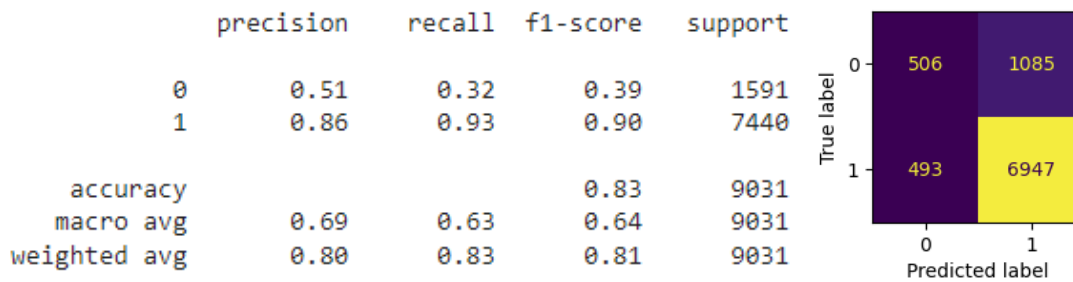


Рисунок 4 – KNearestNeighbors, вхідні дані – матриця Bag-of-words

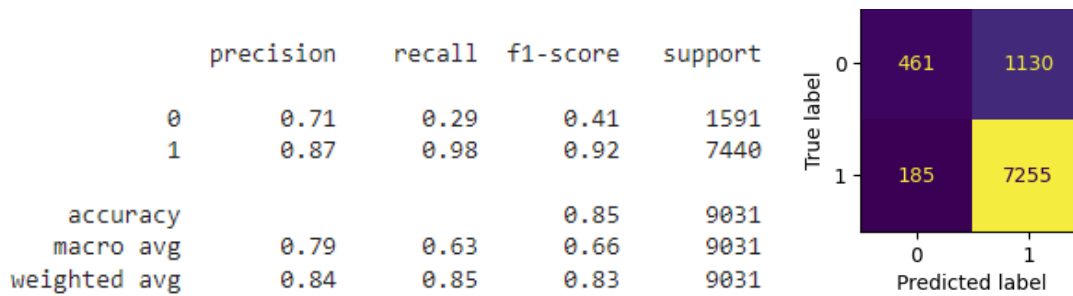


Рисунок 5 – KNearestNeighbors, вхідні дані – матриця TF-IDF

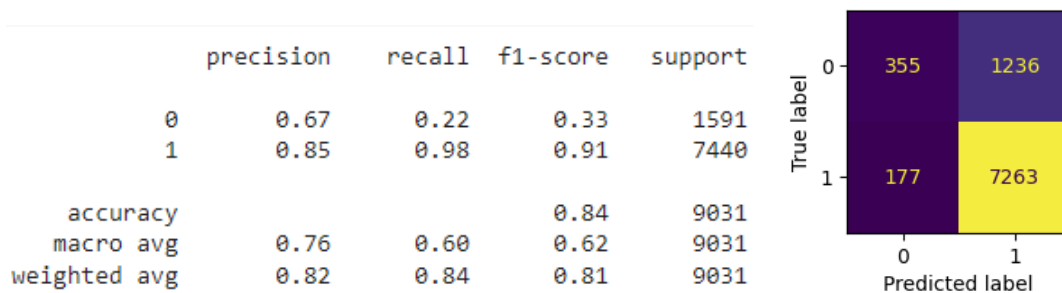


Рисунок 6 – KNearestNeighbors, вхідні дані – матриця word2vec

Метод випадкового лісу (рисунок 7-9).

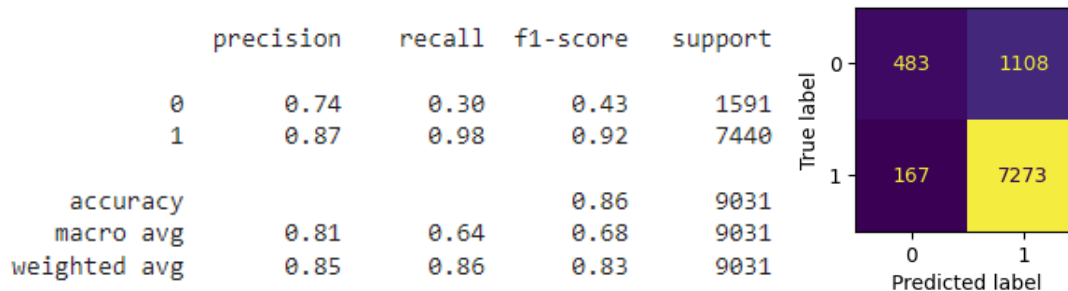


Рисунок 7 – RandomForestClassifier, вхідні дані – матриця Bag-of-words

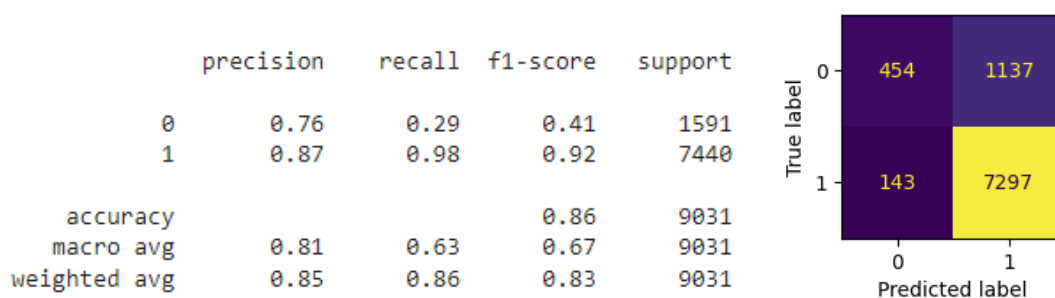


Рисунок 8 – RandomForestClassifier, вхідні дані – матриця TF-IDF

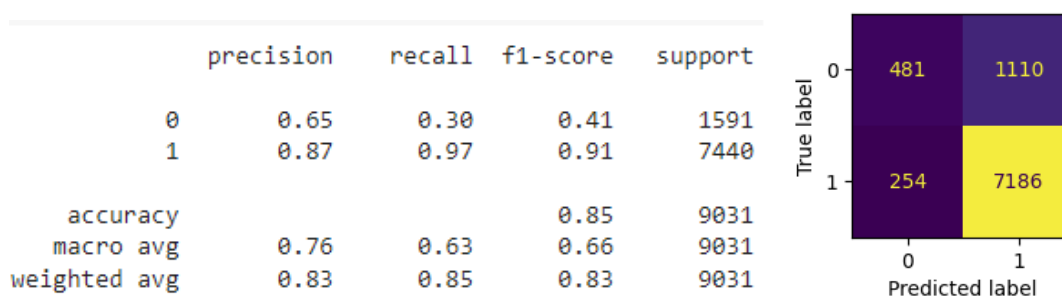


Рисунок 9 – RandomForestClassifier, вхідні дані – матриця word2vec

Метод опорних векторів (рисунок 10-12). Для класифікації було обрано модель LinearSVC через велику розмірність вхідних наборів даних.

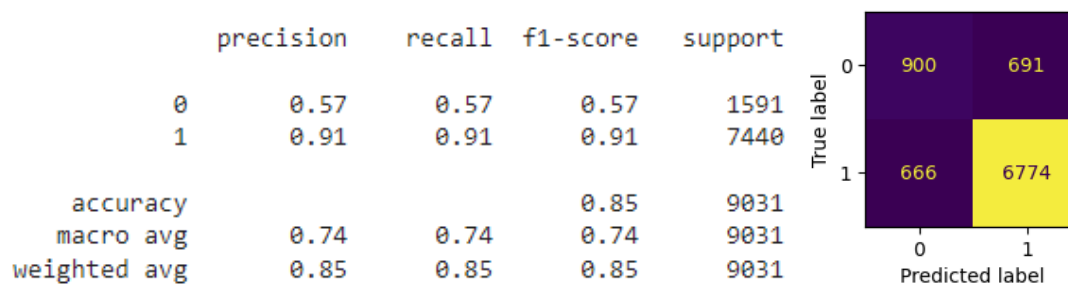


Рисунок 10 – LinearSVC, вхідні дані – матриця Bag-of-words

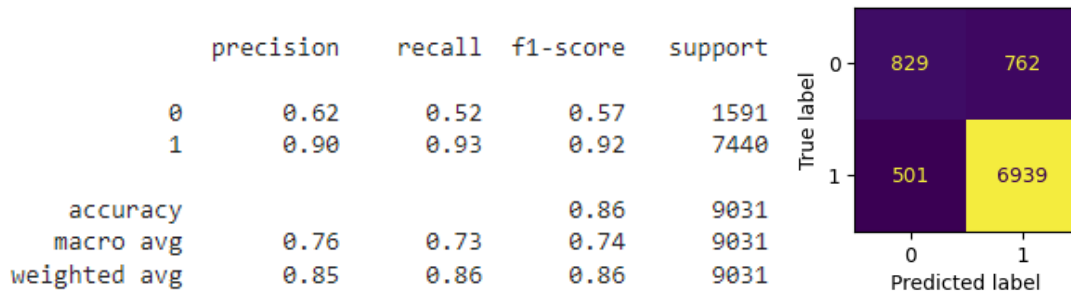


Рисунок 11 – LinearSVC, вхідні дані – матриця TF-IDF

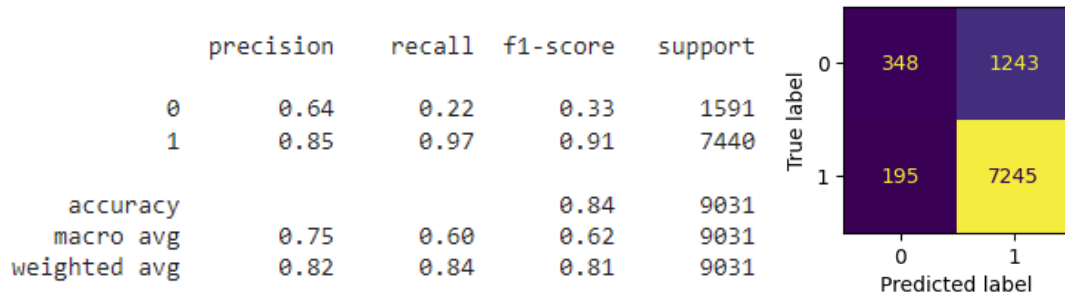


Рисунок 12 – LinearSVC, вхідні дані – матриця word2vec

Наївний байєсівський класифікатор (рисунки 13-15). Для класифікації було обрано модель MultinomialNB через поліноміальну природу розподілу частот слів у текстах.

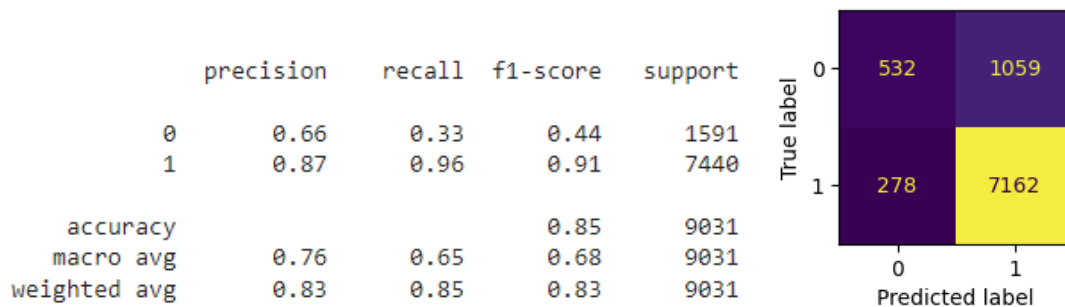


Рисунок 13 – MultinomialNB, вхідні дані – матриця Bag-of-words

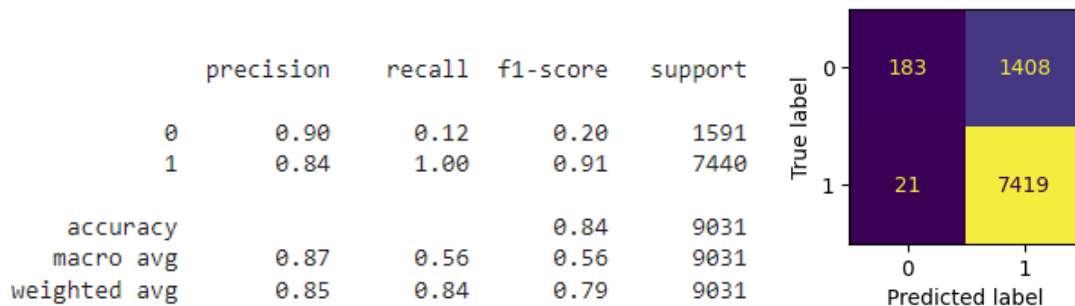


Рисунок 14 – MultinomialNB, вхідні дані – матриця TF-IDF

Порівняння результатів моделей.

За результатами тестування моделей можна зробити такі висновки:

- ефективність класифікації для усіх моделей мало залежить від вибору типу вхідних даних (Bag-of-Words, TF-IDF, word2vec);

- логістична регресія та метод опорних векторів показали трохи кращу ефективність, ніж інші моделі;
- найкращий баєсівський класифікатор показав відносно низьку ефективність при тренуванні на TF-IDF матриці;
- метод k найближчих сусідів показав відносно низьку ефективність при тренуванні на матриці Bag-of-Words.

Враховуючи те, що найкращу ефективність показали моделі на основі логістичної регресії та методу опорних векторів та швидкодію логістичної регресії, найдоцільнішим для розв'язання поставленого завдання є використання саме класифікатора LogisticRegression на матриці Bag-of-Words.

Висновки. За результатами порівняльного аналізу ефективності моделей виокремлено логістичну регресію на вхідних даних Bag-of-Words як найефективнішу модель для задачі бінарної класифікації текстових повідомлень із обраного набору.

Отримані в ході дослідження результати можуть бути використані для розробки систем автоматичного виявлення ознак кібербулінгу в повідомленнях користувачів соціальних мереж та оперативного вживання відповідних заходів.

ЛІТЕРАТУРА / REFERENCES

1. Писаренко О. А. Інтелектуальна система фільтрації коментарів з використанням машинного навчання. – 2019.
2. Іванов О. А. Розробка сервісу для боротьби з кібербулінгом // Автоматизація та комп'ютерно-інтегровані технології у виробництві та освіті: стан, досягнення, перспективи розвитку. – С. 298.
3. <https://www.kaggle.com/datasets/andrewmvd/cyberbullying-classification/>
4. Документація мови програмування Python. [Електронний ресурс] – Режим доступу до ресурсу: <https://docs.python.org/3/>
5. Документація бібліотеки spaCy. [Електронний ресурс] – Режим доступу до ресурсу: <https://spacy.io/usage>
6. Документація бібліотеки Gensim. [Електронний ресурс] – Режим доступу до ресурсу: <https://radimrehurek.com/gensim/>
Документація бібліотеки scikit-learn. [Електронний ресурс] – Режим доступу до ресурсу: <https://scikit-learn.org/stable/>

Received 04.09.2024.
Accepted 09.09.2024.

Application of machine learning methods for analysis of Twitter messages

The paper investigates the problem of binary classification of text messages for the presence of bullying. Bullying on the Internet, in particular in social networks, is a serious threat to the mental health of users. Aggressive, offensive or humiliating messages can cause stress, anxiety, depression or other mental disorders. Because of this, identifying and preventing cyberbullying is a priority for organizations developing communication platforms.

A dataset with Twitter messages was prepared and pre-processed, including cleaning, tokenization, and lemmatization. 3 sets of input data for classification models were created: Bag-of-Words, TF-IDF matrix, word2vec matrix.

Models based on various machine learning methods were built and tested: logistic regression, k nearest neighbors, random forest, support vector, naive Bayesian classifier methods on each of the input data sets.

Based on the results of testing the models, a comparative analysis of their effectiveness was carried out, logistic regression on Bag-of-Words input data was singled out as the most effective model for the task of binary classification of text messages from the selected set.

The results obtained in the course of the study can be used for the development of systems for automatic detection of signs of cyberbullying in the messages of users of social networks and the prompt use of appropriate measures.

Keywords: text classification, natural language processing, social networks, cyberbullying, binary classification, model evaluation, word2vec, Bag-of-Words, TF-IDF.

Герасимчук Павло Вікторович – студент кафедри інформатики та програмної інженерії, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

Ліхоузова Тетяна Анатоліївна – к.т.н., доцент кафедри інформатики та програмної інженерії, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

Олійник Юрій Олександрович – к.т.н., доцент кафедри інформатики та програмної інженерії, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

Herasymchuk Pavlo – student, Department of Informatics and Software Engineering National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute».

Likhouzova Tetiana – PhD, associate professor, Department of Informatics and Software Engineering National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute».

Oliinyk Yurii – PhD, associate professor, Department of Informatics and Software Engineering National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute».

АНАЛІЗ ПІДХОДІВ ДО РЕАЛІЗАЦІЇ СИСТЕМ ЕЛЕКТРОННОГО ГОЛОСУВАННЯ

Анотація. У роботі розглядаються сучасні підходи для побудови систем електронного голосування, такі як блокчейн, а також традиційні криптографічні методи, що включають гомоморфне шифрування, сліпий підпис та докази нульового розголошення. Метою роботи є вибір підходу до побудови систем електронного голосування на основі порівняльного аналізу їх основних характеристик. Вирішені задачі: огляд вимог, узагальненої структури та основних процедур систем електронного голосування; аналіз існуючих видів систем електронного голосування та їх порівняльна характеристика. Виділено список вимог до систем електронного голосування, представлено узагальнені структури основних їх видів. Проведено порівняльний аналіз видів систем електронного голосування на основі відповідності вимогам. Зроблено вибір підходу для подальшої побудови системи.

Ключові слова: електронне голосування, гомоморфне шифрування, блокчейн, докази з нульовим розголошенням, сліпий підпис.

Постановка проблеми. При проектуванні систем електронного голосування важливим аспектом постає формулювання вимог до програмної частини. Рада Європи [1] випустила директиву, яка описує формальні вимоги до таких систем. Директиви описують юридичні аспекти та залишають питання реалізації і криптобезпеки не вирішеними. З розвитком цифрових технологій з'являється значний потенціал для вдосконалення процесів голосування через впровадження електронних систем. З впровадженням електронних систем приходять і нові виклики, особливо у плані забезпечення безпеки даних та недопущення маніпуляцій з результатами виборів. Криптографія відіграє ключову роль у створенні безпечних електронних систем голосування, дозволяючи забезпечити конфіденційність голосів і верифікацію результатів без розкриття особистості виборців.

Одним з передових напрямків [2, 3] у розробці безпечних систем електронного голосування є використання технології блокчейн. Ця технологія дозволяє створити децентралізовану і незмінну базу даних голосів, яка може захистити процес від фальсифікацій та несанкціонованого доступу. Водночас, існують і інші технології та підходи, здатні підвищити ефективність та безпеку електронного голосування, включаючи різноманітні криптографічні протоколи які можуть забезпечити анонімність виборців, верифікацію результатів та неможливість подвійного голосування.

Аналіз останніх досліджень і публікацій. Гомоморфне шифрування використовується у системах електронного голосування для проведення операції підрахунку голосів без розкриття ідентифікатора голосуючого [4, 5]. Сучасні технології доказів з нульовим розголошенням дозволяють неінтерактивно провести операції над зашифрованими даними, зберігаючи криптографічний доказ, що вихідні дані не було змінено [6, 7]. Змішані сітки за допомогою спеціальних протоколів маршрутизації дозволяють сховати ідентифікатор голосуючого [8]. Блокчейн системи використовуються для забезпечення децентралізованого сховища і інфраструктурної організації систем електронного голосування [2, 9].

Системи електронного голосування активно досліджуються, як з боку використання сучасної криптографії [10, 11], так і з боку організації віддаленого доступу та забезпечення державних вимог [12].

Мета дослідження. Метою роботи є вибір підходу до побудови систем електронного голосування на основі порівняльного аналізу їх основних характеристик. Результати аналізу можуть бути використані при реалізації відповідних систем. В роботі відповідно до мети поставлені такі задачі: огляд вимог, узагальненої структури та основних процедур систем електронного голосування; аналіз існуючих видів систем електронного голосування та їх порівняльна характеристика. У роботі не розглядаються конкретні реалізації систем, а тільки підходи до їх реалізацій.

Основний матеріал дослідження. У системах електронного голосування є кілька ключових діючих сторін, кожна з яких має свої функції та права. Основні діючі сторони включають [12]:

- **Виборці** – це особи, які мають право голосу і беруть участь у виборах.
- **Кандидати** – це особи та організації, які беруть участь у виборах і змагаються за голоси виборців.
- **Адміністратори (або посадові особи)** – це органи та особи, відповідальні за організацію і проведення виборів
- **Спостерігачі (Аудитор)** – це особи, які контролюють виборчий процес, забезпечуючи його прозорість та чесність.
- **Ресстратор** – відповідальний за автентифікацію виборців. Вони видають виборцям дані для доступу до системи, такі як приватні та публічні ключі.

Вимоги до системи. Аналіз існуючих підходів до побудови систем доцільно почати з формування списку формальних вимог до систем, за якими можна якісно або кількісно порівняти існуючі системи.

Як пише Ліав Х. [14], в існуючих дослідженнях часто використовують різні терміни для одних і тих самих сутностей. Не всі системи електронного голосування задовольняють однаковий набір властивостей безпеки, в силу їх протиріччя. Наприклад питання перевіряємості (verifiability) та ревізійності (auditability) завжди протиставляється питанню приватності та анонімності. Базуючись на дослідженнях Лі Б. [14] є доцільним вимагати наступні властивості безпеки:

- **Анонімність голосування** - система повинна гарантувати, що ніхто не може простежити, за кого проголосував конкретний виборець.

- **Універсальна перевіряємість** - будь-який зацікавлений учасник процесу (виборець, спостерігач, незалежна організація) може переконатися в правильності підрахунку голосів. Перевіряємість також може бути індивідуальною, коли система дозволяє перевірити кожному голосуючому що його голос був врахований.

- **Чесність** (fairness) - часткові результати не повинні бути опубліковані.

- **Надійність** (robustness) - стійкість системи до зловмисної поведінки учасників процесу.

- **Відсутність квитанцій** (receipt-freeness) - система не повинна надавати виборцю документ який би підтверджував за кого він голосував.

Структура системи електронного голосування. Алгоритм роботи систем електронного голосування можна розбити на три основні фази: початкове налаштування (див. рис. 1), збір голосів (див. рис. 2) та підрахунок і оприлюднення голосів (див. рис. 3).

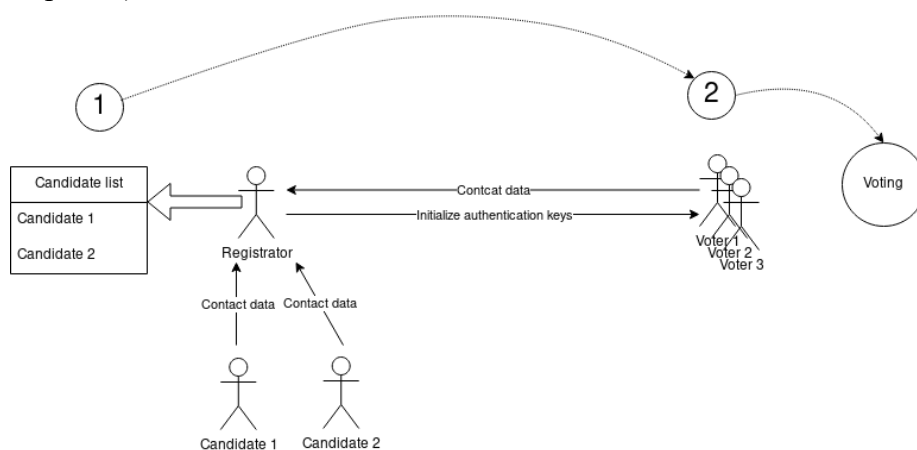


Рисунок 1 - Узагальнена схема процедури ініціалізації системи електронного голосування

У ході початкового налаштування виставляються параметри системи, такі як максимальна кількість кандидатів, строки виборів, а також проходять наступні процеси:

1. **Реєстрація кандидатів (Candidate Registration)** – подача заявок кандидатами для участі у виборах, реєстратор перевіряє дані кандидатів і додає до списку кандидатів.

2. **Реєстрація виборців (Voter Registration)** – збір і верифікація даних виборців, забезпечення їхньої автентифікації в системі. У ході цього шагу також проходить розповсюдження криптографічних ключів, або інших даних необхідних для автентифікації користувачів.

Далі система може переходити до наступного етапу.

Основні кроки (див. рис. 2):

1. У ході процесу збору голосів, важливо щоб усі голосуючі мали доступ до списку кандидатів та параметрів системи.

2. Далі, голосуючі формують свої голоси та автентифікуються у реєстратора для відправки цього голосу, в деяких схемах цей процес автентифікації голосуючого та

відправки голосу може бути реалізований в рамках одного кроку.

3. Після надсилання голосу, голосуючі отримують підтвердження, що їх голос опрацьований.

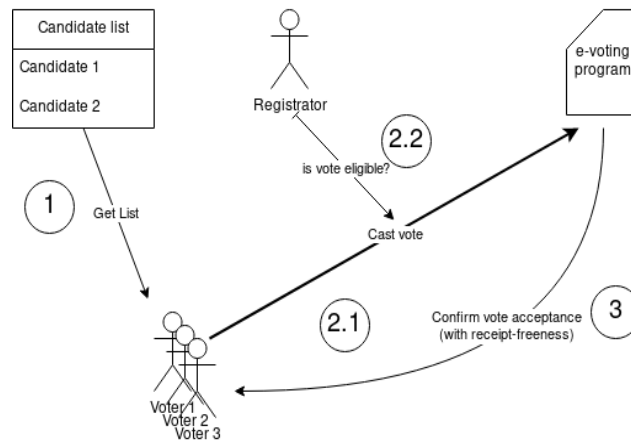


Рисунок 2 - Узагальнена схема процедури збору голосів

На цьому етапі важливо дотримання деяких властивостей:

1. **Анонімність голосу** - сам голос не має містити інформації про ідентифікатор користувача. При цьому очевидним є вимога до системи щодо захисту від подвійного голосування.

2. **Анонімність голосуючих** - система повинна використовувати деякі методи, наприклад змішування [8] щоб відв'язати ідентифікатор користувача від його голосу.

Відсутність квитанцій - при підтвердженні голосу, голосуючі не повинні отримувати так звані "квитанції" - підтвердження за кого саме із кандидатів вони голосували.

У ході процедури фіналізації (див. рис. 3) зашифровані голоси дешифруються сервером, і кожен голосуючий може перевірити що їх голос було враховано.

Програма також формує список результат з фінальною кількістю голосів за кожного кандидата. Спостерігачі можуть перевірити що голоси були чесно підраховані.

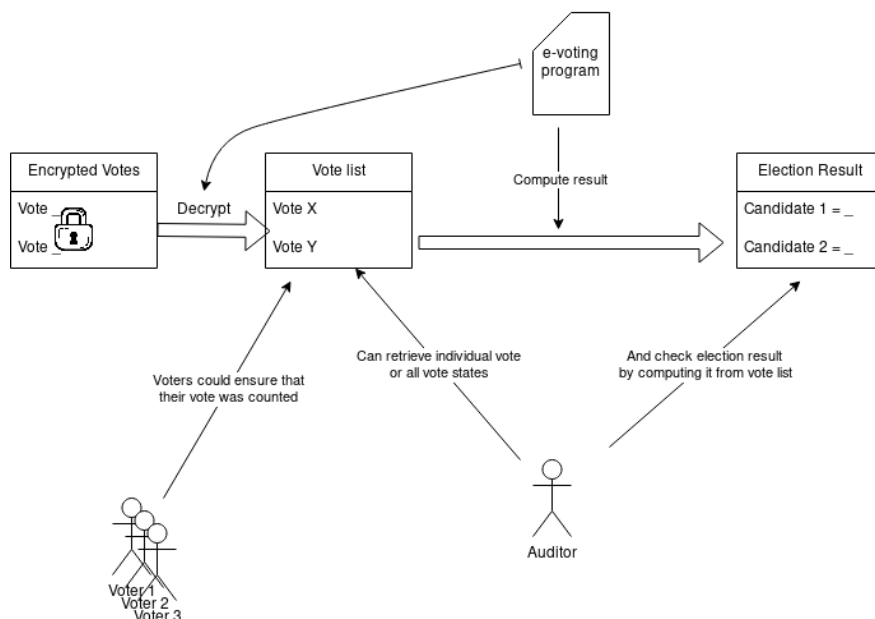


Рисунок 3 - Узагальнена схема підрахунку і оприлюднення голосів

Основні види систем електронного голосування. Гомоморфне шифрування дозволяє учасникам шифрувати свої голоси таким чином, що сервер голосування зможе підрахувати загальний результат без необхідності розшифровувати окремі голоси.

Тобто, для заданих $E(v_1)$, $E(v_2)$, можна підрахувати $E(v_1 \otimes v_2)$ без дешифровки v_1 та v_2 .

У схемі (див. рис. 4) голосуючими є Alice та Bob - вони відправляють свої зашифровані голоси. Система може підрахувати голоси та видати результат. Системи на основі гомоморфного шифрування потребують додатково схеми доказу. Це може бути або доказ нульового розголошення, або довіри до адміністратора під час збору даних, або додаткових схем часткового розшифрування [3].

Система на базі гомоморфного шифрування має наступні властивості:

- **Анонімність голосування** - усі голоси зашифровані на стороні відправника, а отже залишається тільки питання доставки голосу.
- **Універсальна перевіряємість** - після підрахунку голосів, усі голоси доступні у зашифрованому вигляді, а отже усі небайдужі можуть повторити підрахунок самотужки.
- **Відсутність квитанцій** - залежить від схем доказу.
- **Надійність** - стійкість системи до зловмисної поведінки залежить від схем доказу унікальності голосу.
- **Чесність** - можливе тільки при довірі до серверу електронного голосування, лежить за межами питання підходу гомоморфного шифрування.

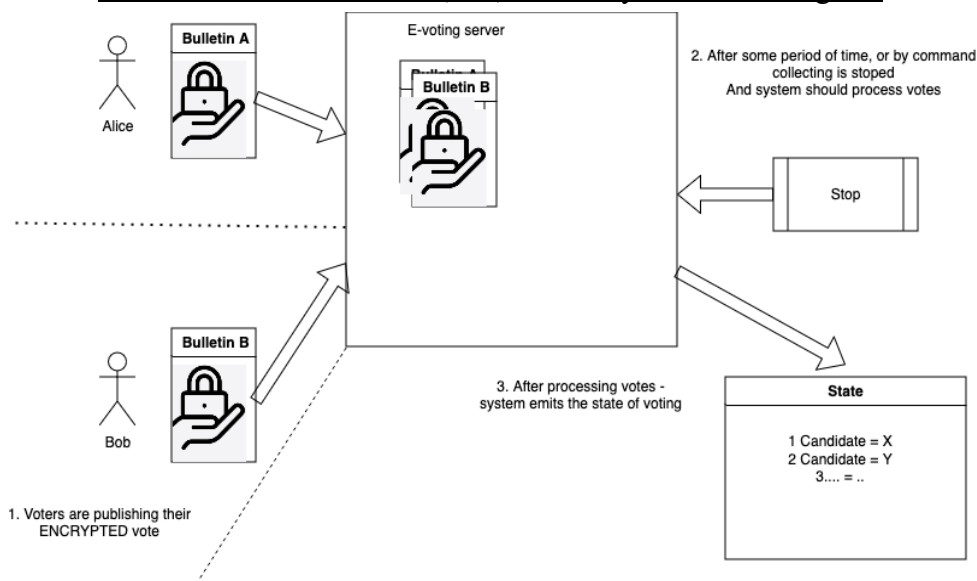


Рисунок 4 – Спрощена структура взаємодії з сервером у системі на базі гомоморфного шифрування

Сліпий підпис дозволяє учасникам довести своє право на голосування без розкриття своєї ідентичності, що допомагає забезпечити анонімність та запобігти подвійному голосуванню (див. рис.5).

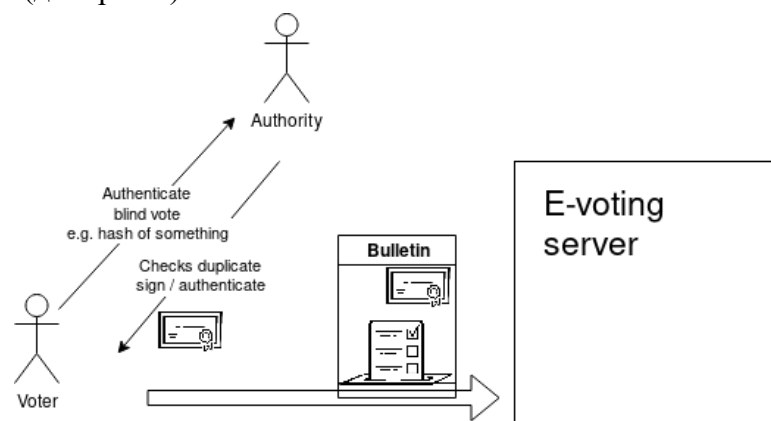


Рисунок 5 – Спрощена структура взаємодії з сервером у системі на базі сліпого підпису

Система на базі сліпого підпису має наступні властивості:

- **Анонімність голосування** - сліпі підписи дозволяють виборцеві засліпити свій бюлетень перед відправкою. Після отримання підпису виборець знімає засліплення, і кінцевий підписаний бюлетень не містить ідентифікатор виборця. Виборчий орган підписує бюлетень без знання його вмісту, тому анонімність виборця зберігається.

- **Універсальна перевіряємість** - сліпі підписи забезпечують, що бюлетені були підписані виборчим органом, але не розкривають інформацію про те, хто подав конкретний бюлетень. Це дозволяє будь-якому учаснику перевірити, що всі бюлетені підписані легітимним органом. Однак, універсальна перевіряємість також вимагає можливості перевірки правильності підрахунку голосів без компрометації анонімності, що потребує додаткових механізмів, таких як криптографічне доказування з нульовим

розголошенням (Zero-Knowledge Proofs).

- **Відсутність квитанцій** - виборець не отримує жодного доказу свого голосу після відправки підписаного бюлетеня до системи. Це унеможлиблює ситуацію, коли виборець може довести третій стороні, за кого він голосував, що запобігає примусу або купівлі голосів.

- **Надійність** - тільки з іншими криптографічними методами.

- **Чесність** - тільки при довірі до адміністратора.

Блокчейн дозволяє побудувати децентралізовану систему, в якій немає довіри до єдиного сервера (див. рис. 6). У цій схемі - Alice та Bob, це учасники, вони публікують свої голоси у блокчейн. Стан системи може в режимі реального часу підраховувати голоси, або робити це за командою адміністратора.

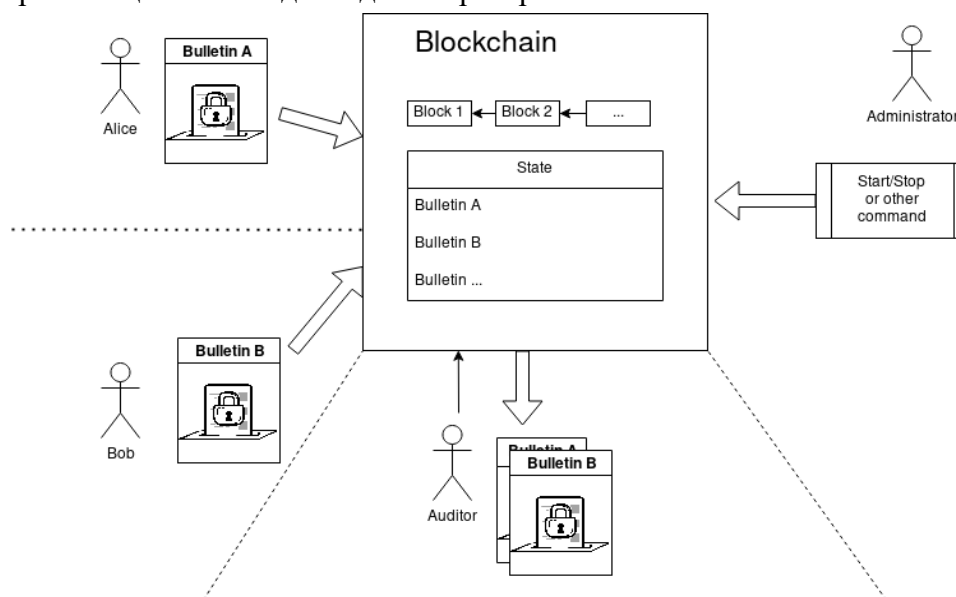


Рисунок 6 – Спрощена структура реалізації системи електронного голосування на базі блокчейну

Система на базі блокчейну має наступні властивості:

- **Анонімність голосування** - може бути забезпечена лише за умови використання псевдонімів або інших методів, що приховують справжню особистість виборців. Без додаткових методів для забезпечення анонімності, таких як шифрування або змішування, блокчейн не може гарантувати анонімність.

- **Універсальна перевіряємість** - кожна транзакція на блокчейні є незмінною і доступною для перевірки. Це дозволяє будь-якому учаснику перевірити, що кожен голос був зарахований правильно і не змінений після подання.

- **Відсутність квитанцій** - для забезпечення відсутності квитанцій потрібні додаткові методи, такі як сліпі підписи або інші криптографічні протоколи.

- **Надійність** - блокчейн забезпечує високу надійність завдяки своїй децентралізованій природі і криптографічному захисту. Всі транзакції є незмінними і захищеними від фальсифікації. Консенсусні механізми, такі як Proof of Work або Proof of Stake, забезпечують цілісність і достовірність даних.

- **Чесність** - блокчейн забезпечує прозорість і неможливість зміни результатів після подання голосу. Проте, без додаткових механізмів автентифікації і контролю, блокчейн не може повністю гарантувати, що кожен голос є легітимним (тобто, поданий дійсним виборцем). Потрібні додаткові заходи для перевірки правомірності виборців і запобігання подвійного голосування.

Докази нульового розголошення використовуються для верифікації того, що голоси відповідають правилам голосування (наприклад, належать до допустимого набору варіантів) без необхідності розкривати голоси (див. рис. 7). У цій схемі Alice та Bob генерують докази для шифрованих голосів. Сервер прийому голосів точно знає що голос ще не був врахований, і належить до належного користувача, а також те що бюлетень не був зіпсований.

Сучасні системи доказів нульового розголошення такі як zk-SNARK та zk-STARK є універсальними та можуть бути повною заміною гомоморфного шифрування [6, 7], тому вони майже повністю копіюють властивості систем на бази гомоморфного шифрування доповнюючи деякі питання автентифікації.

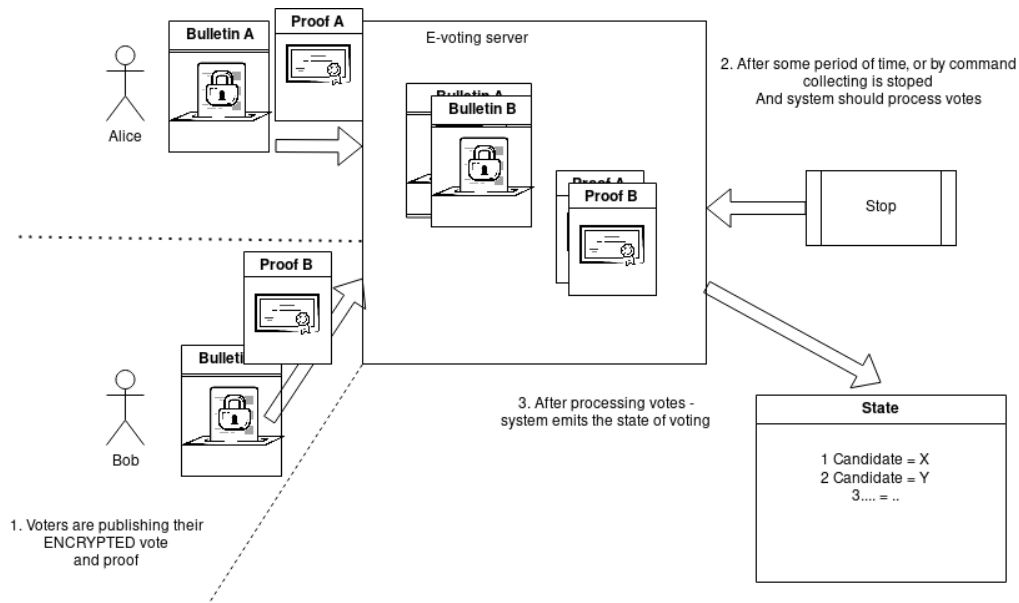


Рисунок 7 – Спрощена структура реалізації системи електронного голосування на базі доказів нульового розголошення

Система на базі доказів нульового розголошення має наступні властивості:

- **Анонімність голосування** - усі голоси зашифровані на стороні відправника, а отже залишається тільки питання доставки голосу.
- **Універсальна перевіряємість** - після підрахунку голосів, усі голоси доступні у зашифрованому вигляді, а отже усі небайдужі можуть повторити підрахунок самотужки.
- **Відсутність квитанцій** - залежно від реалізації, але більшість систем на базі ZKP повністю задовольняють цю вимогу.
- **Надійність** - доказ також автентифікує голосуючого, що можна використовувати для фільтрації небажаних голосів.

• **Чесність** - можливе тільки при довірі до серверу електронного голосування, лежить за межами питання підходу доказів з нульовим розголошенням.

За результатами аналізу відомих видів систем електронного голосування проведено їх порівняльну характеристику (див. табл. 1).

Таблиця 1

Порівняльна характеристика відомих видів систем електронного голосування

	Анонімність	Перевіряємість	Чесність	Надійність	Відсутність квитанцій
Гомоморфне шифрування	+	+	*	-	-
Сліпий підпис	+	+	*	-	+
Блокчейн системи	-	+	-	+	-
Докази нульового розголошення	+	+	*	+	+

Примітка: символами «+» відмічена повна відповідність, «-» невідповідність, «*» часткова відповідність.

За результатами порівняльного аналізу виходить, що системи на базі гомоморфного шифрування та сліпого підпису потребують додаткових механізмів криптографічних доказів, що робить їх використання без доказів нульового розголошення непрактичним. Також, з порівняння видно, що технологія блокчейн може бути використана тільки при використанні інших криптографічних методів. Сучасні докази нульового розголошення забезпечують майже усі властивості. Єдиним питанням у таких систем залишається питання довіри до серверу. Технологія блокчейн вирішує проблему довіри до серверу та робить процес децентралізованим, саме тому доцільним є використання гібридної системи блокчейну та сучасних доказів з нульовим розголошенням.

Висновки. У роботі проаналізовані існуючі підходи до побудови систем електронного голосування. Виділено список вимог до систем електронного голосування, побудовані узагальнені структури основних видів систем. На основі характеристик видів систем проведено їх порівняльний аналіз. Для забезпечення усіх властивостей в кожному з видів систем потрібна реалізація додаткових механізмів. В результаті запропоновано гібридну систему, яка може задовольнити усі необхідні властивості.

ЛІТЕРАТУРА / REFERENCES

1. Recommendation on legal, operational and technical standards for e-voting: Rec of 30.09.2004. URL: [https://www.coe.int/t/dgap/goodgovernance/Activities/Key-Texts/Recommendations/00Rec\(2004\)11_rec_adopted_en.asp](https://www.coe.int/t/dgap/goodgovernance/Activities/Key-Texts/Recommendations/00Rec(2004)11_rec_adopted_en.asp).
2. Hajian M. Berenjestanakiet al, Blockchain-Based e-voting systems: a technology review // Electronics. 2023. Vol. 13, no. 1. P. 17.

3. Y. Zhan et al. Efficient Electronic Voting System Based on Homomorphic Encryption // *Electronics*. 2024. Vol. 13, no. 2. P. 286.
4. A. Abu Aziz A., N. Qunoo H., A. Abu Samra A. Using Homomorphic Cryptographic Solutions on E-voting Systems // *International Journal of Computer Network and Information Security*. 2018. Vol. 10, no. 1. P. 44–59.
5. Shinde S. S., Shukla S., Chitre D. K. Secure E-voting Using Homomorphic Technology // *International Journal of Emerging Technology and Advanced Engineering*. 2013. Vol. 3, no. 8. P. 203–206.
6. Panait A.-E., F. Olimid R. On Using zk-SNARKs and zk-STARKs in Blockchain-Based Identity Management // *Innovative Security Solutions for Information Technology and Communications*. 2020. T. 12596. URL: https://link.springer.com/chapter/10.1007/978-3-030-69255-1_9.
7. Ashur T., Dhooghe S. MARVELlous: a STARK-friendly family of cryptographic primitives // *International Association for Cryptologic Research*. 2018. No. 1098. URL: <https://eprint.iacr.org/2018/1098>.
8. Bibiloni P., Escala A., Morillo P. Vote Validatability in Mix-Net-Based eVoting // *E-Voting and Identity*. 2015. URL: https://link.springer.com/chapter/10.1007/978-3-319-22270-7_6.
9. Jafar U., Ab Aziz M. J., Shukur Z. Blockchain for electronic voting system—review and open research challenges // *Sensors*. 2021. Vol. 21, no. 17.
10. Kho Y.-X., Heng S.-H., Chin J.-J. A Review of Cryptographic Electronic Voting // *Symmetry*. 2022. Vol. 14, no. 5. P. 858.
11. K. M. AboSamra et al. A practical, secure, and auditable e-voting system // *Journal of information security and applications*. 2017. Vol. 36. P. 69–89.
12. Li H., Kankanala A. R., Zou X. A Taxonomy and Comparison of Remote Voting Schemes. *IEEE*. 2014. No. 23.
13. Liaw H.-T. A secure electronic voting protocol for general elections. *Computers & Security*. 2004. Vol. 23, no. 2. P. 107–119.

Received 03.09.2024.
Accepted 09.09.2024.

Analysis of approaches of electronic voting systems implementation

The work examines modern approaches to building electronic voting systems, such as blockchain, which promises to revolutionize the process due to its immutability and decentralization properties, as well as traditional cryptographic methods, including homomorphic encryption, which allows vote counting without the need to decrypt each individual vote. Blind signatures ensure the ability to confirm a vote without disclosing the user's identity, and zero-knowledge proofs allow voting without interacting with the server. The goal of the work is to select an approach for building electronic voting systems based on a comparative analysis of their key characteristics. The solved tasks include reviewing the requirements, generalized structures, and main procedures of electronic voting systems; analyzing the existing types of electronic voting systems and their comparative characteristics. During the work, existing systems and other literature were thoroughly analyzed. The article provides a detailed analysis of the advantages and limitations of these technologies, as well as their suitability for dif-

ferent electoral systems, considering important aspects such as scalability, efficiency, and protection against potential threats. Throughout the work, a list of requirements for electronic voting systems was compiled, the main procedures present in electronic voting systems were outlined, a set of actors in typical electronic voting systems was defined, and the generalized structures of their main types were presented. A comparative analysis of the types of electronic voting systems based on compliance with the requirements was conducted. An approach was chosen for further system development.

Остапець Денис Олександрович – к.т.н., доцент, доцент кафедри «Електронні обчислювальні машини» Українського державного університету науки і технологій, ORCID: 0000-0003-1778-7770

Мотиленко Володимир Артемович – аспірант Українського державного університету науки і технологій, ORCID: 0000-0003-3337-9945

Ostapets Denys – candidate of technical sciences, associate professor, associate professor of «Electronic computers» department of Ukrainian State University of Science and Technologies, ORCID: 0000-0003-1778-7770

Motylenko Volodymyr – graduate student of Ukrainian State University of Science and Technologies, ORCID: 0000-0003-3337-9945

ПРОГНОЗУВАННЯ ПОПУЛЯРНОСТІ МУЗИЧНИХ ТРЕКІВ НА ПЛАТФОРМІ SPOTIFY НА ОСНОВІ ЧИСЛОВИХ МЕТРИК

Анотація. В сучасному світі музика відіграє важливу роль у житті мільйонів людей, а платформи для стрімінгу музики, такі як Spotify, стали невід'ємною частиною сучасної культури. Популярність музичних треків має велике значення для музичної індустрії, впливаючи на доходи артистів та тренди в музичному світі. Прогнозування популярності музичних треків є важливим завданням, яке може допомогти артистам, продюсерам та платформам краще розуміти вподобання слухачів та оптимізувати свої стратегії.

Для прогнозування популярності музичних треків застосовувались моделі *Decision Tree Classifier*, *KNeighbors Classifier*, *XGBoost Classifier* та *Random Forest Classifier*. Загальний аналіз показав, що моделі *XGBoost* та *Random Forest* є найбільш ефективними для прогнозування популярності музичних треків. Вони демонструють високу точність і стійкість до змін у наборі атрибутів, що робить їх придатними для використання у реальних умовах.

Ключові слова: інтелектуальний аналіз даних, класифікація, *knn*, *decision tree*, *random forest*, *extreme gradient boosting*.

На сьогоднішній день музична індустрія є однією з найбільш динамічно-розвиваючих секторів економіки багатьох країн. Музичні стрімінгові сервіси, а в особливості Spotify, відіграють одну з найважливіших ролей у поширенні музики в усьому світі. Саме завдяки таким платформам, користувачі мають доступ до величезної бібліотеки музичних композицій, а музиканти, в свою чергу, можуть ефективно і швидко доносити свою творчість та свої почуття до слухачів.

Популярність музичних треків значною мірою впливає на доходи артистів та музичних лейблів. Визначення популярності треків базується на кількості прослуховувань, вподобань та інших метрик, які відображають взаємодію користувачів з музикою. Успішне прогнозування популярності треків дозволяє музичним платформам оптимізувати свої рекомендаційні системи, а артистам та продюсерам — краще планувати релізи та рекламні кампанії.

Метою дослідження є створення програмного забезпечення для аналізу музичних треків на платформі Spotify, їх класифікація за рівнями популярності та подальше прогнозування рівня популярності на основі числових метрик.

Матеріали та методи. Для виконання дослідження було обрано 4 джерела відкритих даних на сайті <https://www.kaggle.com>, а саме:

- загальна інформація про музичні треки на платформі Spotify: <https://www.kaggle.com/datasets/lehaknarnauli/spotify-datasets?select=tracks.csv>

- загальна інформація по виконавцям на платформі Spotify: <https://www.kaggle.com/datasets/lehaknarnauli/spotify-datasets?select=artists.csv>

- додаткова інформація про музичні треки з 1921 по 2020 роки: <https://www.kaggle.com/datasets/ektanegi/spotifydata-19212020?select=data.csv>

- інформація по прослуховуванням за датами, коли трек входив в топ 200 по прослуховуванням: <https://www.kaggle.com/datasets/dhruvildave/spotify-charts?rvi=1>

Для завантаження даних з датасетів до stage-зони використано скрипти, написані як на чистому SQL, так і з використанням мови програмування Python.

У моделі сховища за типом сніжинка спроектовано одна таблиця фактів та сім таблиць вимірів (рис. 1). Така модель сховища легко може бути розширена новими даними, такими як нові виконавці, нові музичні композиції чи кількість прослуховувань за певний період часу. Модель приведена до третьої нормальної форми, всі зв'язки логічні, є можливість комбінувати таблиці для отримання даних у бажаному форматі, що сильно полегшує роботу з даними: їх зберігання, зміни, обробку чи видалення. На основі даної моделі було створено SQLite [1] сховище даних. Подальша робота зі сховищем, зокрема ETL процеси, проводилася за допомогою Python [2].

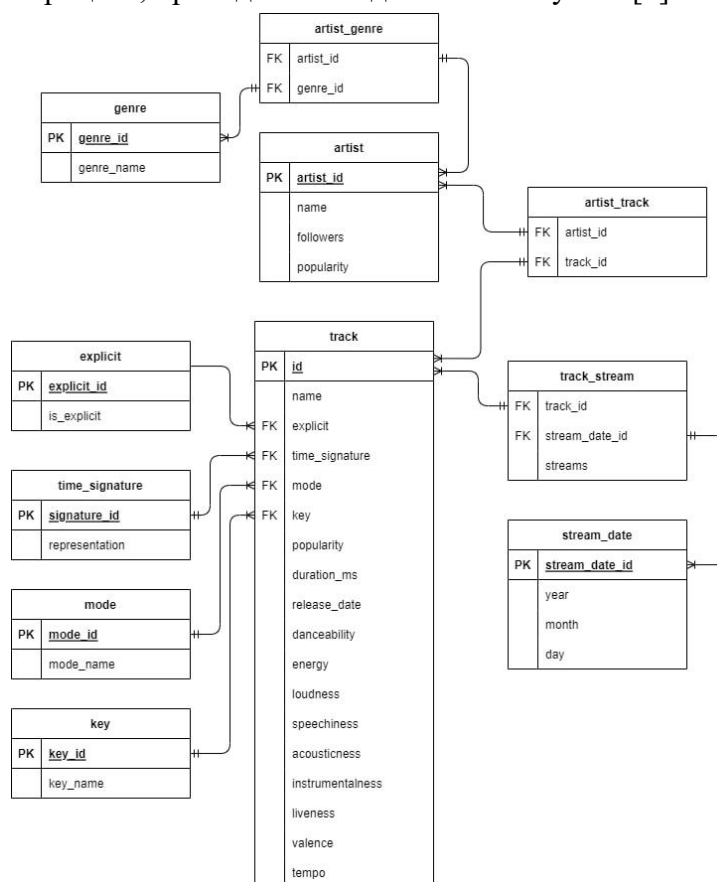


Рисунок 1 – Модель бази даних

Дані було розбито на 3 класи за рівнем популярності (0 – зовсім непопулярний, 1 – середньої популярності, 2 – хіт) (рис.2,3).

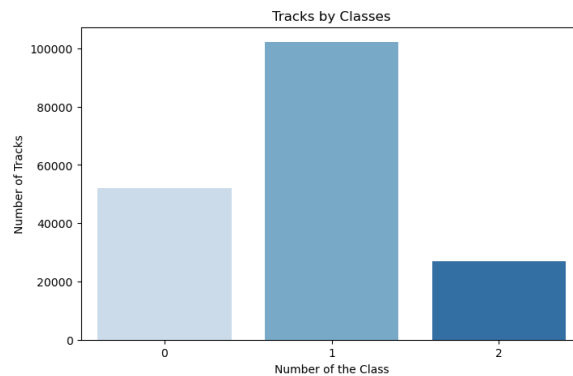


Рисунок 2 – Кількість треків для кожного з класів

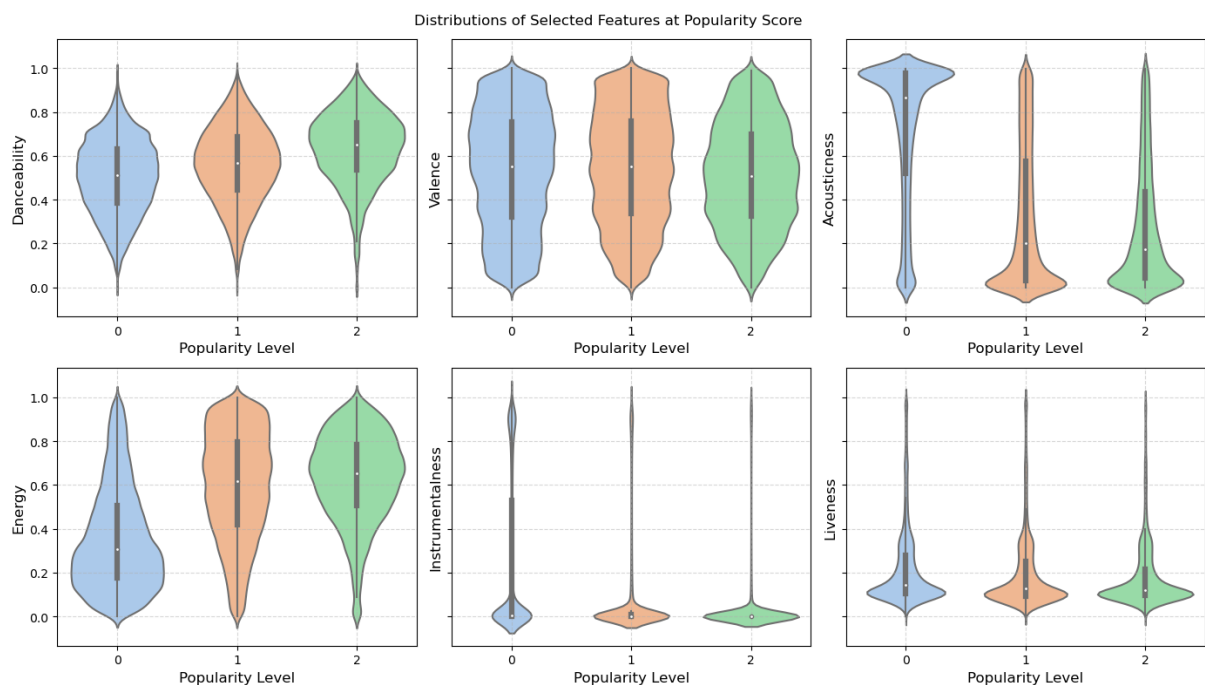


Рисунок 3 – Розподіл рівнів популярності музичних треків між метриками

Результати та основний матеріал дослідження. Для обробки інформації та виявлення в ній моделей та тенденцій для прийняття рішень використано інтелектуальний аналіз даних. Спираючись на числові характеристики інших музичних треків, за допомогою математичних методів можна оцінити майбутні показники популярності, що є актуальним для виявлення тенденцій не тільки для платформи Spotify, а й для багатьох інших музичних платформ та сервісів. В якості моделей для проведення прогнозування популярності музичних треків було обрано наступні: Decision Tree Classifier [3], KNeighbors Classifier [4], XGBoost Classifier [5, 6], RandomForest Classifier [7]. Вибір був зроблений на основі суттєвих факторів, що полегшують проведення класифікації. Дані моделі дозволяють аналізувати великі обсяги даних, виявляти складні залежності та

взаємозв'язки між характеристиками треків, такими як акустичність, темп, валентність, живість тощо.

Для тренування та оцінювання моделей була створена загальна функція, яка спрощує цей процес для різних моделей класифікації. Ця функція приймає наступні параметри: модель, її назву у текстовому вигляді, тренувальну вибірку незалежних змінних, тренувальну вибірку залежної змінної, тестову вибірку незалежних змінних та тестову вибірку залежної змінної. Далі описано кроки, які виконує ця функція:

- модель навчається на тренувальних даних;
- після навчання моделі, на тестових даних виконуються прогнози;
- виконується обчислення метрик якості моделі: accuracy, precision, recall, та f1-score. Метрики обчислюються для кожного з класів окремо, а також розраховуються середні значення для останніх трьох метрик;
- функція виводить обчислені метрики в консоль, графічно зображає їх та виводить матрицю невідповідностей (confusion matrix) для тестових та передбачених даних;
- функція додає обчислені метрики до результуючого масиву. Після тренування всіх моделей та обчислення їх метрик, загальні результати візуалізовані на графіку для порівняння ефективності моделей та вибору найкращої серед них.

Такий підхід дозволяє ефективно проводити порівняльний аналіз моделей класифікації, обраних для прогнозування популярності музичних треків. Результати моделювання представлені на рис.4-7.

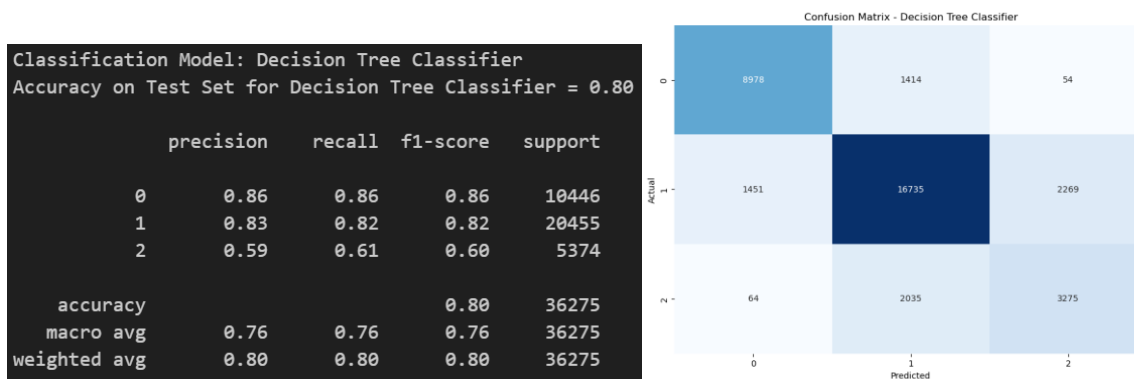


Рисунок 4 – DecisionTree Classifier

Модель показує досить хороші результати, з точністю близько 80% та значенням f1-score 0.76. Проте основною проблемою є її схильність до перенавчання, особливо на великих та складних наборах даних. Це може призводити до поганої генералізації на нових даних.

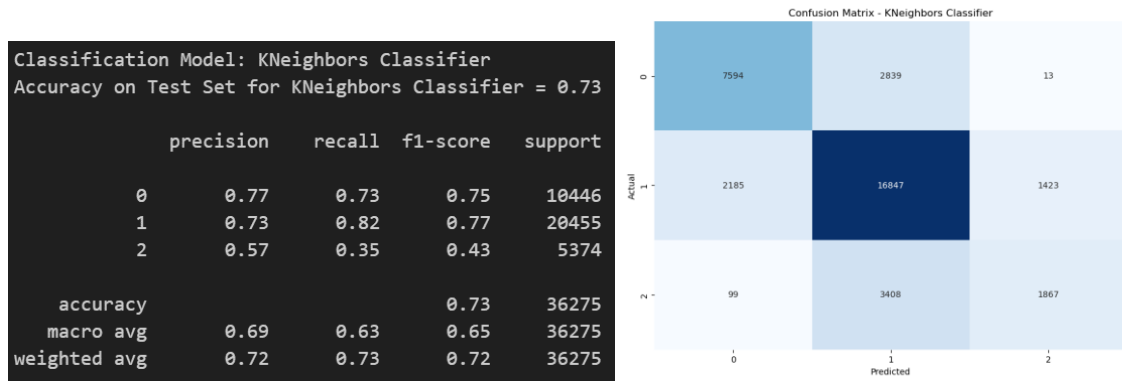


Рисунок 5 – KNearestNeighbors Classifier

Модель KNearestNeighbors Classifier показала середні результати з точністю близько 72.5% та f1-score близько 0.65. Основна проблема цієї моделі полягає в її чутливості до масштабування ознак та вибору значення параметра k (кількість сусідів). Модель може бути менш ефективною на великих наборах даних через високу обчислювальну складність.

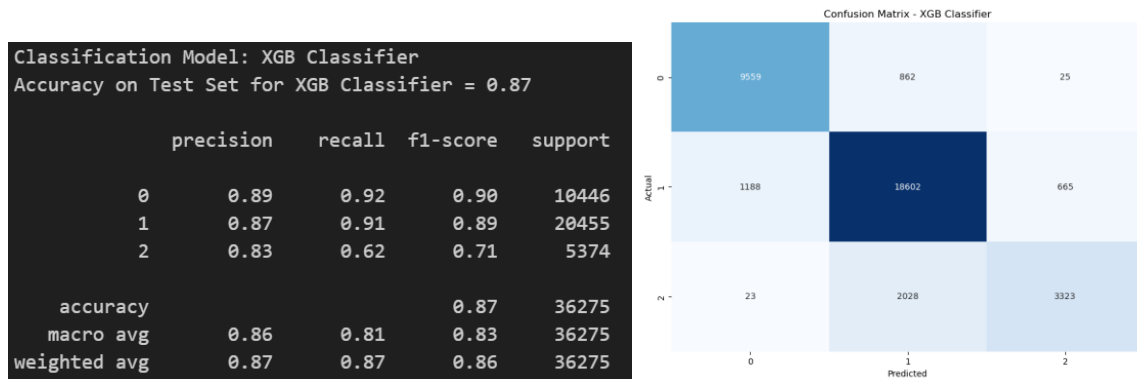


Рисунок 6 – XGBoost Classifier

XGBoost Classifier показав на даний момент найкращі результати серед усіх моделей, з точністю майже 87% та f1-score 0.81. І дійсно, модель має чудову здатність до генералізації та високу стійкість до перенавчання.

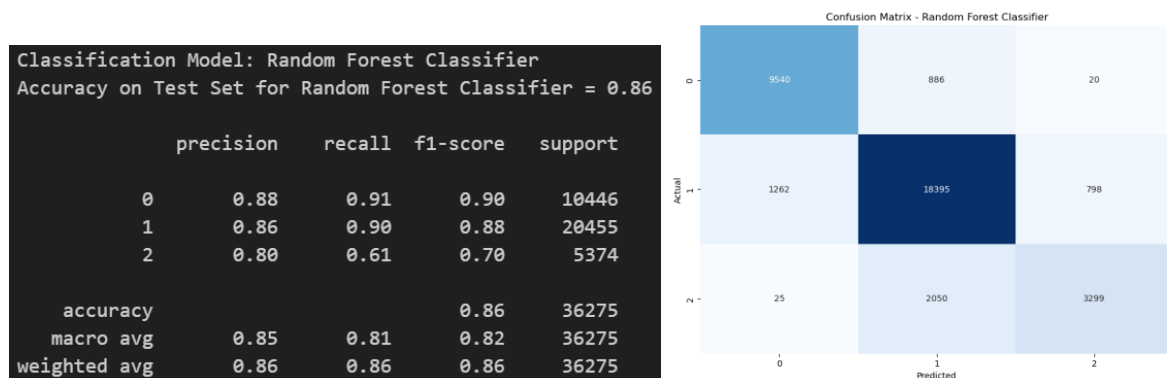


Рисунок 7 – Random Forest Classifier

Random Forest Classifier показав дуже високі результати, з точністю понад 86% та f1-score 0.81. Модель добре справляється зі складними наборами даних і є стійкою до

перенавчання. Проте, обчислювальна складність та час тренування можуть бути досить високими, особливо на великих наборах даних, порівняно з XGBoost Classifier.

Беручи за основу минулі результати було прийнято рішення про підбір параметрів для трьох моделей: KNeighbors Classifier, XGBoost Classifier, RandomForest Classifier. Чому не було включено DecisionTree Classifier замість KNeighbors Classifier? Все доволі просто, занадто велика чутливість до даних та ризик перенавчання робить цю модель поганим рішенням для задачі передбачення рівня популярності музичних треків на основі числових значень треків, адже в реальному світі сховище з музичними треками буде постійно поповнюватися, що може і буде впливати на результати моделі. Тобто навіть якщо зараз модель справляється доволі непогано, немає гарантій, що додавши ще декілька тисяч треків, вона не почне перенавчатися. Знову ж, хоча KNeighbors і не дуже підходить для великих вибірок даних, було прийнято зробити підбір параметрів, адже модель є більш стабільною за DecisionTree Classifier і в перспективі може її обігнати в результуючих метриках.

Для моделі KNeighbors Classifier основними параметрами є `n_neighbors`, `weights`, `algorithm`, `leaf_size` та `p`. Параметр `n_neighbors` визначає кількість сусідів, які враховуються при класифікації, що може впливати на точність і стабільність моделі. Вибір зважування (`weights`) впливає на вагомість близькості сусідів, тоді як алгоритм пошуку (`algorithm`) впливає на швидкість обробки даних. Параметр `leaf_size` визначає розмір листа, що впливає на продуктивність, а параметр `p` обирає метрику відстані між точками: 1 – Manhattan distance, 2 – Euclidean distance. Після підбору параметрів, були отримані наступні результати (рис.8).

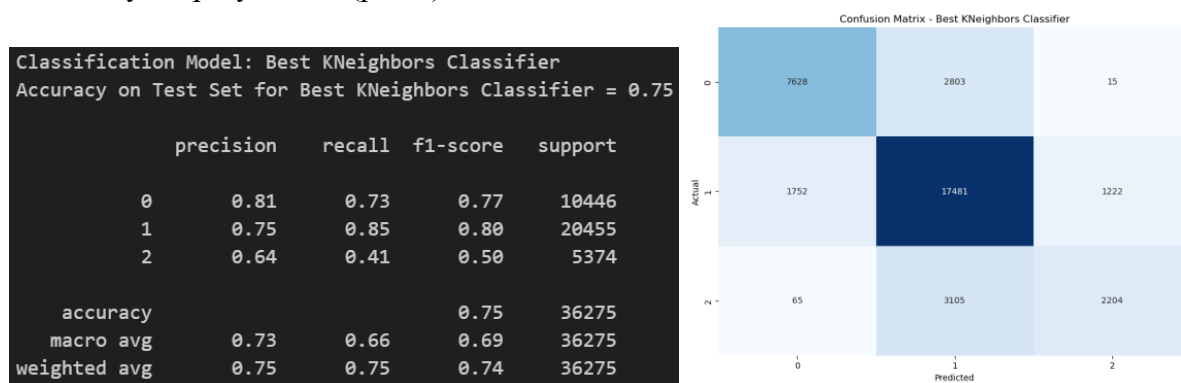


Рисунок 8 – Метрики KNeighbors Classifier з підібраними параметрами

Модель продемонструвала покращення в усіх метриках, що вказує на те, що налаштування параметрів позитивно вплинуло на її продуктивність. Але навіть у такому разі, продуктивність моделі все ще була нижчою порівняно з іншими моделями.

Для моделі XGBoost Classifier ключовими параметрами є `n_estimators`, `max_depth`, `min_child_weight` та `gamma`. Параметр `n_estimators` визначає кількість дерев у моделі, що покращує її продуктивність, але збільшує час навчання. Максимальна глибина дерева (`max_depth`) дозволяє моделювати складніші закономірності, але ризикує перенавчанням, тоді як мінімальна вага листа (`min_child_weight`) робить модель стійкішою до шуму. Параметр `gamma` регулює мінімальне зменшення втрат для розбиття вузла, що

робить модель більш консервативною і знижує ризик перенавчання. Після підбору параметрів, були отримані наступні результати (рис. 9):

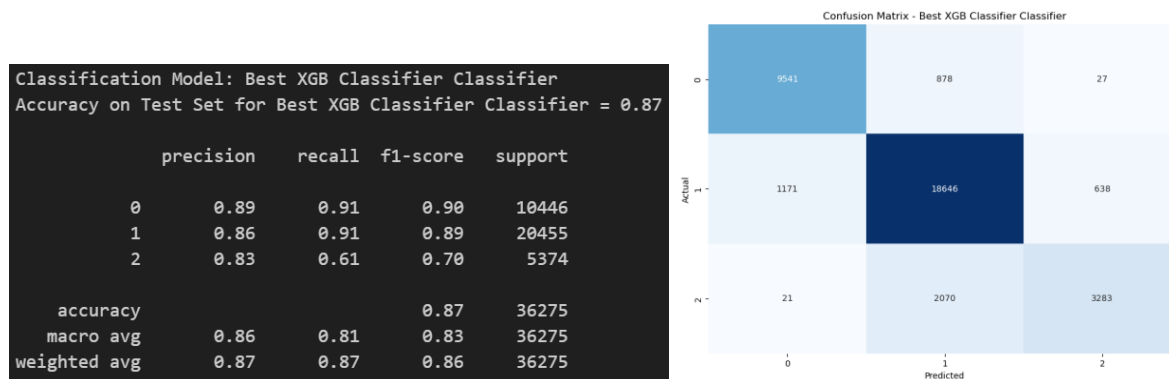


Рисунок 9 – Метрики XGBoost Classifier з підібраними параметрами

Налаштування гіперпараметрів для XGBoost Classifier також призвело до незначного покращення результатів, зробивши цю модель однією з найкращих для даної задачі. Точність склала 87%, а значення f1-score становить 0.84, що є доволі гарним результатом, враховуючи специфіку задачі.

Останньою моделлю, для якої було виконано підбір гіперпараметрів, була RandomForest Classifier. У випадку з цією моделлю, важливими параметрами є `n_estimators`, `max_depth`, `min_samples_split`, `min_saples_leaf` та `bootstrap`. Параметр `n_estimators` встановлює кількість дерев у лісі, що підвищує стабільність, але збільшує час обробки. Максимальна глибина дерева (`max_depth`) контролює перенавчання, дозволяючи моделі вловлювати складні закономірності або знижуючи ризик перенавчання. Мінімальна кількість зразків для розбиття вузла (`min_samples_split`) і мінімальна кількість зразків у листі (`min_saples_leaf`) впливають на гнучкість дерева, а параметр `bootstrap` визначає використання вибірки з поверненням для навчання кожного дерева, що впливає на різноманітність дерев. Після підбору параметрів, були отримані наступні результати (рис. 10):

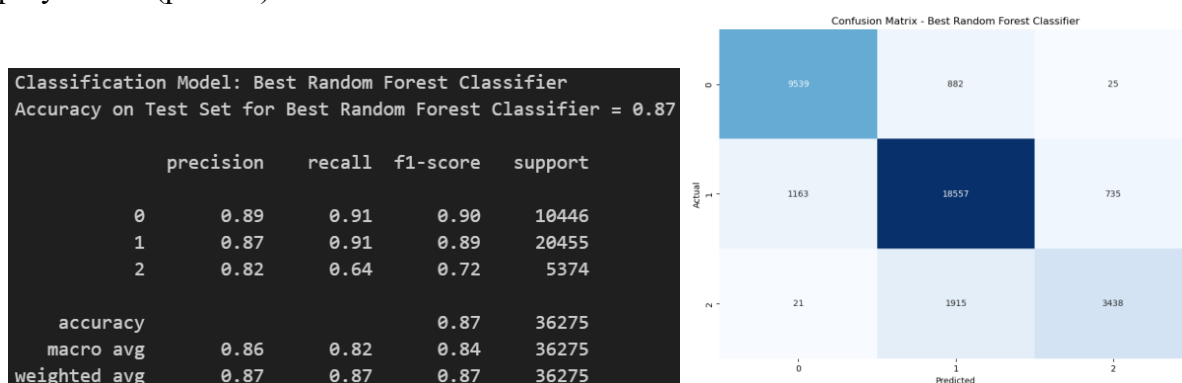


Рисунок 10 – Метрики RandomForest Classifier з підібраними параметрами

Після налаштування гіперпараметрів для Random Forest Classifier, результати стали ще трохи кращими, що підкреслює високу ефективність цієї моделі для задачі кла-

сифікації популярності музичних треків. Незначне покращення присутнє в усіх метриках: accuracy, precision, recall та f1-score.

Для більш наглядного порівняння було створено графік ключових метрик, за якими проводилась оцінка моделей та ROC-крива для кожної з них (рис. 11-12):

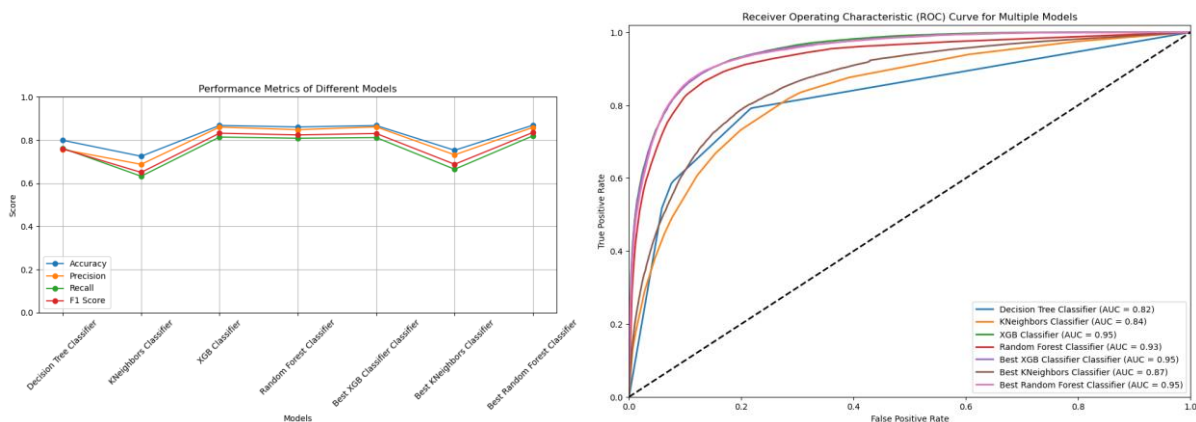


Рисунок 11 – Порівняння оцінок ключових характеристик моделей

Algorithm	Accuracy	Precision	Recall	F1 Score
Decision Tree Classifier	0.799118	0.75659	0.76234	0.759369
KNeighbors Classifier	0.725238	0.68784	0.632668	0.650443
XGB Classifier	0.867926	0.860383	0.814282	0.832014
Random Forest Classifier	0.861034	0.848267	0.808814	0.824178
Best XGB Classifier Classifier	0.86754	0.861327	0.811943	0.830737
Best KNeighbors Classifier	0.752943	0.731846	0.664987	0.688152
Best Random Forest Classifier	0.869304	0.859187	0.820043	0.835757

Рисунок 12 – Порівняння характеристик моделей

Після виконання налаштування гіперпараметрів, XGBoost Classifier та Random Forest Classifier продемонстрували найвищі показники метрик, що робить їх найефективнішими моделями для прогнозування популярності музичних треків у даній курсовій роботі.

Після цього було проведено ідентичний процес, але вже без використання атрибуту “year” для музичних треків. Цей атрибут сильно впливає на популярність треків на платформі Spotify, що є специфікою даної платформи. Після отримання гарних результатів в основній частині, стало цікаво провести схожий алгоритм дій без урахування специфіки платформи, що дозволить розширити застосування моделей на більшу кількість музичних платформ чи стрімінгових сервісів. На рис. 13 можна бачити метрики accuracy, precision, recall та f1-score для кожного з класів, для моделей DecisionTree

Classifier, KNeighbors Classifier, XGBoost Classifier та RandomForest Classifier без урахування підбору гіперпараметрів.

Algorithm	Accuracy	Precision	Recall	F2 Score
Decision Tree Classifier	0.662578	0.605515	0.609909	0.607624
KNeighbors Classifier	0.554101	0.4446	0.421941	0.415558
XGB Classifier	0.75909	0.734379	0.670632	0.691422
Random Forest Classifier	0.737202	0.702892	0.650093	0.665126
Best XGB Classifier Classifier	0.757657	0.733415	0.668804	0.689604
Best KNeighbors Classifier	0.585665	0.506123	0.458361	0.466088
Best Random Forest Classifier	0.758401	0.741727	0.663554	0.686688

Рисунок 13 – Порівняння метрик оцінювання моделей, без використання атрибуту “year”

З результатів видно, що при виключенні року виходу музичного треку з атрибутів, точність та f1-score моделей XGBoost Classifier та Random Forest Classifier залишаються високими. Це свідчить про їх стійкість і ефективність навіть без інформації про рік випуску треку. Натомість, моделі Decision Tree Classifier та KNeighbors Classifier демонструють помітно нижчі результати, що вказує на їхню більшу залежність від повного набору атрибутів. Таким чином, для задачі прогнозування популярності музичних треків у ситуаціях, коли рік виходу невідомий, найкращим вибором є моделі XGBoost Classifier та Random Forest Classifier.

Висновки. Загальний аналіз показав, що моделі XGBoost та Random Forest є найбільш ефективними для прогнозування популярності музичних треків. Вони демонструють високу точність і стійкість до змін у наборі атрибутів, що робить їх придатними для використання у реальних умовах.

ЛІТЕРАТУРА / REFERENCES

1. SQLite. SQLite Home Page. URL: <https://www.sqlite.org/index.html> (date of access: 26.05.2024).
2. Python. The official home page of the Python Programming Language. URL: <https://www.python.org/> (date of access: 26.05.2024).
3. DecisionTree Classifier. scikit-learn. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html> (date of access: 26.05.2024).
4. KNeighbors Classifier. scikit-learn. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html> (date of access: 26.05.2024).

5. XGBoost Documentation. XGBoost Documentation. URL: <https://xgboost.readthedocs.io/en/stable/index.html> (date of access: 25.05.2024).
6. Tabulate Documentation. URL: https://pyneng.readthedocs.io/en/latest/book/12_useful_modules/tabulate.html (date of access: 26.05.2024).
7. RandomForest Classifier. scikit-learn. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html> (date of access: 26.05.2024).

Received 11.09.2024.

Accepted 13.09.2024.

Predicting the popularity of music tracks on Spotify based on numerical metrics

In today's world, music plays an important role in the lives of millions of people, and music streaming platforms such as Spotify have become an integral part of modern culture. The popularity of music tracks is of great importance to the music industry, affecting artists' incomes and trends in the music world. Predicting the popularity of music tracks is an important task that can help artists, producers, and platforms better understand listener preferences and optimize their strategies.

As part of this work, a data storage of music tracks on the Spotify platform has been developed, based on a physical model of the database, the functionality of which is implemented using SQL scripts. Working with the database is presented through the implementation of software for the implementation of ETL processes and intelligent analysis of selected data. The software allows you to classify tracks by the level of popularity (0 - not at all popular, 1 - medium popularity, 2 - hit) using numerical track metrics such as acousticness, tempo, valence, liveness, etc. The role of the data storage management system is SQLite, the programming language for implementing the application is Python.

Different machine learning models are used to predict track popularity, including KNeighbors, Decision Tree, Random Forest, and Extreme Gradient Boosting. Data mining software provides efficient track classification and graphical display, allowing users to easily interpret forecasting results. Libraries used in the work: pandas, numpy, seaborn, matplotlib, tabulate, xgboost, scipy, sqlite3.

The overall analysis showed that the XGBoost and Random Forest models are the most effective for predicting the popularity of music tracks. They demonstrate high accuracy and resistance to changes in the set of attributes, which makes them suitable for use in real conditions.

Keywords: intelligent data analysis, classification, KNeighbors, Decision Tree, Random Forest, Extreme Gradient Boosting.

Бур Антон Олександрович – студент кафедри інформатики та програмної інженерії, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

Ліхоузова Тетяна Анатоліївна – к.т.н., доцент кафедри інформатики та програмної інженерії, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

Олійник Юрій Олександрович – к.т.н., доцент кафедри інформатики та програмної інженерії, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

Bur Anton – student, Department of Informatics and Software Engineering National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute».

Likhouzova Tetiana – PhD, associate professor, Department of Informatics and Software Engineering National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute».

Oliinyk Yurii – PhD, associate professor, Department of Informatics and Software Engineering National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute».

КОРЕЛЯЦІЙНІ ВЛАСТИВОСТІ ГІСТОГРАМ КОЛЬОРУ ПРИ ЗНИЖЕННІ ЯКОСТІ ЗОБРАЖЕНЬ

Анотація. В роботі досліджено можливість використання гістограм кольору, сформованих у колірних моделях HS^* для ідентифікації зображень при накладанні на них спотворень різної природи. Розглянуто спотворення, які можуть виникати в реальних умовах при отриманні зображень, зокрема, поворот, накладання шуму, розмиття і оптична дисторсія. Порівняння гістограм кольору оригінальних та спотворених зображень здійснювалося за допомогою обчислення коефіцієнту їх взаємної кореляції. Виявлено, що в межах того самого зображення кореляція є досить високою, надивлячись на накладені спотворення, що дозволяє допустити застосовність даного підходу для задач розпізнавання образів.

Ключові слова: спотворені зображення, гістограма кольору, колірний простір HS^* , коефіцієнт кореляції, *OpenCV*.

Постановка проблеми. У задачах розпізнавання образів ідентифікація об'єктів на зображеннях здійснюється з використанням ознак різної природи. Сюди входять текстура і колір, виділення контурів, статистичні та семантичні описи тощо [1–3]. Серед методів отримання атрибутів об'єктів поширеним є аналіз гістограм їх інтенсивності. Хоча у цифровій обробці зображень традиційно використовується трьохканальне кольорове представлення RGB , аналіз колірних властивостей доцільно здійснювати в системах HS^* (HSL , HSV та HSI), оскільки вони найкраще відображають спектральне подання кольору [4]. Спільною рисою цих систем є спосіб визначення кольорової координати H (*Hue*), яка в усіх трьох системах є кутовою координатою у циліндричній системі координат. Насиченість S (*Saturation*) задається радіальною відстанню, а зміна яскравості I (*Intensity*), V (*Value*) або L (*Lightness*) визначається висотою. Спосіб обчислення величин S , I , V та L у кожній системі є різним. Таким чином в усіх системах HS^* *Hue* має зміст спектрального складу кольору і може бути використана у ролі спектрального ідентифікатора зображення чи його частини.

Мета дослідження. У роботі досліджувалась можливість використання гістограм кольору, сформованих у колірних просторах HS^* для ідентифікації (порівняння) зображень з накладеними на них спотвореннями.

Об'єкти і методика досліджень. Для досліджень було обрано набір зображень (*dataset*) CQ100: A High-Quality Image Dataset for Color Quantization Research [5]. Даний набір, в першу чергу, служить для вирішення завдань, пов'язаних з квантуванням ко-

люру [6], однак, може бути використаний в інших задачах обробки зображень, таких як фільтрація, сегментація тощо. Базовий (не квантизований) набір містить 100 зображень розміром 768×512 пікселів в *RGB* представлені, з глибиною кольору 24 біти (8 біт на кожен з каналів кольору). Зображення надаються у форматі *png*, який використовує алгоритми стискування без втрат даних.

Проведено дослідження впливу на колірні характеристики зображень різних спотворень, які можуть виникнути при фото/відео зйомці. Зокрема, розглядалися: поворот, дисторсія, розмиття і накладання шумів.

Для цього обчислювалися гистограми кольорової складової (*Hue*) зображень у кольоровому просторі *HSV* для оригінального і спотвореного зображення і здійснювалося їх порівняння шляхом обчислення коефіцієнту кореляції *R* за Пірсоном [7]:

$$R(H_1, H_2) = \frac{\sum_i (H_1(i) - \bar{H}_1)(H_2(i) - \bar{H}_2)}{\sqrt{\sum_i (H_1(i) - \bar{H}_1)^2 (H_2(i) - \bar{H}_2)^2}}, \quad (1)$$

де $H_1(i)$, $H_2(i)$ – значення функції розподілу на *i*-му відліку гистограми, \bar{H}_1 , \bar{H}_2 – середнє значення для 1-ої та 2-ої гистограм. Значення коефіцієнту кореляції змінюється у межах від мінус до плюсу одиниці, що відповідає мінімальній і максимальній кореляції, відповідно. При відсутності кореляції $R = 0$.

При багатьох типах спотворень, зокрема, у нашому випадку при повороті та дисторсії, відбувається зміна геометрії об'єкту. Спотворені зображення займають тільки частину площі основного зображення і, очевидно, для порівняння потрібно розглядати лише цю ділянку.

В бібліотеці *OpenCV* є засоби для виділення області інтересів (англ. *ROI – Region Of Interest*) [8], однак вони передбачають прямокутну форму цієї області. Для виділення на спотворених зображеннях *ROI* не прямокутної форми ми розробили наступний підхід. На оригінальне зображення накладалася маска, отримана з *ROI* спотвореного зображення. Маска отримувалася таким чином: спотворене зображення розмивалося (фільтрувалося методом усереднення), конвертувалося з кольорового представлення *RGB* у представлення відтінки сірого (*GreyScale*), після чого бінаризувалося. Попереднє розмиття здійснювалося для ліквідації чорних ділянок всередині маски, які виникали на місці областей чорного кольору *ROI* зображення. Експериментальним шляхом було встановлено, що чорні ділянки всередині маски зникають, коли значення розміру ядра фільтрації досягає 17 пікселів. Оскільки внаслідок фільтрації відбувається також незначне розмиття контурів зображення, отримана вищезгаданим методом маска накладалася як на оригінальне, так і на спотворене зображення (з якого вона була отримана) – для повного співпадіння *ROI*. Накладання маски здійснювалося побітовою кон'юнкцією зображення і маски.

Процедура порівняння гистограм є практично однаковою для всіх типів спотворень і складається з наступних кроків:

1. Завантаження оригінального зображення в трьохканальному кольоровому представленні.
2. Створення спотвореного зображення. При цьому оригінальне зображення зберігається.
3. Обрахунок маски за спотвореним зображенням. Накладання маски на оригінальне і спотворене зображення.
4. Конвертація замаскованих оригінального та спотвореного зображення з кольорового представлення *BGR* у представлення *HSV*.
5. Розрахунок гістограм замаскованих оригінального та спотвореного зображення (гістограма розраховувалась тільки для каналу кольору (*Hue*)).
6. Порівняння гістограм оригінального та спотвореного зображення за допомогою обчислення коефіцієнту їх взаємної кореляції.

Результати досліджень.

Дослідження зміни коефіцієнту кореляції при повороті зображення. На рисунку 1 наведено середнє значення та медіана коефіцієнту кореляції для 100 зображень набору даних CQ100 при їхній ротації. Вони відображають загальну тенденцію зміни коефіцієнту кореляції залежно від кута повороту, який змінювався від 0 до 360 градусів з кроком 10 градусів проти годинникової стрілки.

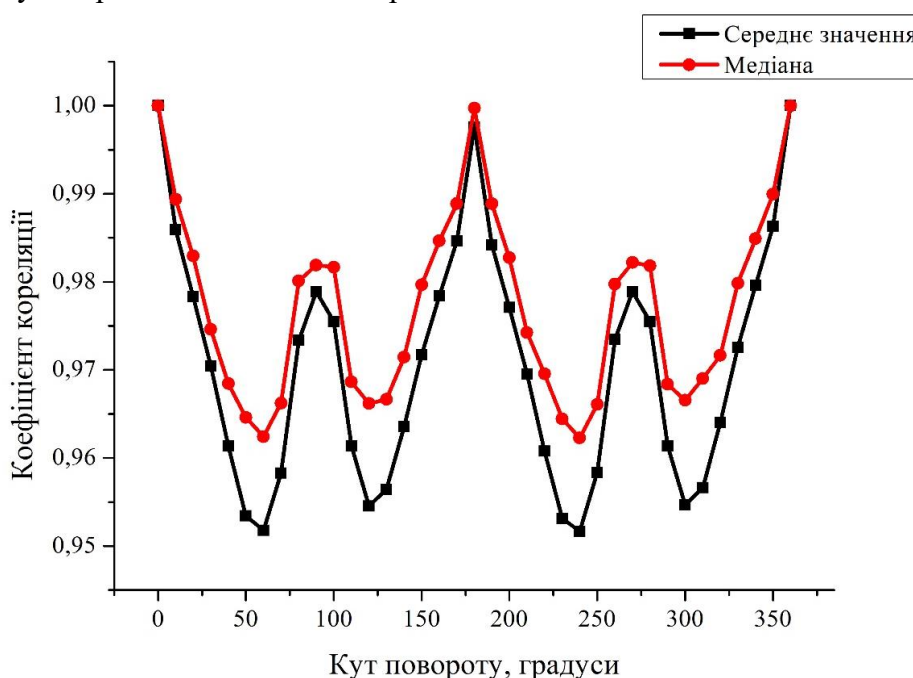


Рисунок 1 – Середнє значення та медіана залежностей коефіцієнту кореляції від кута повороту зображення для набору даних CQ100

З рисунку видно, і це зрозуміло з властивостей симетрії, що хід кривих від 0 до 180 градусів і від 180 до 360 градусів є абсолютно ідентичним.

Від 0 до 60 градусів коефіцієнт кореляції спадає, досягаючи першого, глибшого мінімуму, далі зростає, досягаючи локального максимуму при куті повороту 90 градусів, знов спадає, досягаючи другого мінімуму при 120 градусах повороту і зростає до

значення одиниці (повна кореляція) при куті 180 градусів (перевернуте зображення). Отриманий хід кривих і значення кутів, при яких досягаються максимуму і мінімуми також легко пояснити з міркувань симетрії.

Слід зауважити, що хід кривих середнього значення і медіани коефіцієнту кореляції є характерним для усіх досліджуваних зразків і, загалом, коефіцієнт кореляції змінюється не дуже сильно при повороті зображення. Найменше значення коефіцієнту кореляції 0,753 виявлено для зображення *easter_eggs.png* [5].

Незначне відхилення значення коефіцієнту кореляції від одиниці для більшості тестових зразків може свідчити про можливість застосування методики порівняння зображень на основі обрахунку коефіцієнту їх кореляції [9] при трансформації повороту.

Дослідження зміни коефіцієнту кореляції при зашумленні зображення. Проведено дослідження залежності коефіцієнту кореляції оригінальних та зашумлених зображень для набору даних CQ100. Для дослідження було обрано гаусівський шум [10], оскільки поява такого типу шуму найчастіше трапляється при отриманні зображень.

На рисунку 2 наведено зміну середнього значення та медіани коефіцієнту кореляції для 100 зображень набору даних CQ100 в залежності від ступеня зашумленості. Ступінь зашумленості змінювався від 0,0 до 1,0 з кроком 0,1. При цьому значення параметрів нормального розподілу: математичне сподівання та середньоквадратичне відхилення залишалися постійними і становили 0,0 та 1,0, відповідно.

Для більшості зразків коефіцієнт кореляції зменшується дуже повільно аж до значення ступеня зашумленості 0,9, після чого спостерігається більш інтенсивний спад. Загалом, (рисунком 2) спостерігається незначна зміна коефіцієнту кореляції при накладанні гаусівського шуму на досліджувані об'єкти, хоча існують окремі винятки. Наприклад, для зображення *mandarin_ducks.png* [5] мінімальне значення коефіцієнту кореляції при повному зашумленні становить 0,619.

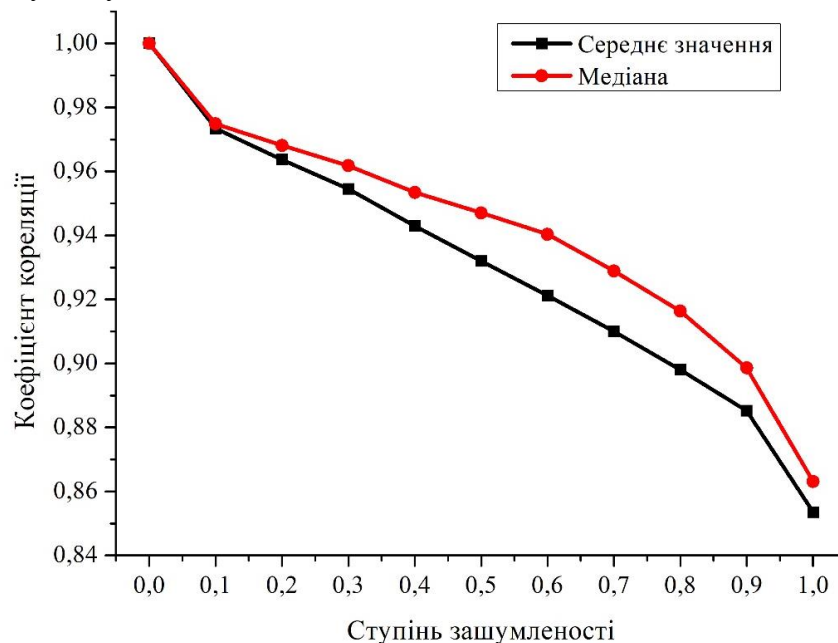


Рисунок 2 – Залежність коефіцієнту кореляції від ступеня зашумленості зображення гаусівським шумом: середнє значення та медіана для набору даних CQ100

Дослідження зміни коефіцієнту кореляції при розмитті зображення. Досліджено зміну коефіцієнту кореляції оригінальних та розмитих гаусівським фільтром зображень залежно від розміру ядра. Розмір ядра змінювався від 1 до 55 з кроком 2 (цей параметр повинен приймати невід'ємні непарні значення). Медіана та середнє значення цих залежностей для набору CQ100 наведені на рисунку 3. При цьому середньоквадратичне відхилення нормального розподілу обчислювалося автоматично на основі розміру ядра за методикою [11].

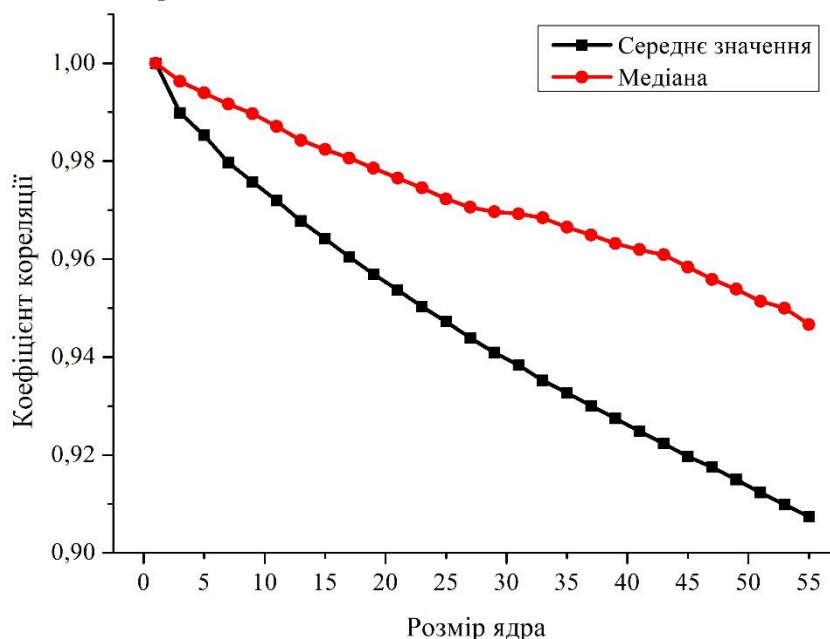


Рисунок 3 – Залежність коефіцієнту кореляції від величина ядра гаусівського фільтра при розмитті зображень: середнє значення та медіана для набору даних CQ100

Отримані результати свідчать, що зміна розміру ядра гаусівського фільтра не сильно впливає на зміну коефіцієнту кореляції, для багатьох зразків його спадання досягає значення 0,8...0,9, в той же час як зміна розміру ядра є суттєвою. Найменше значення коефіцієнту кореляції спостерігалось для зображення *medicine_packets.png* [5] і становило 0,501 при розмірі ядра 55.

Дослідження зміни коефіцієнту кореляції при дисторсії зображення. Дисторсією називають вид аберації оптичних систем, при якій коефіцієнт лінійного збільшення змінюється при віддаленні об'єктів, що відображаються від оптичної осі. У випадку, коли лінійне збільшення зростає при віддаленні елементів зображення від оптичної осі виникає дисторсія типу „подушка”, якщо ж лінійне збільшення зменшується – виникає дисторсія типу „бочка” [12]. Якщо положення точки задано радіус вектором \vec{r} , то у випадку дисторсії третього порядку радіус вектор зображення цієї точки \vec{R} визначається співвідношенням:

$$\vec{R} = b_0 \vec{r} + F_3 r^2 \vec{r}, \quad (2)$$

де b_0 – коефіцієнт лінійного збільшення, r – довжина вектора \vec{r} , F_3 – коефіцієнт дисторсії третього порядку (дисторсії вищих порядків не розглядалися). Коли коефіцієнт F_3 має той самий знак, що й коефіцієнт b_0 , виникає дисторсія типу „подушка”, в протилежному випадку – типу „бочка”.

Досліджено залежність коефіцієнту кореляції оригінальних зображень з набору даних CQ100 та їх дисторсій типу „бочка” та „подушка” від величини коефіцієнту F_3 зі співвідношення (2). На рисунку 4 наведено відповідні залежності середнього значення та медіани при дисторсії типу „бочка” (в цьому випадку коефіцієнти F_3 є додатними). На рисунку 5 подано аналогічні залежності для дисторсії типу „подушка” (в цьому випадку коефіцієнти F_3 є від’ємними). Значення коефіцієнту F_3 змінювалося від 0,0 до $\pm 10,0$ з кроком $\pm 0,5$.

З рисунків 4 і 5 видно, що коефіцієнт кореляції зменшується при зростанні модуля параметра F_3 . Для дисторсії типу „бочка” зменшення є плавним: спочатку більш різким до значень F_3 в околі 1,0...1,5 з подальшим виходом на майже горизонтальну ділянку. Для дисторсії типу „подушка” спочатку спостерігається різке падіння коефіцієнту кореляції з чітко вираженим мінімумом в околі $F_3 = -1,0$ з подальшим ростом і виходом на майже горизонтальну ділянку. Очевидно, що найбільших змін коефіцієнт кореляції зазнає в околі значень модуля параметру F_3 0,0...2,0.

Мінімальне значення коефіцієнту кореляції 0,302 при дисторсії типу „бочка” спостерігалось для зображення *birthday_balloons.png* (при $F_3 = 10$) і 0,144 при дисторсії типу „подушка” для зображення *veggie_pizza.png* [5] (при $F_3 = -0,5$).

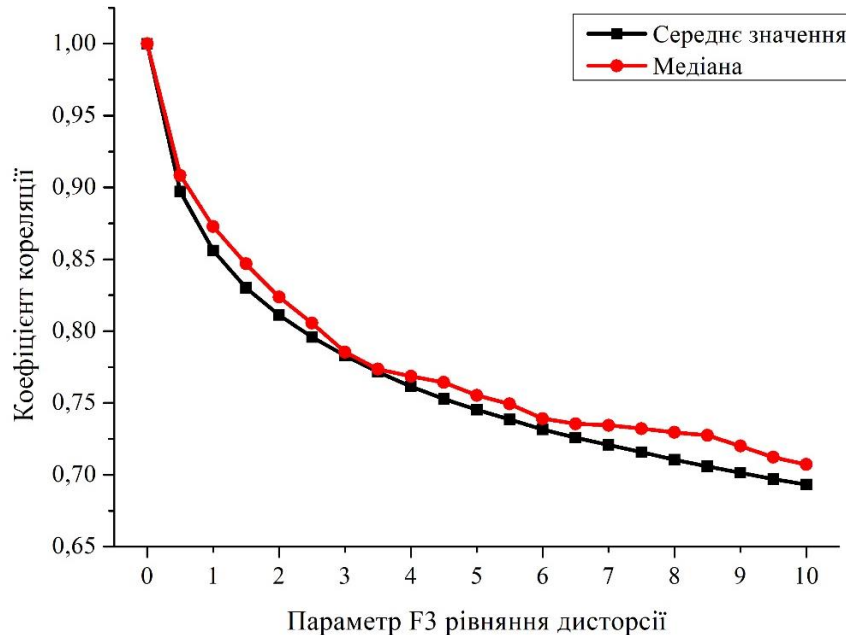


Рисунок 4 – Залежність коефіцієнту кореляції від значення параметру F_3 для дисторсії типу „бочка”: середнє значення та медіана для набору даних CQ100

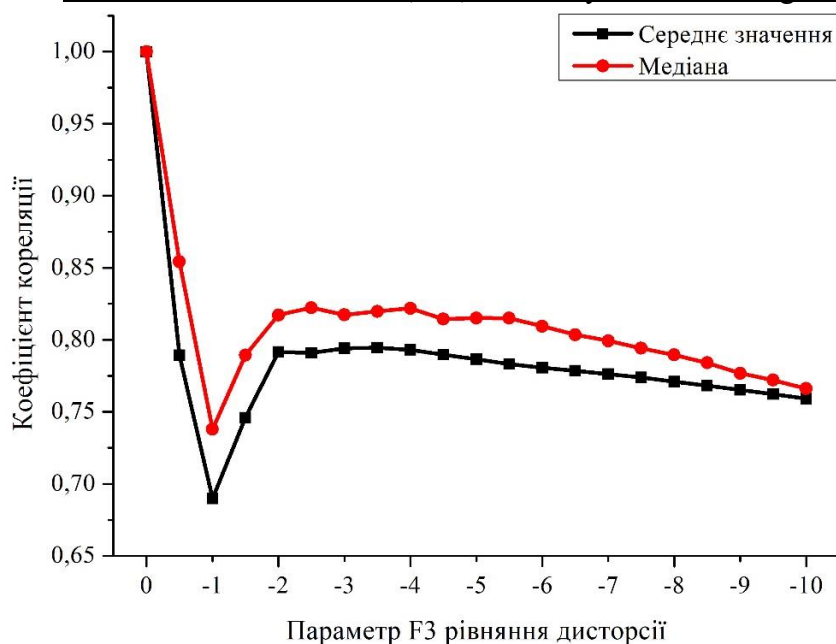


Рисунок 5 – Залежність коефіцієнту кореляції від значення параметру F_3 для дисторсії типу „подушка”: середнє значення та медіана для повного набору даних CQ100

Висновки. Результати проведених досліджень дозволяють припустити допустимість використання гістограм кольору для ідентифікації зображень чи їх фрагментів навіть при досить високих ступенях їх спотворення завадами різної природи, які є типовими за реальних умов реєстрації зображень.

Також можна стверджувати про можливість використання коефіцієнту кореляції, як параметру для порівняння гістограм, недивлячись на відносну простоту його обчислення. Такий підхід дозволить створювати більш швидкодіючі системи розпізнавання образів, порівняно з системами, заснованими на складніших методах порівняння.

ЛІТЕРАТУРА / REFERENCE

1. Burger W., Burge M.J. Principles of Digital image processing: Core Algorithms // Springer. – 2009. – 344 P.
2. Gonzalez R.C., Woods R.E. Digital Image Processing // [4th ed.] – Pearson Education Ltd. – 2017. – 1192 P.
3. Duda R.O., Hart P.E., Stock D.G. Pattern Classification // [2nd ed.] – Wiley. – 2000. – 688 P.
4. Agoston M.K. Computer Graphics and Geometric Modeling: Implementation and Algorithms // [5th ed.] – Springer. – 2005. – 928 P.
5. Celebi M.E., Pérez-Delgado M.-L. CQ100: A High-Quality Image Dataset for Color Quantization Research // Mendeley Data, V2. – 2022. DOI: 10.17632/vw5ys9hfxw.2
6. Celebi M.E. Forty years of color quantization: a modern, algorithmic survey // Artif. Intell. Rev. – 2023. – Vol. 56. – pp. 13953–14034. DOI: 10.1007/s10462-023-10406-6.
7. Bland M. An Introduction to Medical Statistics // [4th ed.] – Oxford University Press. – 2015. – 446 P.

8. OpenCV Documentation. Basic Operations on Images. URL: https://docs.opencv.org/4.x/d3/df2/tutorial_py_basic_ops.html (дата звернення/Accessed: 15.08.2024).
9. Furgala Yu., Velhosh A., Velhosh S., Rusyn B. Using Color Histograms for Shrunk Images Comparison // 2021 IEEE XIIth International Conference on Electronics and Information Technologies (ELIT). – 2021. – pp. 130-133. DOI: 10.1109/ELIT53502.2021.9501117
10. Miller S., Childers D. Probability and Random Processes With Applications to Signal Processing and Communications // [2nd ed.] – Academic Press. – 2012. – 522 P.
11. OpenCV Documentation. Smoothing Images. URL: https://docs.opencv.org/4.x/d4/d13/tutorial_py_filtering.html (дата звернення/Accessed: 15.08.2024).
- Kingslake R., Johnson R.B. Lens Design Fundamentals // [2nd ed.] – Academic Press. – 2009. – 576 P.

Received 11.09.2024.
Accepted 15.09.2024.

Correlation Properties of Color Histograms in the Case of Image Quality Decreasing

In image recognition tasks, objects are identified by examining various features such as texture, color, contour detection, and statistical or semantic descriptions. One widely used approach for extracting image attributes is the analysis of intensity histograms. While the traditional RGB color model is commonly used in digital image processing, it is often more effective to analyze color properties in HS systems (such as HSL, HSV, and HSI) since these systems more closely resemble the spectral representation of color. A key characteristic shared by these three systems is the use of the H (Hue) coordinate, which is represented as an angular value within a cylindrical coordinate system. The paper investigates the possibility of using color histograms generated in HS* spaces for identifying images that have undergone various types of distortions. The CQ100: A High-Quality Image Dataset for Color Quantization Research was chosen for the research. The non-quantized section of the CQ100 dataset consists of 100 RGB images in PNG format, each with a resolution of 768×512 pixels and a color depth of 24 bits. The study examines how different distortions, which can occur during real-time photo and video capture, affect the color properties of images. Specifically, the research focuses on distortions caused by rotation, noise, blurring, and optical aberration. Histograms were compared using the Pearson cross-correlation coefficient, and the findings reveal that the correlation remains high for the same image despite the applied distortions. Conversely, the correlation coefficient between different images is low for most of the studied objects. These results suggest that color histograms could be effectively used for image identification tasks, even when images are significantly distorted, as is common in image registration processes. The applicability of correlation detection as a method for histograms comparison is considered regardless of the relative simplicity of its calculation. This approach could contribute to the development of faster image recognition systems.*

Key words: distortion of images, color histogram, HS color spaces, Pearson correlation, OpenCV.*

Вельгош Андрій Сергійович – аспірант кафедри оптоелектроніки та інформаційних технологій факультету електроніки та комп'ютерних технологій Львівського національного університету імені Івана Франка.

Фургала Юрій Михайлович – к.ф.м.н., доцент, декан факультету електроніки та комп'ютерних технологій Львівського національного університету імені Івана Франка, доцент кафедри оптоелектроніки та інформаційних технологій факультету електроніки та комп'ютерних технологій Львівського національного університету імені Івана Франка, ORCID: 0000-0001-7316-7424.

Velhosh Andriy – Postgraduate Student of the Department of Optoelectronics and Information Technologies, Faculty of Electronics and Computer Technologies, Ivan Franko National University of Lviv.

Furgala Yuriy – Candidate of Physical and Mathematical Sciences, Associate professor, Dean of Faculty of Electronics and Computer Technologies, Ivan Franko National University of Lviv, Associate Professor of the Department of Optoelectronics and Information Technologies, Faculty of Electronics and Computer Technologies, Ivan Franko National University of Lviv, ORCID: 0000-0001-7316-7424.

О.О. Яворська, Л.А. Новіков, М.Ю. Іконніков

ЗМІНА КОНЦЕНТРАЦІЇ ПИЛУ У ГАЗОПИЛОВІЙ ХМАРІ ПРИ ВИБУХОВИХ РОБОТАХ У ЗАЛІЗОРУДНОМУ КАР'ЄРІ

Анотація. Відкрита розробка родовищ супроводжується викидами у повітря шкідливих фракцій пилу, що призводить до порушення екології навколишнього середовища і підвищує ризик хронічних захворювань. У статті проведена оцінка масової концентрації фракції пилу в газопиловій хмарі на динамічній стадії її формування після масового вибуху у залізорудному кар'єрі. Встановлені діапазони зміни висоти підйому і концентрації фракції пилу. Наведена залежність вмісту фракцій пилу від висоти їх підйому та отримано рівняння регресії.

Ключові слова: кар'єр, вибухова речовина, гірська порода, газопилова хмара, динамічна стадія, фракція пилу, концентрація, висота підйому.

Постановка проблеми. Відкрита розробка родовищ супроводжується виділенням в повітря шкідливих фракцій пилу та газів. Виділення пилу відбувається при роботі кар'єрного транспорту, бурових та вибухових роботах. Техногенне забруднення повітря твердими частинками пилу діаметрами від 0,001 до 2,5 мкм (PM 2.5) та від 2,5 до 10 мкм (PM 10) є основним фактором ризику хронічних захворювань. Середньодобова концентрація цих частинок у повітрі повинна становити не більше ніж 0,025 мг/м³ (PM 2.5) та 0,05 мг/м³ (PM 10) [1].

При масових вибухах у кар'єрах відбувається викид твердих та газоподібних продуктів детонації, включаючи частинки зруйнованих гірських порід, що знаходяться поруч із зарядом вибухової речовини (ВР). Відбувається формування газопилової хмари, в якій концентрація пилу може досягати більше 4000 мг/м³ (для міцних гірських порід). Фракції пилу з розмірами до 10 мкм можуть досягати висоти до 2 км і переносяться на відстань до 10 км і більше. Шкідливі речовини, що виділяються під час масових вибухів, забруднюють атмосферу кар'єрів та прилеглих до них територій. Зокрема встановлено, що в умовах міста Кривий Ріг запиленість повітря на межі санітарно-захисних зон може перевищувати значення гранично-допустимих концентрацій у 17-200 разів. При цьому відсоток перевищення рівнів гранично допустимих концентрацій (разове значення 0,5 мг/м³) на відстанях від 4 до 19 км від масових вибухів зазвичай становить 100% [2]. Тому актуальним питанням є дослідження процесів зміни концентрації фракцій пилу в газопиловій хмарі та їх розповсюдження за межі кар'єру.

Аналіз останніх досліджень і публікацій. У роботах [2-6] та інших розглянуто процес формування та динаміка розповсюдження частинок пилу після масових вибухів

у кар'єрах. Зокрема у роботі [2] встановлено, що у газопиловій хмарі фракції пилу з розмірами від 1 до 10 мкм складають більше 30% від загальної маси. У роботі [3] отримано співвідношення для визначення часу та висоти підйому сферичних частинок пилу діаметром до 100 мкм на динамічній стадії формування хмари. Роботи [7, 8] присвячені моделюванню забруднення кар'єрного простору пилом при масових вибухах. Зокрема у роботі [7] встановлено, що за відсутності вітру швидкість міграції частинок пилу досить низька. При цьому фракції пилу з розмірами до 40 мкм осідають досить повільно. У роботі [8] досліджено схеми розсіювання газопилових хмар при проведенні вибухових робіт у глибокому кар'єрі. Встановлено, що швидкість вітру, його напрямок та положення джерела вибуху впливають на схему дифузії та концентрацію частинок пилу. У роботі [9] проведено аналіз складу пилової хмари при масовому вибуху у вапняковому кар'єрі «Taffs Well» (Уельс). Встановлено, що червонувато-сірі хмари складаються переважно з частинок доломітової вапнякової породи, а більш світліші містять частинки вибухового згоряння. У роботі [10] наведено математичну модель руху частинок пилу після детонації заряду ВР. Здійснено моделювання форми газопилових хмар у стані ударного, грибоподібного та дифузійного руху. Також реалізовано моделювання руху частинок пилу у реальному часі.

Мета роботи – оцінка середньої концентрації фракції пилу в газопиловій хмарі на динамічній стадії її формування після масового вибуху у залізорудному кар'єрі.

Викладення основного матеріалу дослідження. При масових вибухах у кар'єрах формування газопилової хмари відбувається у дві основні стадії: динамічна (до 1 с) та теплова (до 60 с) [3].

На першій стадії після утворення тріщин у блоці гірських порід під впливом вибуху відбувається інтенсивний вихід в атмосферу продуктів детонації. Висхідний рух частинок пилу та шматків зруйнованої гірської породи (включаючи продукти детонації) відбувається під дією динамічного імпульсу. На другій стадії на частинки пилу діє підйомна сила, яка викликана різницею температур зовні і всередині газопилової хмари (ежекційні потоки повітря). В даному випадку висота підйому хмари буде значно вище. При цьому у верхній частині хмари (наприкінці стадії) будуть міститися фракції пилу переважно до 10 мкм.

На рис. 1 наведено вигляд газопилової хмари на динамічній стадії її формування у залізорудному кар'єрі та схема фракційного розподілу частинок пилу за висотою підйому.

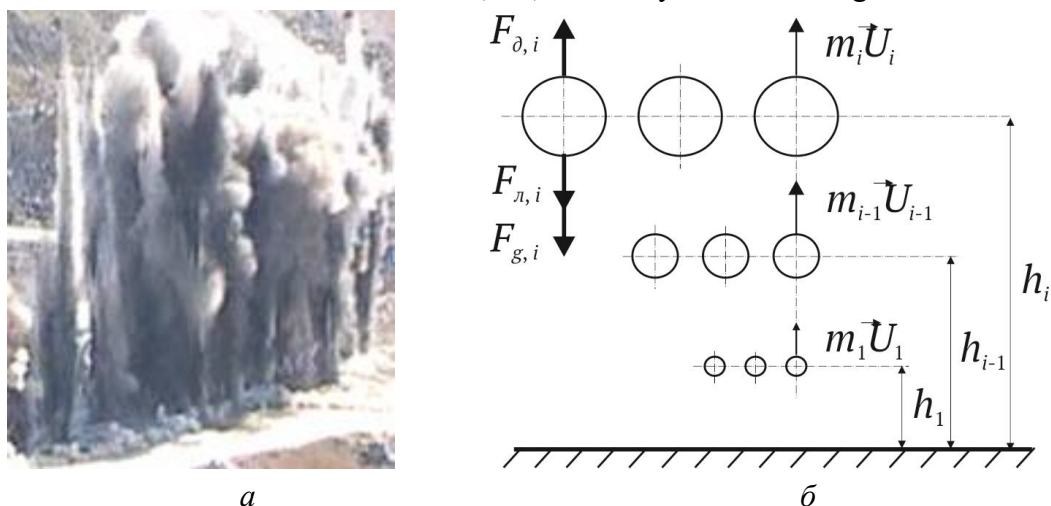


Рисунок 1 – Газопилова хмара на динамічній стадії її формування (а) та схема фракційного розподілу частинок пилу за висотою підйому (б): h_i – висота підйому фракції пилу, м; m_i – маса фракції, кг; \vec{U}_i – вектор швидкості фракції пилу, м/с; $F_{\Delta,i}$ – сила інерції, Н; $F_{L,i}$ – сила лобового опору, Н; $F_{g,i}$ – сила тяжіння, Н

Відповідно до рис. 1, б висхідний рух сферичних частинок пилу під дією сили інерції описується рівнянням

$$\frac{dU_i}{dt_i} = -g - \xi_i \frac{\rho_{zc} U_i^2}{2} \frac{\pi d_i^2}{4} \frac{1}{m_i}, \quad (1)$$

де ρ_{zc} – щільність газової суміші у хмарі, кг/м³; ξ_i – коефіцієнт лобового опору фракції пилу; t_i – час підйому фракції пилу, с; g – прискорення вільного падіння, м/с².

На динамічній стадії енергія вибуху надає сферичним частинкам пилу динамічний імпульс

$$p_i = m_i U_i \approx 0,167 \pi d_i^3 \rho U_i, \quad (2)$$

де ρ – щільність гірської породи, кг/м³; d_i – діаметр частинки пилу, м.

Прийmemo припущення про відсутність впливу температурного фактору (ежекційні потоки повітря у хмарі). Тоді можна казати, що частинки пилу з більшою масою і однаковою щільністю матимуть більший імпульс і відповідно висоту підйому на динамічній стадії. Таким чином буде спостерігатися фракційний поділ частинок пилу.

У роботі [3] були прийняті припущення про сферичність частинок пилу та їх обтікання газовим потоком згідно із законом Стокса. З рівняння (2) було отримано формулу для визначення висоти підйому частинок пилу діаметром до 100 мкм під дією динамічного імпульсу

$$h_i = \frac{u_r \rho d_i^2}{18 \mu_{zc}} \ln \left(\frac{g \rho d_i^2}{g \rho d_i^2 + 18 u_r \mu_{zc}} \right), \quad (3)$$

де u_r – швидкість виходу залишкових продуктів детонації, м/с; μ_{zc} – динамічна в'язкість газової суміші, Па·с.

Викид газу та пилу у продуктах детонації свердловинних зарядів відбувається приблизно з однаковою швидкістю, тобто $U_i \approx u_r$.

Для визначення середньої концентрації пилу C_{cp} (мг/м³) у газопиловій хмарі скористаємося емпіричною формулою

$$C_{cp} = 3,25q_y N^{1,68}, \quad (4)$$

де q_y – питома кількість пилу, яка виділяється під час вибуху 1 м³ гірської маси, кг/м³; N – сумарна потужність заряду ВР, т.

Концентрація фракції пилу у газопиловій хмарі визначається за формулою

$$C_i \approx n_i C_{cp}, \quad (5)$$

де n_i – процентний вміст фракцій.

Прийmemo наступні вихідні дані: щільність матеріалу частинок пилу – 3440 кг/м³ (залізна руда); ВР – «Анемікс»; потужність (загальна маса) заряду ВР – 300 т; насипна щільність набійки – 1800 кг/м³; діаметр свердловини – 0,25 м; довжина заряду ВР та набійки – відповідно 11 та 7 м; глибина кар'єру – 295 м; температура атмосферного повітря – 272 К; температура газоподібних продуктів детонації у хмарі – 555 К; вміст оксидів азоту та монооксиду вуглецю – відповідно 2,9 і 248 мг/м³; питома кількість пилу при масовому вибуху – 0,030 кг/м³.

Крім того, скористаємося результатами експериментальних досліджень вмісту фракції пилу у газопилових хмарах після масових вибухів у кар'єрах Інгулецького гірничо-збагачувального комбінату (м. Кривий Ріг) [2] (табл. 1).

Таблиця 1

Вміст фракцій пилу у газопиловій хмарі після масового вибуху

Середній діаметр фракцій $d_{cp,i}$, мкм	1 (0÷1,4)	3 (1,4÷4,2)	7 (4,2÷10)	13 (10÷15)	23 (15÷30)	38 (30÷45)	73 (45÷100)
Процентний вміст фракцій n_i , %	10,67	11,24	12,46	14,15	15,88	17,37	18,24

Розрахунки за формулою (3) показують, що висота підйому фракцій пилу із середнім діаметром від 1 до 73 мкм змінюється у діапазоні відповідно від 0,11 до 26,211 м. При цьому час підйому фракцій пилу прямо пропорційний їх масі (або діаметру).

З формули (4) отримаємо, що середня концентрація пилу у хмарі дорівнює 1367 мг/м³.

Розрахунки за формулою (5) показують, що для заданого діапазону зміни діаметрів фракцій пилу величина їх концентрації змінюється у діапазоні відповідно від 145,86 до 249,34 мг/м³. Зростання концентрації при збільшенні діаметру фракції пилу пояснюється фізико-механічними властивостями гірської породи і відповідно особливостями її дроблення при детонації свердловинних зарядів ВР. У свою чергу процес дроблення залежить від потужності та форми заряду.

На рис. 2 наведено залежність концентрації фракцій пилу від висоти їх підйому на динамічній стадії формування газопилової хмари.

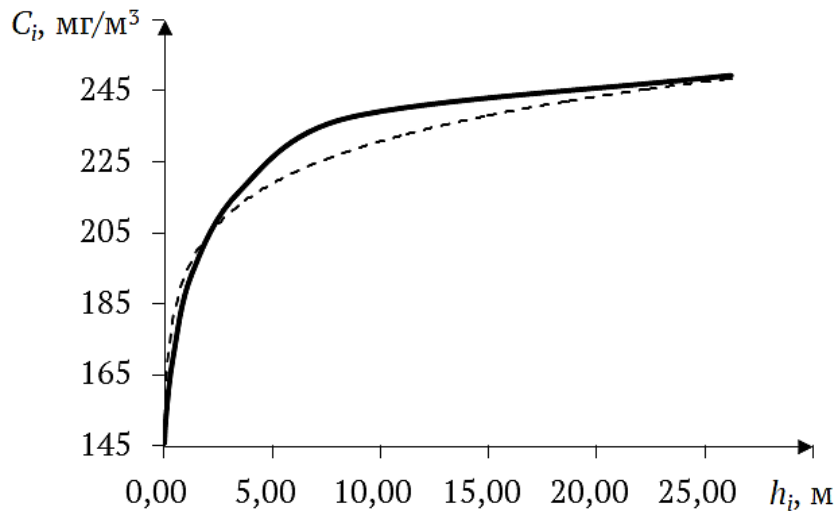


Рисунок 2 – Залежність концентрації фракцій пилу від висоти їх підйому:
----- - крива регресії

Аналіз рис. 2 показує, що найбільш виражене збільшення концентрації фракцій пилу спостерігається в діапазоні висот їх підйому від 0,11 до 6 м. Згідно з формулою (3) це відповідає фракціям пилу із середнім діаметром від 1 до 31 мкм. Надалі спостерігається зниження інтенсивності зростання концентрації ($31 < d_i \leq 73$ мкм).

Крива регресії на рис. 2 описується статичною залежністю з коефіцієнтом детермінації $R^2 = 0,9556$

$$C_i = 193,53 h_i^{0,0765}. \quad (4)$$

Висновки. На динамічній стадії формування газопилової хмари висота підйому частинок пилу із середнім діаметром від 1 до 73 мкм змінюється у діапазоні від 0,11 до 26,211 м, а величина їх концентрації – у діапазоні від 145,86 до 249,34 мг/м³. Вміст фракцій пилу прямо пропорційний їх середньому діаметру. Це пояснюється фізико-механічними властивостями гірської породи і відповідно особливостями її дроблення енергією вибуху. Для вказаних вихідних даних найбільш виражене збільшення концентрації спостерігається для фракцій пилу діаметром від 1 до 31 мкм на висотах їх підйому від 0,11 до 6 м.

ЛІТЕРАТУРА

1. Babii K.V. Studying the impact of single blasts in open pits on the Air dustiness in the city of Krivyi Rih / K.V. Babii, L.A. Novikov, M.O. Batur // Geo-Technical Mechanics: Interdepartmental collection of scientific papers / Institute of Geotechnical Mechanics named by N. Poliakov of NAS of Ukraine. – Dnipro, 2022. – no. 160. – P. 33-43. – URL: <https://doi.org/10.15407/geotm2022.160.033>.
2. Способи і засоби підвищення екологічної безпеки масових вибухів в залізорудних кар'єрах за пиловим чинником / В.Е. Колесник, А.А. Юрченко, А.А. Литвиненко, А.В. Павличенко. – Дніпропетровськ: Літограф, 2014. – 112 с.
3. Novikov L.A. Determination of the lifting height of dust particles after a mass explosion in an iron ore open pit / L.A. Novikov, K.S. Ishchenko, L.O. Lohvyna // Geo-Technical Mechanics: Interdepartmental collection of scientific papers / Institute of Geotechnical

Mechanics named by N. Poliakov of NAS of Ukraine. – Dnipro, 2023. – no. 164. – P. 126-134. – URL: <https://doi.org/10.15407/geotm2023.164.126>.

4. Exploratory investigation on the suppression efficiency of fugitive dust emitted from coal stockpile: Comparison of innovative atomizing and traditional spraying technologies / P. Yen, C. Yuan, J. Ceng, K. Chiang, Y. Tseng, K. Soong, M. Jeng // *Process Safety and Environmental Protection*. – vol. 154. – P. 348-359. –

URL: <https://doi.org/10.1016/j.psep.2021.08.026>.

5. Юрченко А.А. Формування та розповсюдження пилової хмари після масового вибуху в залізорудному кар'єрі / А.А. Юрченко, Д.В. Кулікова // *Збірник наукових праць НГУ*. – 2018. – № 56. – С. 253-264. – URL: <https://ir.nmu.org.ua/handle/123456789/154056>.

6. Гурин А.А. О подъеме пылегазового облака под метательным воздействием взрывчатых веществ / А.А. Гурин, В.И. Деньгуб, Т.В. Деньгуб // *Гірничий вісник*. – 2016. – Вип. 101. – С. 162-164. – URL: http://nbuv.gov.ua/UJRN/girvi_2016_101_37.

7. Numerical simulation of blasting dust pollution in open-pit mines / Z. Huang, S. Ge, D. Jing, L. Yang // *Applied ecology and environmental research*. – 2019. – vol. 17. – Issue 5. – P. 10313-10333. – URL: http://dx.doi.org/10.15666/aer/1705_1031310333.

8. Dispersion Pattern of Blasting Dust in Deep Open-Pit Mines / Z. Chen, C. Du, J. Wang, Y. Wang // *ACS Omega*. – 2023. – vol. 8. – Issue 34. – P. 31353–31364. – URL: <https://doi.org/10.1021/acsomega.3c03528>.

9. Jones T. Richards. Primary blasting in a limestone quarry: Physicochemical characterization of the dust clouds / T. Jones, A. Morgan, R. Richards // *Mineralogical Magazine*. – 2003. – vol. 67. – Issue 2. – P. 153-162. –

URL: <https://doi.org/10.1180/0026461036720092>.

10. Jing Z. Simulated study of the blasting smog-dust movement on account of particle system / Z. Jing // *Journal of University of Science and Technology Beijing*. – 2007. – vol. 67. – Issue S2. – P. 25-29. – URL: <http://doi.org/10.13374/j.issn1001-053x.2007.s2.077>.

REFERENCES

1. Babii K.V. Studying the impact of single blasts in open pits on the Air dustiness in the city of Krivyyi Rih / K.V. Babii, L.A. Novikov, M.O. Batur // *Geo-Technical Mechanics: Interdepartmental collection of scientific papers / Institute of Geotechnical Mechanics named by N. Poliakov of NAS of Ukraine*. – Dnipro, 2022. – no. 160. – P. 33-43. – URL: <https://doi.org/10.15407/geotm2022.160.033>.

2. Sposoby i zasoby pidvyshchennia ekolohichnoi bezpeky masovykh vybukhiv v zalizorudnykh karierakh za pylovym chynnykom / V.E. Kolesnik, A.A. Jurchenko, A.A. Litvinenko, A.V. Pavlichenko. – Dnipropetrovsk: Litohraf, 2014. – 112 p.

3. Novikov L.A. Determination of the lifting height of dust particles after a mass explosion in an iron ore open pit / L.A. Novikov, K.S. Ishchenko, L.O. Lohvyna // *Geo-Technical Mechanics: Interdepartmental collection of scientific papers / Institute of Geotechnical Mechanics named by N. Poliakov of NAS of Ukraine*. – Dnipro, 2023. – no. 164. – P. 126-134. – URL: <https://doi.org/10.15407/geotm2023.164.126>.

4. Exploratory investigation on the suppression efficiency of fugitive dust emitted from coal stockpile: Comparison of innovative atomizing and traditional spraying technologies / P. Yen, C. Yuan, J. Ceng, K. Chiang, Y. Tseng, K. Soong, M. Jeng // *Process Safety and Environmental Protection*. – vol. 154. – P. 348-359. –
URL: <https://doi.org/10.1016/j.psep.2021.08.026>.
5. Yurchenko A.A. Formuvannia ta rozpovsiudzhennia pylovoi khmary pislia masovoho vybukhu v zalizorudnomu karieri / A.A. Yurchenko, D.V. Kulikova // *Zbirnyk naukovykh prats NHU*. – 2018. – no. 56. – P. 253-264. –
URL: <https://ir.nmu.org.ua/handle/123456789/154056>.
6. Gurin A.A. O pod"eme pylegazovogo oblaka pod metatel'nym vozdeystviem vzryvchatykh veshchestv / A.A. Gurin, V.I. Den'gub, T.V. Den'gub // *Hirnychiy visnyk*. – 2016. – issue 101. – P. 162-164. – URL: http://nbuv.gov.ua/UJRN/girvi_2016_101_37.
7. Numerical simulation of blasting dust pollution in open-pit mines / Z. Huang, S. Ge, D. Jing, L. Yang // *Applied ecology and environmental research*. – 2019. – vol. 17. – Issue 5. – P. 10313-10333. – URL: http://dx.doi.org/10.15666/aer/1705_1031310333.
8. Dispersion Pattern of Blasting Dust in Deep Open-Pit Mines / Z. Chen, C. Du, J. Wang, Y. Wang // *ACS Omega*. – 2023. – vol. 8. – Issue 34. – P. 31353–31364. –
URL: <https://doi.org/10.1021/acsomega.3c03528>.
9. Jones T. Richards. Primary blasting in a limestone quarry: Physicochemical characterization of the dust clouds / T. Jones, A. Morgan, R. Richards // *Mineralogical Magazine*. – 2003. – vol. 67. – Issue 2. – P. 153-162. –
URL: <https://doi.org/10.1180/0026461036720092>.
10. Jing Z. Simulated study of the blasting smog-dust movement on account of particle system / Z. Jing // *Journal of University of Science and Technology Beijing*. – 2007. – vol. 67. – Issue S2. – P. 25-29. – URL: <http://doi.org/10.13374/j.issn1001-053x.2007.s2.077>.

Received 15.09.2024.
Accepted 18.09.2024.

Changes in dust concentration in a gas-dust cloud during blasting in an iron ore pit

In open-pit extraction of minerals emit large amounts of harmful dust fractions and gas into the atmosphere. This leads to atmospheric pollution in open pit and sanitary protection zones. The largest dust emissions are observed during blasting operations in open pits. This is especially the cases when unfavorable meteorological conditions occur, means of dust suppression are not sufficiently effective, blasting technology was violated. Dust particles are known to be able to propagate long distances from the source of the explosion. Dust fractions up to 10 μm diameter are the most dangerous, as they are the main risk factor for the development of chronic respiratory diseases. In particular, according to European standards, the average daily concentration of dust particles with a diameter between 2.5 and 10 μm should not exceed 0.05 mg/m^3 .

The paper estimates mass concentrations of dust fractions in the gas-dust cloud at the dynamic stage of its formation after a mass explosion in an iron ore open pit. The following assumptions were made in the calculations: no influence of the temperature factor (ejected air flows); dust particles are spherical; flow of dust fractions by the gas flow follows the Stokes law. The ranges of variation of ascent height and mass concentrations of dust fractions

at the dynamic stage of gas-dust cloud formation are established. In particular, the lift height of the dust fraction is directly proportional to the value of the dynamic impulse. The difference in mass concentrations of dust fractions in the gas-dust cloud is due to the physical and mechanical properties of rocks destroyed by boreholes explosive substance charge. When determining the mass concentrations of dust fractions, the results of experimental studies were used, which were carried out in the open pit of Ingulec Mining and Processing Combine. The graphical dependence of average mass concentrations of dust fractions depending on the height of their ascent at the dynamic stage of gas-dust cloud formation is presented, the regression equation in the form of a degree function is obtained. The results of calculations have shown that the greatest intensity of mass concentration increase in the gas-dust cloud at the dynamic stage of its formation is characteristic of dust fractions with diameter up to 31 μm . The intensity of growth of mass concentrations of larger dust fractions decreases markedly.

Яворська Олена Олександрівна – д.т.н., доцент, професор кафедри охорони праці та цивільної безпеки Національного технічного університету «Дніпровська політехніка».

Новіков Леонід Андрійович – к.т.н., науковий співробітник відділу геомеханічних основ відкритої розробки родовищ Інституту геотехнічної механіки ім. М.С. Полякова НАН України.

Іконніков Максим Юрійови – к.т.н., доцент, професор кафедри охорони праці та цивільної безпеки Національного технічного університету «Дніпровська політехніка».

Yavorska Olena Oleksandrivna, Doctor of Technical Sciences (Dr. habil. Tech.), Professor of the Department of Labour Protection and Civil Safety, Dnipro University of Technology.

Novikov Leonid Andriiovych, Candidate of Technical Sciences (Ph.D. Tech), Researcher in Department of Geomechanics of Mineral Opencast Mining Technology, the M.S. Poliakov Institute of Geotechnical Mechanics of the NAS of Ukraine.

Ikonnikov Maksym Yuriyovych, Candidate of Technical Sciences (Ph.D. Tech), Professor of the Department of Labour Protection and Civil Safety, Dnipro University of Technology

ІНТЕГРОВАНЕ НЕЙРОМЕРЕЖЕВЕ МОДЕЛЮВАННЯ ЕЕГ ДЛЯ ДІАГНОСТИКИ ПОРУШЕНЬ МОЗКОВОЇ ДІЯЛЬНОСТІ

Анотація. У статті розглядається структуроване багатоступеневе моделювання ЕЕГ засобами прикладної математики з налаштування простору вхідних параметрів для нейромережевого прогнозування. Також, проводиться аналіз підходів та моделей на їх точність у визначенні відповідних ознак сигналу та пристосованості до реальних даних.

Ключові слова: нейромережа, електроенцефалографія, часовий ряд, прогнозування, функція активації, хаотична динаміка.

Постановка проблеми. Прогрес розвитку інтелектуальних здібностей, формування мозку, як складної структури взаємозалежних нейронів стає стрімкішим з кожним століттям, що межує із поліпшенням всіх аспектів життєдіяльності. Припускаючи еволюцію геному людини як безпосередній фактор формування мозкової структури, виникає необхідність у створенні категоричного та гнучкого інструменту аналізу взаємозалежної поведінки нейронів. Наразі, існує велика бібліотека досліджень, які частково або повністю описують та пропонують алгоритми подібного аналізу [1,2,3], однак не мають достатньої пристосованості до реальних подій або обмежують розмах точності до необґрунтовано значних меж. Тому, в даній роботі пропонується до огляду більш гнучкий метод аналізу та прогнозування відповідних даних електроенцефалограми (ЕЕГ).

Мета дослідження. Розробити ряд методів та підходів для гнучкого аналізу ЕЕГ з використанням математичного моделювання, як для попереднього відтворення зв'язків та ідентифікації глобальних параметрів, так і для налаштування параметрів нейромережі, для більш точного прогнозування.

Огляд проблеми. Електроенцефалографія дозволяє неінвазивно реєструвати електричну активність мозку, надаючи цінну інформацію про його стан і функції. Цей метод широко застосовується для діагностики та аналізу неврологічних розладів, таких як епілепсія, шизофренія та аутизм.

Важливим етапом в обробці ЕЕГ є виділення релевантних ознак або маркерів для конкретного застосування. Для цього було розроблено безліч методів, від класичного спектрального аналізу до нелінійних підходів і сучасних алгоритмів глибокого навчання. Оскільки мозок функціонує як складна мережа нейронів, що взаємодіють і синхронізуються, то використання метрик функціональної зв'язності як ознаки для аналізу

ЕЕГ-сигналів, пошук параметрів поведінки моделі та використання адаптивних, гнучких підходів з використанням нейромережевого моделювання залишається актуальним.

Опис наданих даних ЕЕГ. В роботі були використані дані пацієнта з ознаками порушення мозкової активності, однак без епілептичних нападів та в порівнянні використано базу даних з дитячої лікарні в Бостоні, пацієнтів з важко виліковними судомами. А саме:

Пацієнт 1. 8 років, хлопчик, тимчасові аномалії – 20.06.2024

Пацієнт 2. 11 років, дівчина, часті судомні напади – 08.06.2010

Алгоритм попередньої обробки ЕЕГ сигналів. Попередня обробка ЕЕГ є відповідальним етапом у підготовці її до подальшого аналізу. Це включає в себе ряд методів, спрямованих на зниження шуму і видалення артефактів, щоб забезпечити готовність чистого сигналу до подальших кроків. Початковим етапом цього процесу є видалення шумів, що походять від зовнішніх електромагнітних полів. Далі необхідно усунути артефакти руху, оскільки вони можуть мати негативний вплив на сигнал ЕЕГ. Одним з найбільш популярних інструментів для видалення артефактів і шумів з нелінійних сигналів ЕЕГ є метод незалежних компонент (Independent Component Analysis, ICA) [4]. Завдяки ICA вдається ефективно розкласти сигнал на незалежні компоненти і ідентифікувати ті з них, що пов'язані з артефактами відповідної ознаки. При роботі з багатоканалними сигналами ICA розділяє багатокомпонентний сигнал ЕЕГ на незалежні частини, тим самим видаляючи шум і перешкоди, викликані морганням, рухами очей, серцевими скороченнями і м'язовою активністю.

Наступним кроком було застосовано ряд фільтрів різного типу, що включає застосування смугового фільтра для виділення частотного діапазону, який містить релевантну інформацію, та режекторного фільтра для зниження впливу електромагнітних перешкод від мережі живлення [5]. Параметри фільтрів можуть бути гнучко налаштовані відповідно до вимог конкретного дослідження та змінені відповідно спостережуваним висновкам про вихідний сигнал. В даній роботі смуговий фільтр було встановлено в діапазоні 0,5-30 Гц, режекторний 50 Гц з вікном розмаху 5 Гц. Отримані результати візуалізовано порівнянням оригінальних даних з обробленими (рис. 1-2).

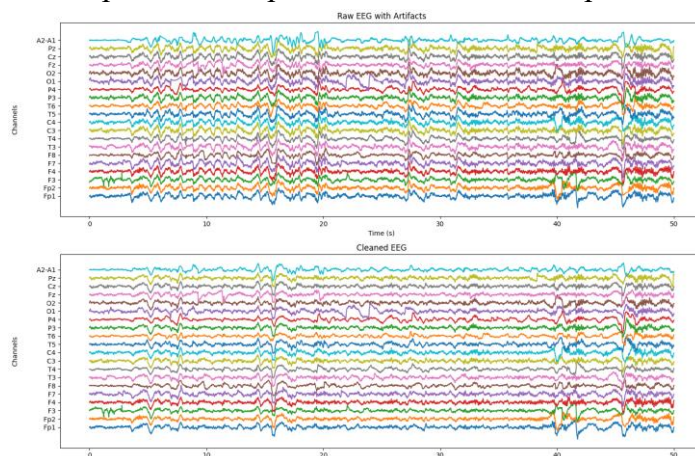


Рисунок 1 – Порівняння вхідних даних з очищеними від артефактів для першого пацієнта

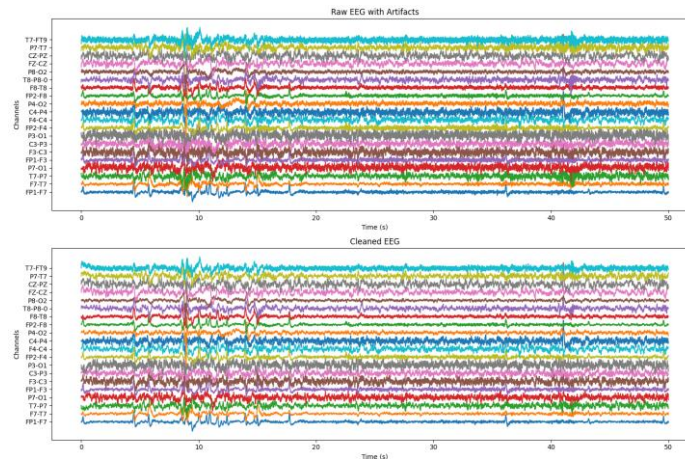


Рисунок 2 – Порівняння вхідних даних з очищеними від артефактів для другого пацієнта

Для виявлення інших типів артефактів застосовується метод на основі амплітудного порогу. Хоча цей підхід є відносно простим, він дозволяє виявляти значні відхилення у сигналі, проте його можливості можуть бути розширені шляхом застосування більш складних алгоритмів, таких як класифікація артефактів з використанням машинного навчання.

Для подальшої обробки даних, сигнал було розділено на сегменти однакової довжини та проаналізовано частотний вміст сигналу.

Спектральний аналіз є широко використовуваним методом вилучення цінної інформації з сигналів ЕЕГ [6]. Аналізуючи спектральну щільність потужності (спектр потужності) сигналу, спектральний аналіз може дати уявлення про частотний склад або розподіл потужності сигналу по частоті. Ця інформація допомагає зрозуміти основні нейронні процеси, відповідальні за генерацію сигналу, і виявити закономірності, які можуть бути пов'язані з конкретними когнітивними станами або поведінкою

З порівняння показників спектру було визначено, що пацієнт з активним та прогресуючим захворюванням мозку має більш виражені піки активності в частоті 15-20 Гц та близько 30 Гц, на відміну від іншого пацієнта з більш гладкою поведінкою частоти, яка демонструє найбільшу потужність сигналу в низькочастотному діапазоні та різкий спад після 40 Гц. Для візуалізації даних та оцінки якості їх обробки було використано інструменти з бібліотеки Matplotlib (рис. 3-4).

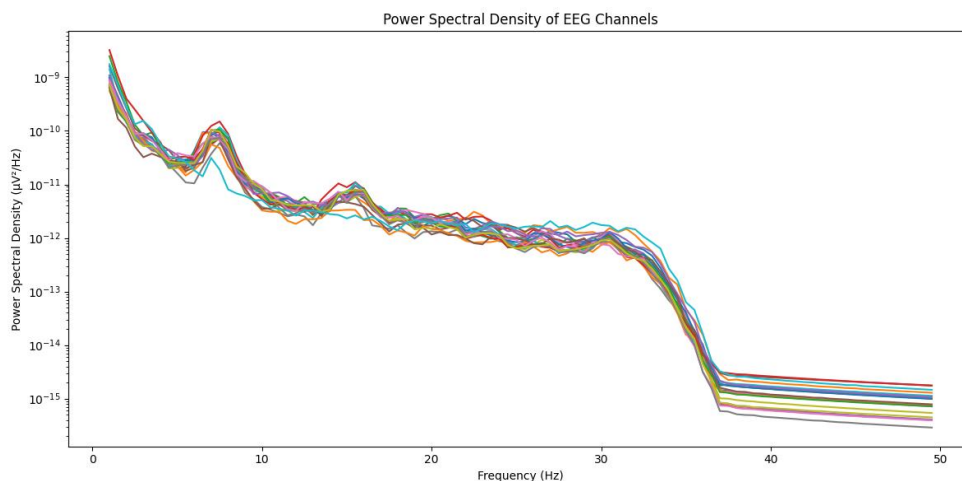


Рисунок 3 – Візуалізація спектральної потужності першого пацієнта

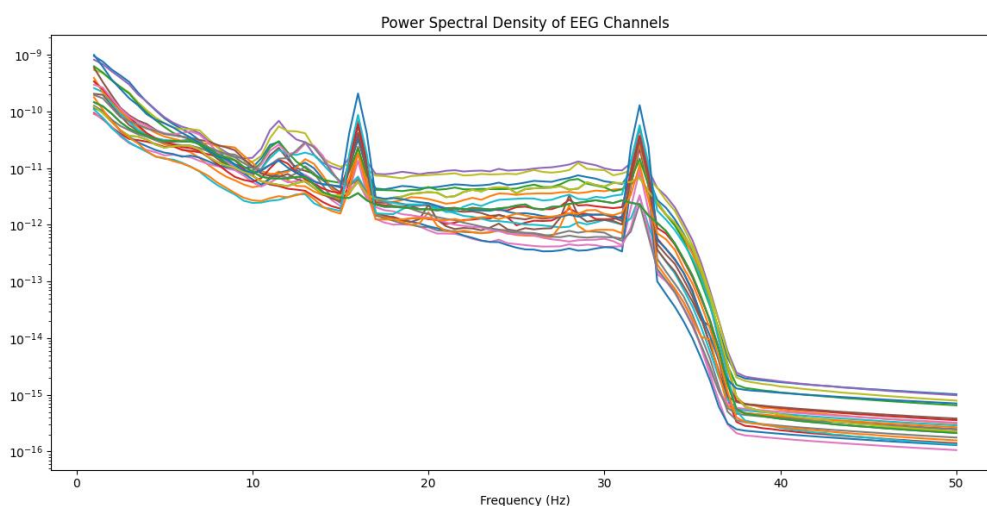


Рисунок 4 – Візуалізація спектральної потужності другого пацієнта

Варто зазначити, що алгоритм обробки може бути вдосконалений шляхом впровадження адаптивної фільтрації для більш ефективного видалення артефактів, а також за рахунок застосування методів машинного навчання для автоматичної класифікації артефактів. Крім того, перспективним є додавання аналізу зв'язності між каналами ЕЕГ і автоматичний вибір параметрів фільтрації на основі характеристик сигналу.

Незважаючи на відсутність консенсусу щодо оптимального алгоритму попередньої фільтрації та методів видалення артефактів, поточні дослідження продовжують підвищувати якість даних ЕЕГ та підвищувати точність аналізу. Використання різних методів, може ефективно пом'якшити артефакти в сигналах ЕЕГ, забезпечуючи всебічне використання даних в широкому діапазоні.

Використання адаптивних алгоритмів глибокого навчання. В даному розділі розглянемо наступний етап роботи з даними, ціллю якого є інтеграція математичних моделей хаотичної динаміки для пошуку нелінійних зв'язків між сигналами. Для огляду та порівняння були використані різноманітні моделі з їх специфікою пристосова-

ності та визначені показники функції втрат та точності моделювання для висновку щодо їх подальшого застосування.

Слід зазначити, що пошук стабільних станів здорової людини стає задачею менш актуальною та вже отримав певний рівень автоматизації, однак розуміння прояву аномальної активності нейронів, що призводить до так званого нейрозапалення є більш складним процесом.

Складність досліджуваних процесів, що взаємодіють у різних часових масштабах, робить розуміння епілептогенезу великим викликом. У такій системі з декількома нелінійними взаємодіючими процесами математичне моделювання є корисним інструментом для кращого розуміння динаміки системи [7]. Крім того, моделювання допомагає систематизувати та пояснити велику кількість спостережень, отриманих у клінічних дослідженнях

Сигнал ЕЕГ несе цінну інформацію в певних частотних діапазонах: альфа (8-13 Гц), бета (14-40 Гц), тета (4-8 Гц), дельта (0,5-3 Гц), гамма (вище 40 Гц). Кожна з цих частот має свої унікальні характеристики і області застосування. Оскільки частотні діапазони і їх характеристики були широко вивчені, то множину цих значень можна використовувати в якості вхідного вектору для певних моделей.

Представлені математичні моделі були пристосовані для глибокого аналізу динаміки ЕЕГ сигналів, де кожна модель намагається забезпечити унікальний погляд на природу досліджуваних процесів. Авторегресійна модель (AR) дозволяє виявити лінійні залежності в часових рядах ЕЕГ та прогнозувати їх поведінку на основі попередніх значень, що особливо корисно для виявлення повторюваних патернів активності мозку. Прихована марковська модель (НММ) ефективно ідентифікує дискретні стани мозкової активності та переходи між ними, що важливо для розуміння зміни режимів функціонування мозку.

Модель Лоренца демонструє високу ефективність у моделюванні хаотичної динаміки ЕЕГ сигналів, дозволяючи виявити детерміновані хаотичні процеси в мозковій активності. Модель Рьослера, що має простішу структуру порівняно з моделлю Лоренца, але також здатна генерувати хаотичну поведінку, допомагає виявити більш тонкі особливості нелінійної динаміки ЕЕГ. Модель Ван дер Поля, що описує автоколивання, особливо корисна для аналізу ритмічної активності мозку та її автоматичних коливань. Модель ФітцХью-Нагумо, спочатку розроблена для опису нейронної активності, відмінно підходить для моделювання збудження та релаксації нейронних популяцій, що відображаються в ЕЕГ сигналах. Модель Кермака-МакКендріка дозволяє аналізувати поширення активності між різними ділянками мозку за аналогією з епідеміологічними процесами. Модель ФітцХью-Нагумо-Даффінга поєднує властивості нейронної динаміки з нелінійними коливаннями, що робить її особливо цінною для аналізу складних патернів ЕЕГ.

Нижче продемонстроване візуальне порівняння моделей для очищених даних та окремо виділені їх фазові портрети (рис. 5 – 8).

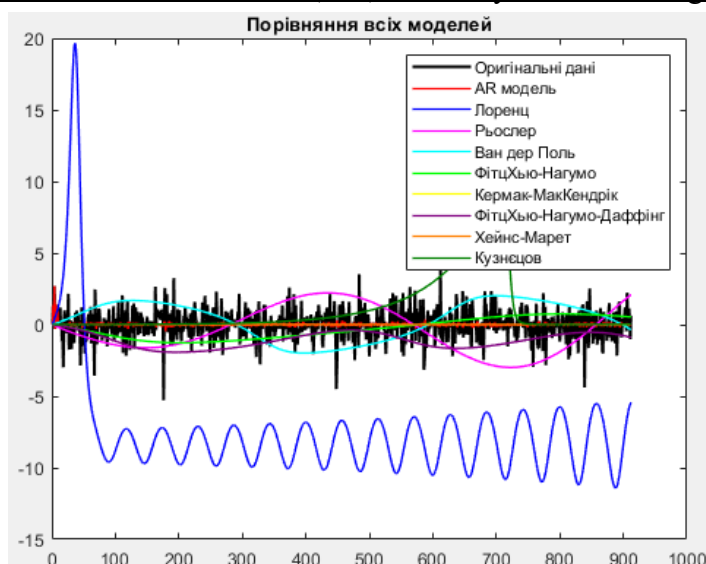


Рисунок 5 – Порівняння результатів моделювання на основі даних першого пацієнта

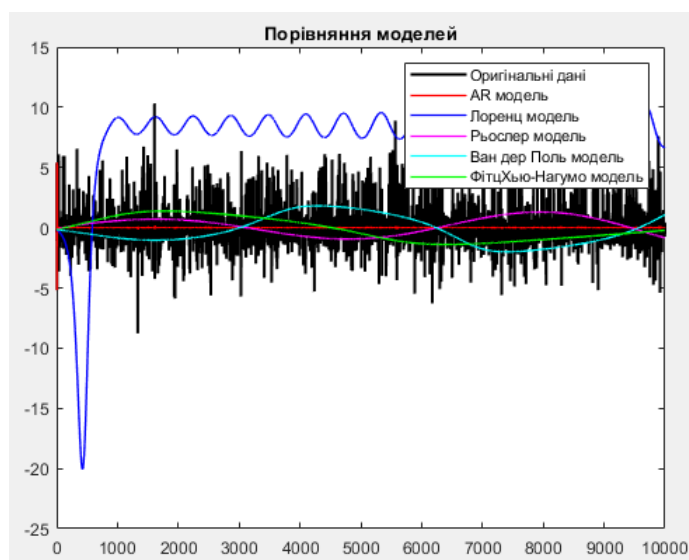


Рисунок 6 - Порівняння результатів моделювання на основі даних другого пацієнта

Аналіз результатів моделювання з використанням певних метрик дозволяє виявити найбільш адекватну модель для конкретного типу ЕЕГ сигналу та краще зрозуміти механізми, що лежать в основі спостережуваної мозкової активності. До таких можемо віднести обчислення показника Ляпунова, які дозволяють оцінити ступінь хаотичності досліджуваних процесів, кореляційна розмірність характеризує складність атрактора та може вказувати на кількість незалежних змінних, необхідних для опису динаміки системи, середньоквадратична помилка (MSE) слугує об'єктивним критерієм для порівняння точності відтворення експериментальних даних [8]. Для прикладу нижче було розраховано середньоквадратичну помилку пацієнтів, яка демонструє перевагу AR моделювання, що було підтверджено візуальною інтерпретацією раніше.

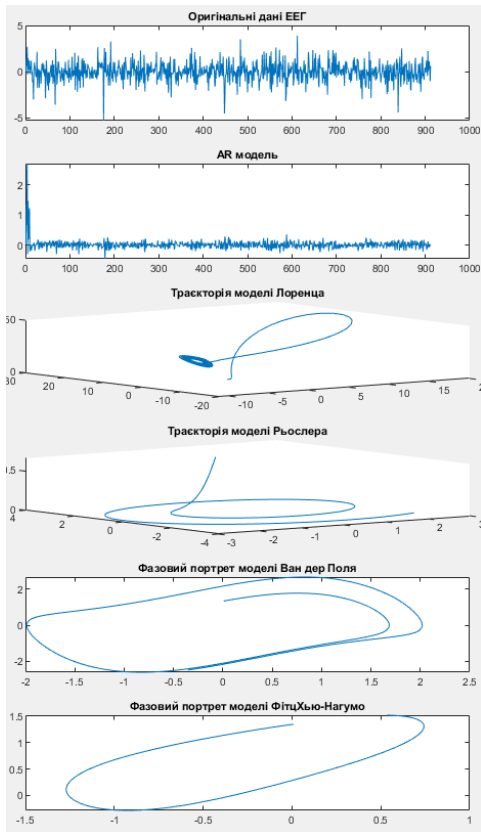


Рисунок 7 – Фазові портрети моделей на основі даних першого пацієнта

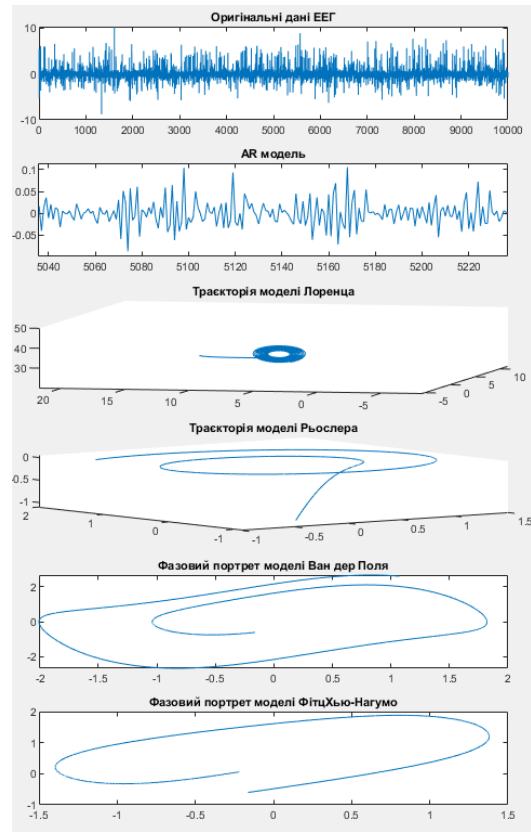


Рисунок 8 - Фазові портрети моделей на основі даних другого пацієнта

Для 1го пацієнта:

AR MSE: 0.974036

Lorenz MSE: 73.473649

Rosler MSE: 3.729282

Van der Pol MSE: 2.779805

FitzHugh-Nagumo MSE: 1.585269

Kermack-McKendrick MSE: 0.998940

FitzHugh-Nagumo-Duffing MSE: 2.511791

Для 2го пацієнта:

AR MSE: 0.990802

Lorenz MSE: 73.568802

Rosler MSE: 1.518555

Van der Pol MSE: 2.368941

FitzHugh-Nagumo MSE: 1.929700

Kermack-McKendrick MSE: 1.036738

FitzHugh-Nagumo-Duffing MSE: 2.212152

Оптимізація прогнозу EEG даних через попереднє налаштування LSTM архітектури. У цьому розділі дослідження запропоновано підхід до аналізу та прогнозування часових рядів електроенцефалограм на основі модифікованої архітектури Long Short-Term Memory (LSTM) [9]. Модель LSTM – це рекурентна нейронна мережа, вдос-

конталена моделлю RNN. LSTM може ефективно передавати та виражати інформацію в більш тривалих часових рядах. Це успішно вирішує проблему градієнтного зникнення RNN, спричинену додаванням мережевих шарів і плином часу, роблячи її більш прийнятною. Унікальний стан нейронних клітин топології нейронної мережі LSTM, яка записує та передає дані за допомогою комірок пам'яті, відрізняє її від інших методів глибокого навчання. Стандартна LSTM архітектура включає вхідний шар, один або декілька прихованих LSTM-шарів, а також вихідний шар. Кожен LSTM-шар містить три типи воріт: вхідні, які контролюють потік нової інформації; ворота забування, що визначають, яку інформацію зберігати або відкидати; та вихідні ворота, що вирішують, яку інформацію передавати далі по мережі. Така структура дозволяє ефективно працювати з довгими послідовностями, зберігаючи релевантну інформацію і ігноруючи непотрібні дані.

В більшості подібних досліджень модифікація будови шарів та комбіноване застосування комплексного аналізу не призводило до втішних результатів прогнозування, яке зводилось до стаціонарного стану. Тому, були застосовані попередні процедури визначення параметрів LSTM моделі для досягнення прогнозованими значеннями хаотичної та природньо відповідної поведінки.

По-перше, для ініціалізації вхідних ваг мережі проведено ряд математичних операцій над вхідними даними [10, 11], що надає мережі початкове розуміння структури вхідних даних, прискорюючи збіжність і покращуючи точність прогнозування на ранніх етапах навчання. По-друге, впроваджується нестандартна активаційна функція, що поєднує синусоїдальну компоненту та гаусівський шум, яка надає моделі стохастичності та нерегулярності. Ця функція відрізняється від традиційних активаційних функцій (сигмоїди чи гіперболічного тангенса) тим, що створює більш хаотичну і нелінійну динаміку, що є важливим для моделювання складних біологічних систем, таких як мозкова активність (рис. 9-10).

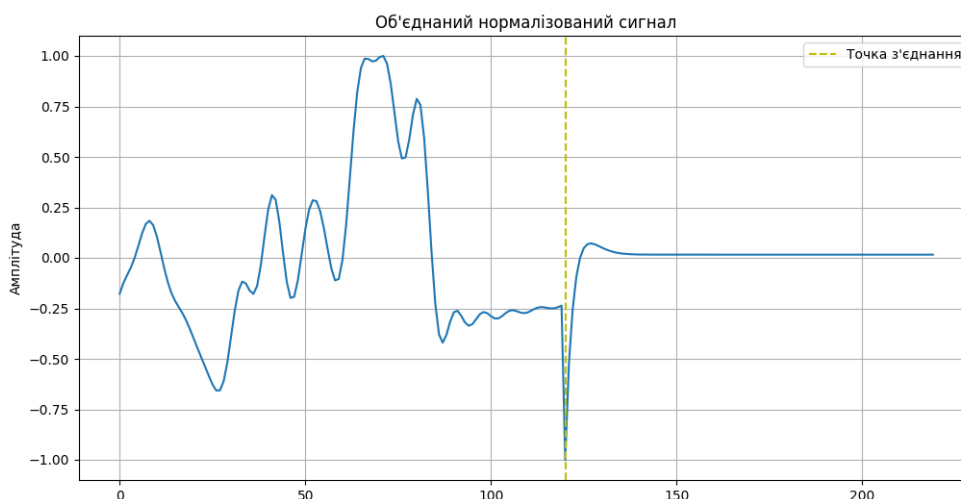


Рисунок 9 – Використання гіперболічного тангенса для прогнозування даних

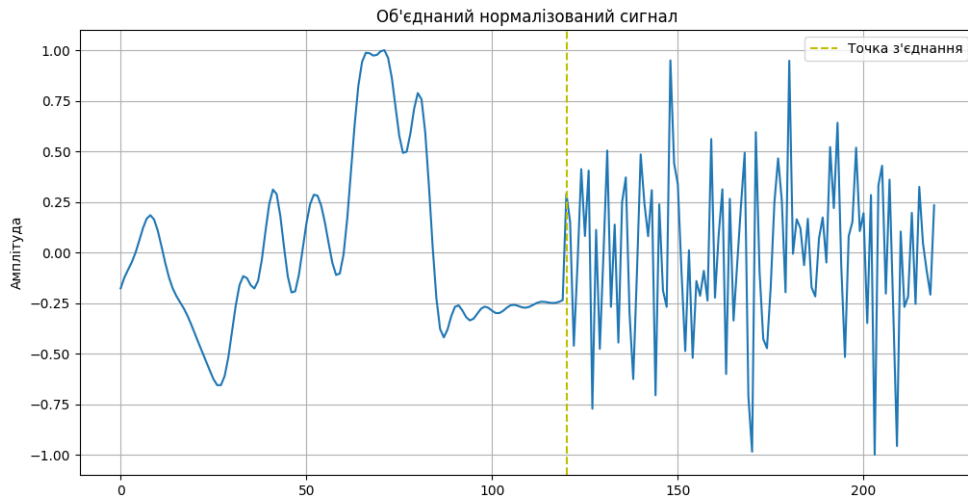


Рисунок 10 – Використання гаусівського шуму з синусоїдним доданком для прогнозування даних

З отриманих результатів можемо зробити висновок, що запропонований підхід дозволив моделі краще відображати складні нелінійні залежності в ЕЕГ-сигналах та підвищило її здатність моделювати хаотичні процеси, що відкриває поле пошуку та потенціал для можливості виявлення аномалій мозку, таких як епілептичні напади або ознаки нейродегенеративних захворювань.

Подальші дослідження можуть включати оптимізацію гіперпараметрів моделі для різних типів ЕЕГ-даних, порівняння з іншими методами виявлення аномалій, інтеграцію з методами попередньої обробки сигналів, а також розширення моделі для аналізу багатоканальних ЕЕГ.

Висновки. 1. Було розглянуто алгоритм попередньої обробки ЕЕГ сигналів з фільтрацією, видалення артефактів, сегментацією та введенням спектрального аналізу. Використання спеціалізованих бібліотек, методу незалежних компонент (ІСА) для видалення артефактів очей, амплітудного порогу для виявлення інших артефактів, та методу Велча для обчислення спектральної щільності потужності забезпечує ефективну та надійну обробку даних для подальшої автоматизації клінічної практики. Алгоритм демонструє здатність виявляти відмінності між пацієнтами з різним станом здоров'я мозку, що підтверджується візуалізацією результатів, і має потенціал для подальшого вдосконалення шляхом впровадження методів машинного навчання.

2. Також, було описано застосування математичного моделювання, зосереджуючись на пошуку нелінійних зв'язків між сигналами за допомогою хаотичної динаміки. Представлений підхід використовує комплекс взаємодоповнюючих моделей, включаючи авторегресійну модель, приховану марковську модель, моделі Лоренца, Рьослера, Ван дер Поля, ФітцХью-Нагумо та інші. Кожна модель надає можливість синтезувати параметри динаміки ЕЕГ сигналів, дозволяючи виявляти різні аспекти мозкової активності. Результати моделювання візуалізуються та порівнюються для двох пацієнтів, включаючи часові ряди та фазові портрети. Оцінка ефективності моделей здійснюється за допомогою середньоквадратичної помилки (MSE). Представлені значення MSE для

різних моделей дозволяють порівняти їх точність у відтворенні експериментальних даних для обох пацієнтів, що має важливе значення для діагностики та лікування неврологічних розладів. Для описаних даних найнижчі показники похибки має авторегресійна модель, яка найточніше передбачає поведінку даних, що стверджує відповідна візуалізація.

3. В останньому розділі було представлено підхід для прогнозування ЕЕГ даних, використовуючи модифіковану архітектуру LSTM. Ключовими особливостями є застосування попередніх процедур для визначення параметрів моделі, ініціалізація вхідних ваг на основі попереднього моделювання сигналу та впровадження нестандартної активаційної функції, що поєднує синусоїдальну компоненту з гаусівським шумом. Ця модифікована архітектура демонструє підвищену здатність моделювати складні нелінійні залежності в ЕЕГ-сигналах, відкриваючи нові можливості для виявлення аномалій мозкової активності. Подальші перспективи включають оптимізацію гіперпараметрів, порівняння з іншими методами та розширення моделі для аналізу багатоканальних ЕЕГ, що потенційно може значно вплинути на діагностику та дослідження неврологічних розладів.

ЛІТЕРАТУРА

1. Rupak Kumar Das., Dale Dowling., Arshia Khan. (2023). A Survey on EEG Data Analysis Software. *Sci*, 5(2):23-23. doi: 10.3390/sci5020023
2. Vicchietti ML, Ramos FM, Betting LE, Campanharo ASLO. Computational methods of EEG signals analysis for Alzheimer's disease classification. *Sci Rep*. 2023 May 20;13(1):8184. doi: 10.1038/s41598-023-32664-8
3. Giovanni Chiarion., Laura Sparacino., Yuri Antonacci., Luca Faes., Luca Mesin. (2023). Connectivity Analysis in EEG Data: A Tutorial Review of the State of the Art and Emerging Trends. *Bioengineering*, doi: 10.3390/bioengineering10030372
4. Seonjoo Lee, Haipeng Shen, Young K. Truong. Nonparametric Independent Component Analysis for the Sources with Mixed Spectra. doi: 10.48550/arxiv.2212.06327
5. Zhong-Ke Gao., Yanli Li., Yu-Xuan Yang., Na Dong., Xiong Yang., Celso Grebogi. (2020). A Coincidence-Filtering-Based Approach for CNNs in EEG-Based Recognition. *IEEE Transactions on Industrial Informatics*, doi: 10.1109/TII.2019.2955447
6. Maliheh Miri., Vahid Abootalebi., Hamid Saeedi-Sourck., Dimitri, Van De Ville., Hamid Behjat. (2023). Spectral Representation of EEG Data using Learned Graphs with Application to Motor Imagery Decoding. *bioRxiv*, doi: 10.1101/2022.08.13.503836
7. Marius Georgescu., Laura Haidar., Alina-Florina Serb., Daniela Puscasiu., D. Georgescu. (2021). Mathematical Modeling of Brain Activity under Specific Auditory Stimulation.. *Computational and Mathematical Methods in Medicine*, doi: 10.1155/2021/6676681
8. Martine, Canada., Ness. Einevoll., Fellin. Panzeri. (2020). Computation of the electroencephalogram (EEG) from network models of point neurons. *bioRxiv*, doi: 10.1101/2020.11.02.364802

9. Raden Aurelius, Andhika. Viadinugroho., Rosadi. Dedi. (2021). Long Short-Term Memory Neural Network Model for Time Series Forecasting: Case Study of Forecasting IHSG during Covid-19 Outbreak. doi: 10.1088/1742-6596/1863/1/012016
10. О.А. Інкін, О.В. Погорєлов. (2024). Моделювання ЕЕГ за допомогою глибоких нейронних мереж. doi:10.34185/1562-9945-3-152-2024-06
11. Belozyorov V. Y., Inkin O. A. (2023). Systems of singular differential equations as the basis for neural network modeling of chaotic processes. Journal of Optimization, Differential Equations and Their Applications, 31(2), 24. <https://doi.org/10.15421/142309>

REFERENCES

1. Rupak Kumar Das., Dale Dowling., Arshia Khan. (2023). A Survey on EEG Data Analysis Software. Sci, 5(2):23-23. doi: 10.3390/sci5020023
2. Vicchiatti ML, Ramos FM, Betting LE, Campanharo ASLO. Computational methods of EEG signals analysis for Alzheimer's disease classification. Sci Rep. 2023 May 20;13(1):8184. doi: 10.1038/s41598-023-32664-8
3. Giovanni Chiarion., Laura Sparacino., Yuri Antonacci., Luca Faes., Luca Mesin. (2023). Connectivity Analysis in EEG Data: A Tutorial Review of the State of the Art and Emerging Trends. Bioengineering, doi: 10.3390/bioengineering10030372
4. Seonjoo Lee, Haipeng Shen, Young K. Truong. Nonparametric Independent Component Analysis for the Sources with Mixed Spectra. doi: 10.48550/arxiv.2212.06327
5. Zhong-Ke Gao., Yanli Li., Yu-Xuan Yang., Na Dong., Xiong Yang., Celso Grebogi. (2020). A Coincidence-Filtering-Based Approach for CNNs in EEG-Based Recognition. IEEE Transactions on Industrial Informatics, doi: 10.1109/TII.2019.2955447
6. Maliheh Miri., Vahid Abootalebi., Hamid Saeedi-Sourck., Dimitri, Van De Ville., Hamid Behjat. (2023). Spectral Representation of EEG Data using Learned Graphs with Application to Motor Imagery Decoding. bioRxiv, doi: 10.1101/2022.08.13.503836
7. Marius Georgescu., Laura Haidar., Alina-Florina Serb., Daniela Puscasiu., D. Georgescu. (2021). Mathematical Modeling of Brain Activity under Specific Auditory Stimulation.. Computational and Mathematical Methods in Medicine, doi: 10.1155/2021/6676681
8. Martine, Canada., Ness. Einevoll., Fellin. Panzeri. (2020). Computation of the electroencephalogram (EEG) from network models of point neurons. bioRxiv, doi: 10.1101/2020.11.02.364802
9. Raden Aurelius, Andhika. Viadinugroho., Rosadi. Dedi. (2021). Long Short-Term Memory Neural Network Model for Time Series Forecasting: Case Study of Forecasting IHSG during Covid-19 Outbreak. doi: 10.1088/1742-6596/1863/1/012016
10. Inkin O.A., Pogorelov O.V. (2024). EEG modeling using deep neural networks. doi:10.34185/1562-9945-3-152-2024-06
11. Belozyorov V. Y., Inkin O. A. (2023). Systems of singular differential equations as the basis for neural network modeling of chaotic processes. Journal of Optimization, Differential Equations and Their Applications, 31(2), 24. <https://doi.org/10.15421/142309>

Received 17.09.2024.
Accepted 21.09.2024.

Integrated neuronetwork modeling of EEG for diagnostic disorders of brain activity

The article discusses the structured multistage modeling of EEG by means of applied mathematics to customize the input parameter space for neural network prediction. Also, the approaches and models are analyzed for their accuracy in determining the relevant signal features and adaptability to real data.

The activity of potentials during brain activity is a biological process that depends on many factors and hides a space of parameters, the search for which and their definition can open us up to a new perspective on the nature and activity of the human brain. A rational way of obtaining data on brain activity is a non-invasive electroencephalogram, which registers the potential difference on the electrodes relative to the base. For further processing of the received data, it is necessary to remove noise and artifacts from them. In this work, a standard algorithm of frequency filtering and filtering from noise caused by the power grid is used. After processing the data, an overview of mathematical models is offered, which with a certain degree of accuracy try to simulate the behavior of the signal or the peak moments of certain features. Added to this is the use of the LSTM model to predict the further behavior of the signal with the preliminary introduction of chaos into the model due to the modified activation function (Gaussian noise) and the input modeling of the weights.

Keywords: neural network, electroencephalography, time series, prediction, activation function, chaotic dynamics.

Білозьоров Василь Євгенович - доктор фізико-математичних наук, професор кафедри комп'ютерних технологій Дніпровського національного університету імені Олеся Гончара.

Інкін Олександр Андрійович – аспірант, кафедри комп'ютерних технологій Дніпровського національного університету імені Олеся Гончара.

Belozyorov Vasily - Doctor of Physical and Mathematical Sciences, Professor, Department of computer Technologies, Oles Honchar Dnipro National University.

Inkin Oleksandr – postgraduate student, department of computer technologies, Oles Honchar Dnipro National University.

THE APPLICATION OF SPECTRAL ANALYSIS OF EEG DATA FOR THE IDENTIFICATION OF OPEN AND CLOSED EYE STATES

Annotation. The article examines the analysis of electroencephalogram (EEG) data for the classification of open and closed eye states using the Fast Fourier Transform (FFT). It is shown that this method demonstrates stable recognition accuracy at the level of 70-80% in distinguishing between open- and closed-eye states, demonstrating its effectiveness in classifying biomedical signals. General information about EEG is described, points for their reading, in particular about the “10-10 system”, information about the main types of brain rhythms is given. Modern methods for analyzing EEG data were also reviewed, highlighting three main approaches: spectral analysis, recurrence analysis, and machine learning methods. Software was developed for classification of information presented in the form of EEG time series obtained in the state of open and closed eyes. The software was developed in Python utilizing the PyRQA library.

Key words: spectral analysis, electroencephalography, time series, Fast Fourier Transform, brain rhythms, frequency artifacts, PyRQA.

Introduction and purpose. The analysis of electroencephalographic (EEG) signals is one of the leading methods of researching brain activity. Among the current tasks in this field is the classification of states of consciousness, in particular, the recognition of the states of open and closed eyes, which is important for understanding cognitive processes. Identifying distinct differences in brain activity during these states enables a deeper investigation into the dynamics of brain rhythms and their relationship to cognitive functions such as perception, concentration, and relaxation.

Today, several basic approaches are used to analyze EEG signals, including spectral analysis, recurrent analysis, and machine learning methods. Spectral analysis, in particular the Fast Fourier Transform (FFT), allows us to distinguish the main rhythms of the brain, such as alpha, beta and theta rhythms, which are associated with different states of consciousness. Recurrent analysis provides a deeper exploration of the nonlinear and chaotic characteristics of EEG signals, while machine learning techniques open up possibilities for building accurate classification models based on large data sets. In this study, spectral analysis was chosen as one of the most effective and affordable approaches for recognizing eye conditions.

The purpose of the article is, firstly, to review approaches to the analysis of EEG signals and to justify the choice of spectral analysis for further classification of open and closed eyes. The frequency characteristics of the signals, in particular the activity of the alpha rhythm,

which increases when the eyes are closed, are estimated on the basis of the FFT. Additionally, noise and artifacts are filtered to ensure signal purity. The study was performed on data from the “EEG Motor Movement/Imagery Dataset”, which includes 109 participants and allows obtaining statistically significant results for further application in neuroscience and cognitive research.

Reference Analysis. The article by Farzad et al. (2023) [1] is a systematic review investigating the presence of chaos in electroencephalogram (EEG) signals associated with human activity. The authors emphasize the importance of chaos theory for understanding the complex dynamics of brain activity and analyze the various methods used to study these phenomena. The review presents the results of many studies focusing on cognitive functions, and also found that the most common methods of chaos analysis are correlation dimension, fractal analysis and various entropy algorithms. The authors conclude that further study of the chaotic dynamics of the brain can improve our understanding of cognitive activity and contribute to the development of neuroscience.

The article by Wang et al. (2010) [2] investigates the relationship between the chaotic characteristics of EEG signals and high-level intellectual brain activity. Using phase space reconstruction to analyze one-dimensional and multidimensional time series, the authors study EEG signals during five types of conscious activity: relaxation, verbal multiplication, writing a letter, visualization of an object in three-dimensional space, and number representation. The results of the analysis of chaotic characteristics, including determinism, phase graphs and power spectra, indicate the presence of chaos in consciousness.

Statistical data confirm that the central tendencies are similar to the phase graphs, and the spectra indicate small differences between the frequency channels of conscious activities. Differences in the approximated entropy indicate a higher level of innovativeness among the subjects. The correlation dimension and the Lyapunov exponent indicate brain activity in attractors with fractional dimensions. A non-linear quantitative criterion using neural networks shows positive results in the classification of different types of activity, with arithmetic-related tasks standing out more clearly compared to other types of activity.

A paper by Gallego-Rudolf et al. (2022) [3] investigates the effect of ballistocardiography artifact (BCG) on EEG spectral characteristics during functional magnetic resonance (MR) scanning (fMRI). The authors evaluate the effectiveness of seven BCG correction methods to preserve the spectral properties of the EEG and examine the reactivity of the posterior alpha band during the eye-closing-opening (EC-EO) task.

The study included EEG recordings from 20 healthy young adults outside the MR environment and during fMRI data acquisition. The average artifact subtraction (AAS) method was found to be effective in removing the gradient artifact, while BCG correction was performed using AAS, independent component analysis (ICA), and their combinations. The authors compare the spectral power of traditional frequency bands with corrected EEG-fMRI data and off-scanner recordings.

They found that the BCG artifact significantly distorts the data in all frequency ranges, with residual artifacts remaining even after all correction methods. The results showed that

EEG reactivity to the EC-EO task is better preserved with ICA-based correction methods, especially when extracting alpha power fluctuations. The authors emphasize that current solutions to eliminate BCG artifacts have limited effectiveness, recommending the improvement of existing approaches and combining them with hardware solutions to improve the quality of the EEG signal during simultaneous recording.

The article by Makeig et al. (1996) [4] is devoted to the application of the independent component analysis (ICA) algorithm for source separation of electroencephalographic (EEG) data. Due to the distance between the skull and the brain and the different electrical conductivity, EEG data collected from the surface of the head contains activity generated in a large area of the brain, resulting in spatial blurring. However, it is not accompanied by significant time delays, making ICA appropriate for blind source separation in EEG data.

The authors demonstrate that ICA separates source identification from source localization. Early results show that ICA training is independent of random initial values and can be used to segregate artifactual components such as electromagnetic noise and eye movements. In addition, ICA is able to isolate superimposed EEG phenomena, including alpha and theta bursts.

Instabilities in EEG and behavioral state can be monitored with ICA through changes in residual correlation between output channels. The paper highlights the possibilities of ICA to improve EEG analysis and understanding of brain processes related to perception and attention.

An article by Noor et al. (2022) [5] discusses the importance of electroencephalography (EEG) data analysis and processing in brain activity studies. The researchers emphasize that there are many standard procedures for obtaining informative results in EEG analysis, but the methods used in these procedures can vary significantly depending on the preferences of the researchers and the specifics of the experiments.

One of the main challenges is that traditional manual analysis methods are resource-intensive, and researchers often focus only on the small subset of brain signals that are most relevant to the study. In response to these challenges, the authors propose an automated method for classifying eight different EEG bands (very low, delta, theta, alpha-1, alpha-2, beta-1, beta-2, and gamma) by crossing the FFT with three by machine learning methods: KNN, SVM and ANN.

Research results show that the FFT+SVM method achieves an impressive 100% accuracy, successfully classifying signals into all eight EEG bands. This approach not only improves the efficiency of EEG analysis, but also opens up new possibilities for automating data processing in neuroscience, which can greatly simplify the work of researchers in this field.

Spectral characteristics of time series. Spectral analysis is a method that facilitates the investigation of signal frequency components, particularly in the context of EEG. FFT is one of the main tools of spectral analysis, which allows you to move from the time domain to the frequency domain. This transformation helps reveal the brain's rhythms, which are essential for understanding neural functionality and states.

FFT [6] allows you to efficiently calculate the discrete Fourier transform (DFT), which is defined by the following formula:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i \frac{2\pi}{N} kn}, \quad k = 0, 1, \dots, \frac{N}{2},$$

where:

- X_k is the complex frequency for the k -th component,
- x_n is an input signal in the time domain with N discrete points,
- N is the number of points in the analyzed signal.

FFT implements an algorithm that reduces the computational complexity of DFT from $O(N^2)$ to $O(N \log N)$, making it particularly useful for processing large data sets such as EEG signals. When using FFT to analyze EEG signals, a frequency spectrum is obtained that shows which rhythms dominate brain activity at the time of recording. From the obtained spectrum, you can calculate:

- **Signal strength** in each frequency band, which helps to identify dominant rhythms.
- **Power spectral density**, which allows you to determine how much energy is in each frequency, which is important for analyzing changes in brain activity in different states.

The relationship between the index k and the frequency in Hz following the discrete Fourier transform is governed by the sampling frequency f_s and the number of points N in the signal:

$$f_k = \frac{k \cdot f_s}{N},$$

where f_k is the frequency in Hz corresponding to the index k , f_s is the sampling frequency in Hz. Thus, each index k in the spectrum corresponds to the frequency f_k , which increases linearly from 0 to $f_s/2$ Hz.

EEG signals reflect the brain's electrical activity, which consists of different frequency components, forming brain rhythms classified based on their frequency ranges. In particular, **delta rhythms (0.5–4 Hz)** are observed during deep sleep and indicate the body's restoration and regeneration; **theta rhythms (4–8 Hz)** are associated with light sleep, meditation and creative states; **alpha rhythms (8–12 Hz)** occur during quiet activity and are often associated with relaxation and rest; **beta rhythms (12–30 Hz)** are observed during active thinking, attention, and mental effort. And **gamma rhythms (greater than 30 Hz)** are associated with cognitive functions such as attention, memory and awareness.

However, when analyzing EEG signals, it is important to take into account not only brain rhythms, but also various factors that can distort them. Among such factors is frequency artifacts, which means the presence of unwanted frequency components or random frequency changes in the signal. frequency artifacts can come from a variety of sources and significantly affect signal quality and data analysis accuracy.

Among the main sources of frequency artifacts can be distinguished:

1. **Electromagnetic interference.** Electronic devices, wires or other electrical devices can emit electromagnetic fields that enter the signal reading system and change its frequency characteristics.

2. **Artifacts from muscles and movements.** Muscle movements or physical activity can create noise in the measured signal, especially if the probe is sensitive to mechanical vibrations or movements.

3. **Electrical network.** frequency artifacts can originate from the electrical network, especially from interference at frequencies of 50 or 60 Hz used in various countries to power electrical devices.

In this article, it is worth considering alpha rhythms in more detail, since the state of open and closed eyes refers to the calm activity that is inherent in the alpha rhythms of the brain.

The influence of the condition of the eyes on the activity of alpha rhythms in the brain is an important aspect for understanding the psychophysiological processes of a person. With closed eyes, the brain generally shifts towards a relaxed state. In this state, the alpha rhythm becomes more noticeable, in particular in the region of the back of the brain. It promotes deep relaxation and improves the ability to concentrate. Conversely, with the eyes open, the brain is active and focused on processing visual information and interacting with the environment. In this case, alpha rhythms usually decrease or may disappear, as the brain is busy performing various tasks and perceiving external stimuli. This contrast between states with individual eyes helps in understanding how the human brain reacts to different conditions and affects our perception and psychophysiological state.

Features of alpha waves:

1. **Frequency and location.** Alpha waves are most often found in the back of the brain (occipital region), although they can be present in other parts of the brain as well.

2. **Psychophysiological conditions.** Alpha waves become more prominent during rest and mindless contemplation, with the eyes closed, and during mild relaxation without active mental or physical activity.

3. **Association with relaxation.** A high amplitude of alpha rhythms in the back of the brain may indicate a state of relaxation. This may be useful for research into stress, anxiety, and sleep disorders, as changes in alpha activity may reflect changes in a person's psychophysiological state.

Points for reading EEG. EEG uses special points to place electrodes on the surface of the scalp, which enables precise measurements of cerebral electrical activity. These points are determined according to the international system of electrode placement, which is known as the "10-10" system.

A continuation of this approach is the "10-10" system (fig. 1), which is the standard for placing electrodes in EEG studies. This system got its name from the percentage distances between the electrodes, which are measured as a percentage of the size of the head. It ensures standardized and reproducible electrode positioning, which is critical for comparing results between different studies and patients.

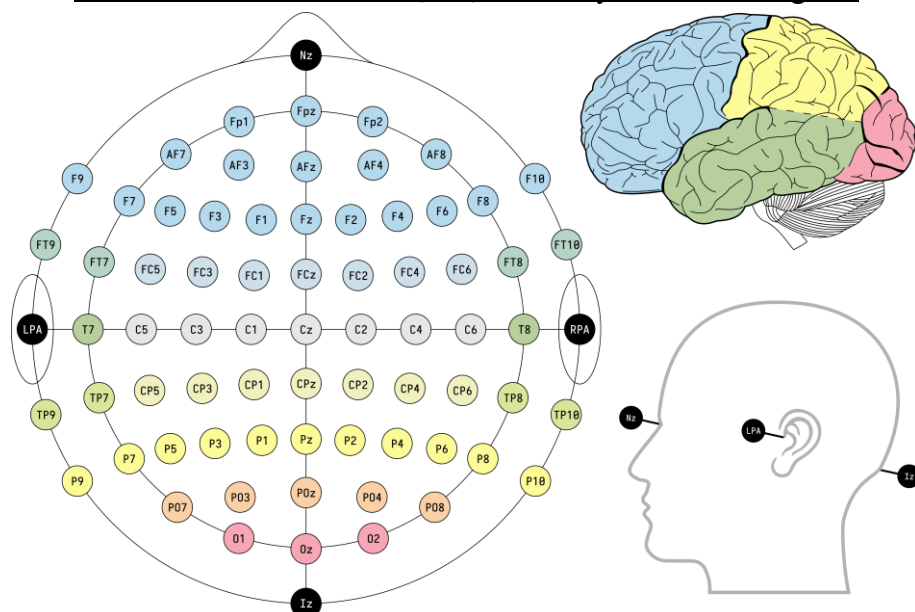


Figure 1 - “10-10” system

Within the framework of the “10-10” system, there are several main points on which electrodes are placed for reading EEG signals. These points include:

1. **Fp (frontal polar):**
 - Fp1: frontal polar point on the left.
 - Fp2: frontal polar point on the right.
2. **F (frontal):**
 - F3, F4: left and right frontal points.
 - F7, F8: lower forehead points left and right.
 - Fz: middle frontal point.
3. **C (central):**
 - C3, C4: left and right center points.
 - Cz: middle center point.
4. **P (parietal):**
 - P3, P4: parietal points on the left and right.
 - Pz: middle parietal point.
5. **O (occipital):**
 - O1, O2: occipital points on the left and right.
6. **T (temporal):**
 - T3, T4: left and right temporal points.
 - T5, T6: lower temporal points on the left and right.
7. **A (auricular):**
 - A1, A2: points on the left and right ears, often used as reference electrodes.
8. **Nasion:** Point on bridge of nose.
9. **Inion:** A point on the back of the head, where the occipital bone protrudes.

These points play an important role in ensuring the accuracy and reliability of the obtained data, which is critical for successful analysis of the electrical activity of the brain.

Investigation of the state of human eyes according to known spectral characteristics. When analyzing the frequency spectrum, we face the problem of discretization of frequencies, which affects the accuracy of determining frequency ranges. Because of this, the frequencies cannot be analyzed in a continuous physical sense, as the frequency limits change depending on the chosen sampling step. Since our task is to select certain frequency ranges and eliminate noise at other frequencies, it is necessary to apply filters with increased accuracy of limiting frequency components. For this, during the research, the Hann window function [7] was chosen for a limitation of frequency ranges, which avoids sharp transitions and provides a smoother smoothing of the spectrum.

Let's build from the well-known Hann window function, which looks like this:

$$w[n] = 0.5 \left[1 - \cos\left(\frac{2\pi n}{N}\right) \right] = \sin^2\left(\frac{\pi n}{N}\right), \quad 0 \leq n \leq N,$$

modified function $w(n, \Delta f)$, which will smooth the required range $[0, \Delta f]$:

$$w(f) = 0.5 \cdot \left[1 - \cos\left(\frac{2\pi f}{2\Delta f}\right) \right].$$

Let's build functions for the low-frequency H_{low} and high-frequency filters H_{high} that will allow us to process the signal within the specified range of the frequencies f_{low} Hz and f_{high} Hz. For this, we use gradual sinusoidal frequency limiting, which allows to reduce the impact of artifacts arising from sharp transitions in the frequency spectrum.

$$H_{low}(f_k, f_{low}) = \begin{cases} 0, & f_k < f_{low} - \Delta f \\ w(f_k - f_{low} + \Delta f), & f_{low} - \Delta f < f_k < f_{low} \\ 1, & f_{low} < f_k \end{cases}$$

$$H_{high}(f_k, f_{high}) = \begin{cases} 1, & f_k < f_{high} \\ 1 - w(f_k - f_{high}), & f_{high} < f_k < f_{high} + \Delta f \\ 0, & f_{high} + \Delta f < f_k \end{cases}$$

By combining both filters, we get the general filter function:

$$H(f_k, f_{low}, f_{high}) = H_{low} * H_{high}$$

Thus, it is possible to construct a filter that provides a smooth sinusoidal transition of values at the boundaries of the given range $[f_{low}, f_{high}]$. In further experiments, the value $\Delta f = 2$ will be used, which allows us to achieve sufficient accuracy in signal processing, reducing the influence of noise and artifacts. This provides more stable results and better data processing quality in the frequency domain. As noise, we will consider frequencies from 59 to 61 Hz, and those close to 0 Hz, that is, according to the increasing filter from 0 to 2 Hz:

$$H_f(f_k) = 1 - H_{low}(f_k, 2) - H(f_k, 59, 61)$$

Then, let's assume that the noise-free value of the DFT is defined by the formula:

$$Z_k = X_k * H_f(f_k) \tag{1}$$

The EEG Motor Movement/Imagery Dataset [8] was used for analysis, which contains more than 1,500 one- and two-minute EEG recordings from 109 participants. As part of the experiment, participants performed various motor and imagination tasks, during which 64-channel EEG signals were recorded using the BCI2000 system [9]. Each participant completed 14 experimental trials, including two one-minute baseline trials (one with eyes open and one with eyes closed) and three two-minute trials for each of the four tasks.

The task consisted of the following: participants opened and closed their fist according to the appearance of a target on the left or right side of the screen, or imagined performing this movement. Additionally, participants opened and closed both fists or both feet depending on the position of the target.

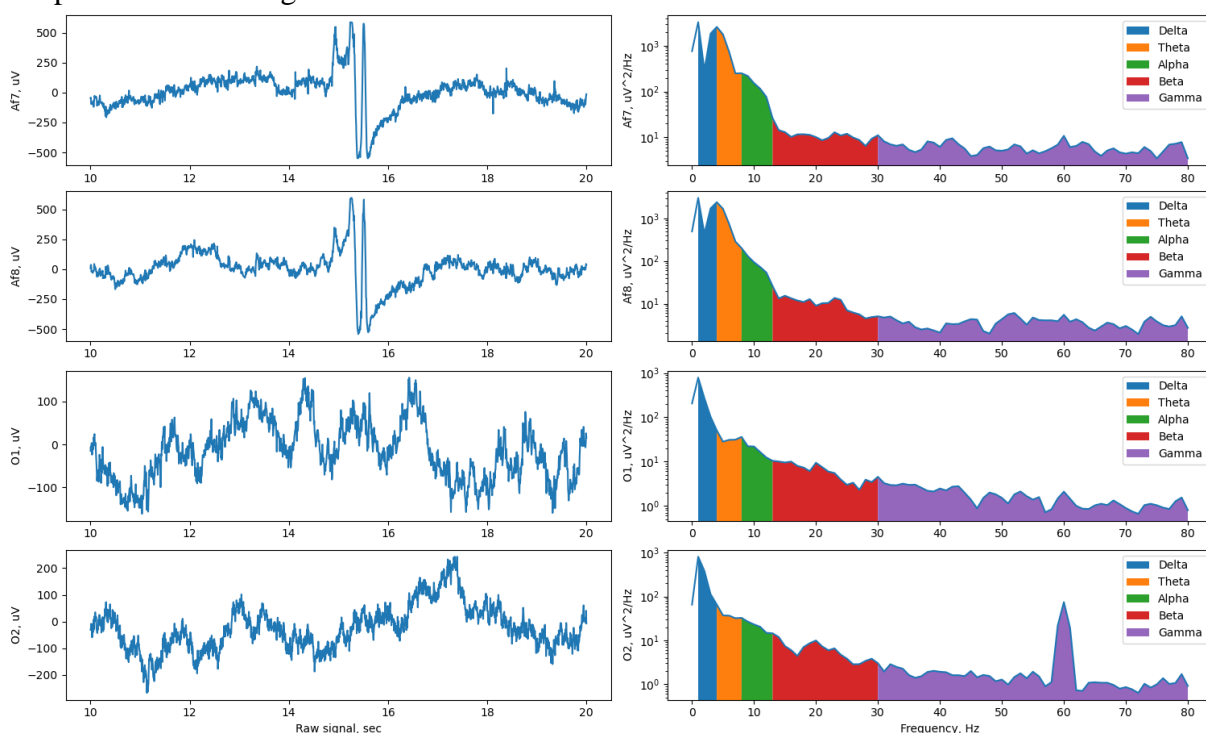


Figure 2 - Brain rhythms (opened eyes, S043R01.edf)

The recordings are in EDF+ format, containing 64 EEG signals recorded at 160 Hz, as well as an annotation channel. Annotations include codes denoting different states: T0 for rest, T1 for initiation of movement (real or imagined) of the left fist or both fists, and T2 for movement of the right fist or both legs. Recordings are made using electrodes according to the international 10-10 system, that excludes some specific electrodes. The EDF+ format is adapted for use with the PhysioToolkit software [10].

Each participant is labeled S001 to S109, and each participant has 14 files labeled R01-R14 in which the corresponding EEG data is recorded. Thus, for each write iteration, there are SsssRrr.edf format files. In the future, we will denote as *s* the number of the participant, and *r* the number of the stage. For our study, we used the following stages: 1 - no task, eyes open, 2 - no task, eyes closed.

The PyRQA library [11] was used for the analysis, which provides efficient execution of recurrence analysis on long time series thanks to the use of the OpenCL framework.

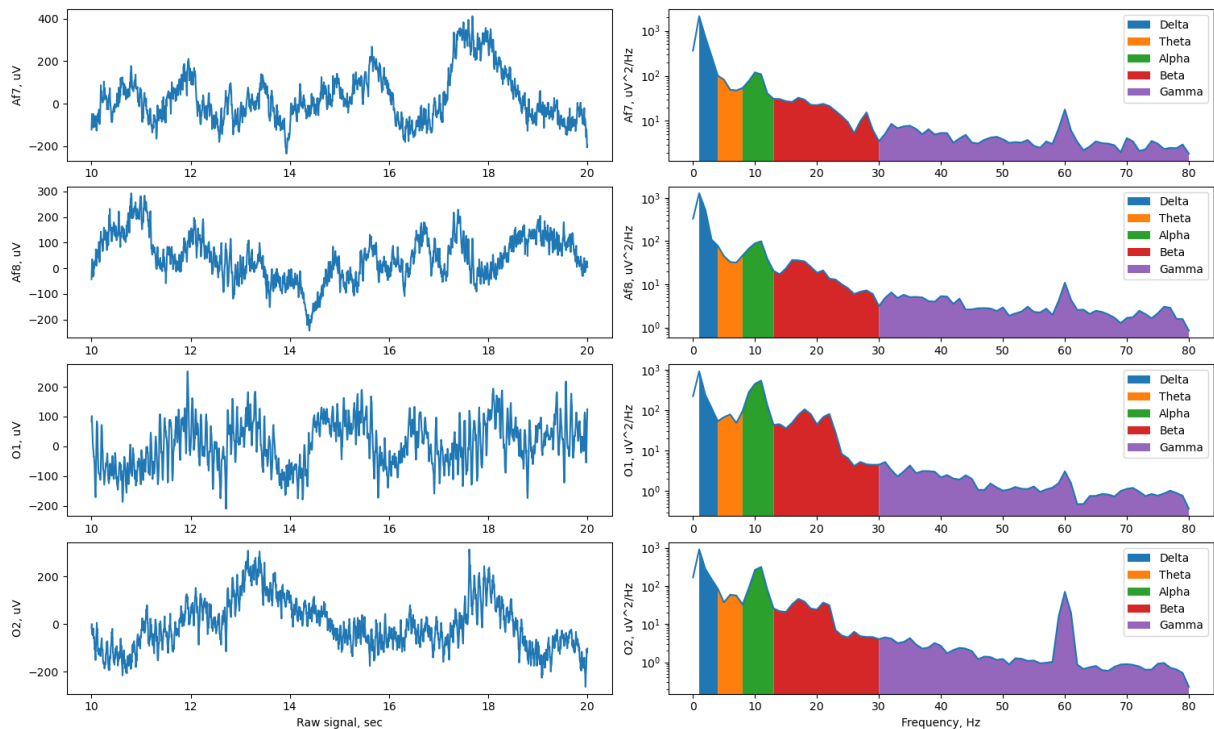


Figure 3 - Brain rhythms (closed eyes, S043R02.edf)

Figures 2 and 3 present the EEG signals of participant S043 from the studied data set, as well as the power of each frequency calculated using the Welch method [12] of spectral density calculation. These graphs clearly illustrate the differences between open and closed eyes. In particular, increased activity in the alpha rhythm is noticeable during eyes closed, where a peak frequency of about 10 Hz is observed, which is typical for this rhythm. In some cases, similar phenomena can be observed in beta rhythm. In addition, the signals from the electrodes located in the frontal area (points Af7 and Af8) show sharp fluctuations associated with the participant's blinking when the eyes are open. These artifacts originate from ocular muscle activity rather than cerebral signals. In the back regions of the head, in particular in the occipital region, blinking does not occur, since these points are located at a considerable distance from the eyes. In addition, the spectra clearly show a peak at a frequency of 60 Hz, which corresponds to the noise of the power grid.

During the construction and analysis of spectral diagrams, it was found that when the eyes are closed, there is a decrease in the activity of delta and theta rhythms, while the activity of alpha and beta rhythms increases. To solve the problem of recognizing the state of the eyes, the power in the range from 8 Hz to 21 Hz, which corresponds to the alpha rhythm and part of the beta rhythms, as well as the total signal power, was calculated. The ratio of these powers can serve as a marker to distinguish open from closed eye states.

Using the noise-free signal (1), we construct a filter for the required frequency range $H_e(f_k) = H(f_k, 7, 22)$ and a power function for the given range P_e and of the entire possible frequency range P :

$$P = \sum_{k=0}^{N/2} |Z_k|^2$$

$$P_e = \sum_{k=0}^{N/2} |Z_k * H_e(f_k)|^2$$

$$R_e = \frac{P_e}{P}$$

In our case, we work with data relative to time, so it is advisable to divide the time series into several segments of the minimum length without losing information. It was established that the optimal length of segments for calculations is 5-10 seconds. At shorter time intervals, significant fluctuations of values between segments are observed relative to time. This is clearly visible in figures 4 and 5, where spectrograms of the participant are shown using STFT. In the figure 4, where the segments are 1 second each, the values are very varied, and mathematically it is difficult to determine a clear pattern. Instead, the 10-second segments show the most powerful frequencies that remain stable over time.

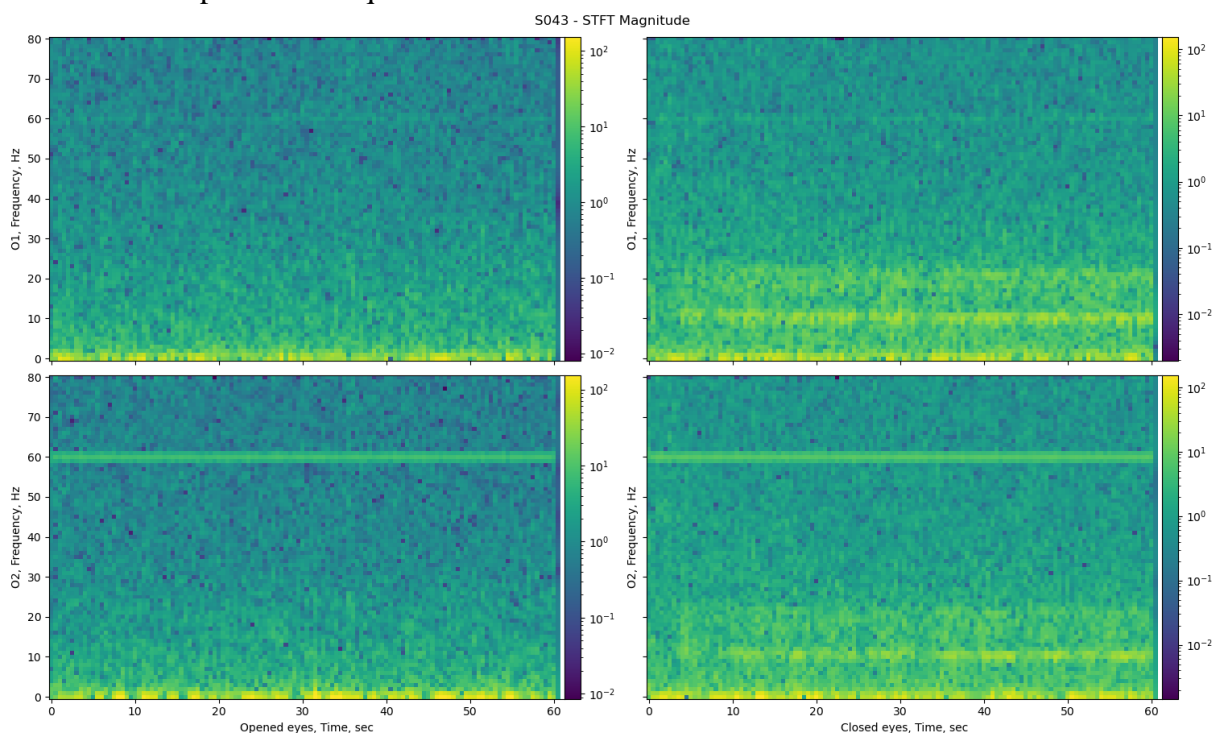


Figure 4 - Spectrograms of participant S043 with 1 second segments

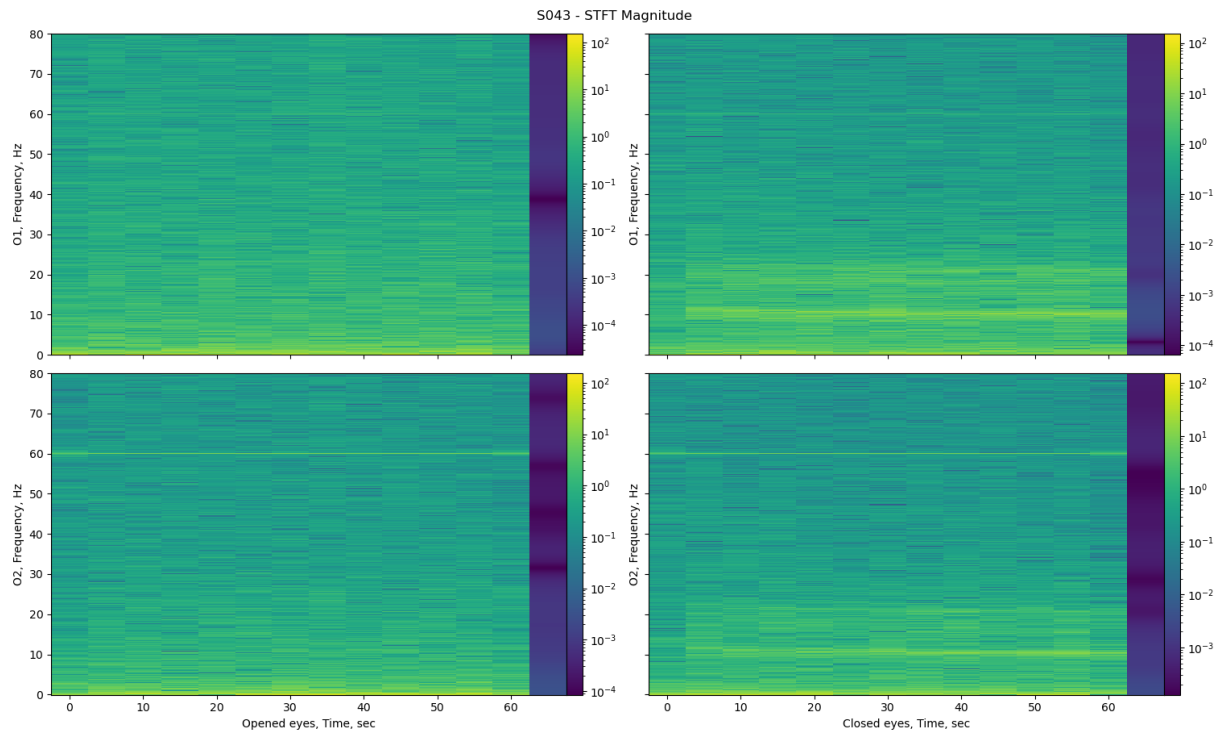


Figure 5 - Spectrograms of participant S043 with 10 second segments

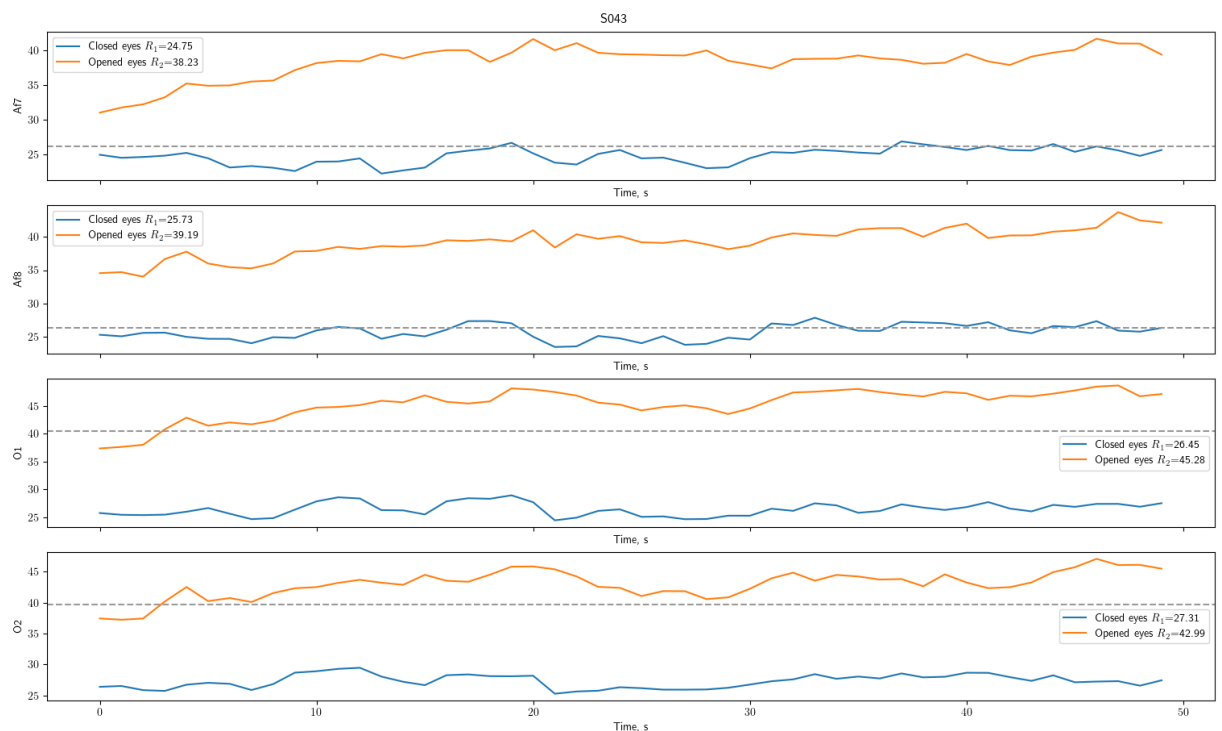


Figure 6 - Percentage of 8-21 Hz band power relative to total power

Figure 6 presents graphs converted to percentages, where the dashed line represents the average value of this ratio over the entire dataset. As can be seen, for participant S043, the alpha rhythm shows a clear trend over the entire time interval: with eyes closed the power is above average, while with eyes open it is lower.

In order for identifying open- and closed-eye states, let's calculate for each point of the brain some threshold value among all participants of the dataset R_{avg} , above which it can be considered that the participant has closed eyes, and below which the eyes are open. And for each participant, we will calculate R_{sr} , where s is the participant's number, and r is the stage number, that is, $r = 1$ for open eyes and $r = 2$ for closed eyes. In this way, we can build the simplest classifier. Let's calculate the classification accuracy for open-eye A_o and closed-eye A_c states, and take the general one as a minimum of these values $A = \min(A_o, A_c)$. We will also calculate the average absolute difference between R_{avg} and R_{sr} for each participant:

$$D_{avg} = \frac{1}{n} \sum_{s=0}^n |R_{avg} - R_{sr}|$$

This formula shows how close participants' classification values are to the common threshold value, and lower values may indicate better classification accuracy. Here is a table 1 containing the classification results of the eye state (open or closed) for different points in the brain, including the classification accuracy, the mean threshold value, and the mean absolute difference between the participants' values and the total threshold.

Table 1

Open and closed eyes classification results

Point	A	A_o	A_c	R_{avg}	D_{avg}
Fc3	75.45%	79.09%	75.45%	33.94%	7.60%
Fc1	75.45%	76.36%	75.45%	34.23%	7.51%
Fcz	75.45%	75.45%	75.45%	34.67%	7.65%
Fc2	75.45%	78.18%	75.45%	34.39%	7.84%
Fc4	75.45%	76.36%	75.45%	32.99%	7.91%
Fc6	73.64%	80.00%	73.64%	30.14%	7.83%
Cz	73.64%	73.64%	74.55%	35.72%	7.07%
C6	73.64%	78.18%	73.64%	31.53%	6.65%
Cp4	73.64%	73.64%	73.64%	36.61%	7.09%
Fp1	77.27%	83.64%	77.27%	27.10%	10.42%
Fpz	79.09%	86.36%	79.09%	28.00%	10.75%
Fp2	79.09%	87.27%	79.09%	26.93%	10.69%
Af7	78.18%	84.55%	78.18%	26.16%	9.87%
Af3	76.36%	84.55%	76.36%	29.28%	10.29%
Afz	80.00%	84.55%	80.00%	31.53%	10.41%
Af4	79.09%	84.55%	79.09%	29.01%	10.95%
Af8	80.00%	87.27%	80.00%	26.47%	10.27%
F5	73.64%	74.55%	73.64%	29.28%	8.67%
F3	78.18%	82.73%	78.18%	32.04%	9.07%

Point	A	A_o	A_c	R_{avg}	D_{avg}
F1	79.09%	80.91%	79.09%	33.09%	9.02%
Fz	79.09%	80.91%	79.09%	33.45%	9.11%
F2	80.00%	81.82%	80.00%	32.66%	9.61%
F4	80.91%	82.73%	80.91%	32.11%	9.96%
F6	78.18%	80.91%	78.18%	29.01%	10.06%
F8	74.55%	83.64%	74.55%	27.40%	8.99%
Ft8	74.55%	75.45%	74.55%	28.39%	7.43%
Po3	74.55%	77.27%	74.55%	39.52%	8.85%
Po4	74.55%	79.09%	74.55%	39.27%	9.57%
Po8	74.55%	81.82%	74.55%	38.71%	10.41%
O1	73.64%	73.64%	73.64%	40.51%	9.29%
Oz	74.55%	79.09%	74.55%	40.19%	9.02%
O2	76.36%	76.36%	76.36%	39.81%	9.73%

Thus, we can conclude that with careful selection of the dataset and the application of appropriate data processing methods, spectral analysis allows for the effective classification of states of brain activity, such as open and closed eyes. The use of correct filters and window functions to select the necessary frequency components allows you to minimize the impact of noise and ensures stable recognition accuracy. Therefore, proper processing of EEG data and adaptation of spectral analysis to specific tasks increase the possibilities of using this method for reliable analysis of brain states.

Conclusions. In the results of the conducted research, spectral analysis was used to classify the states of open and closed eyes based on data from the “EEG Motor Movement/Imagery Dataset”. Hann’s window function was used for the analysis, which ensured a smooth limitation of frequency components and minimized frequency fluctuations between frequency segments. Time segments with a duration of 5-10 seconds turned out to be optimal for calculations, since shorter intervals showed significant fluctuations in spectral characteristics.

The most accurate results were obtained for points in the frontal area, in particular **Fp2**, **Af7** and **Af8**, where the classification accuracy reached **79.09%-80.00%** for closed eyes and **87.27%** for open. In addition, significant results were obtained for points in the occipital region, particularly **O1** and **O2**, where accuracy reached **73.64%** for closed eyes and **76.36%** for open eyes. To eliminate artifacts, filters for frequency artifacts in the range of 59-61 Hz (mains noise) and low-pass filters for frequencies around 0 Hz were used.

REFERENCES

1. S. Kargarnovin, C. Hernandez, F. V. Farahani, and W. Karwowski Evidence of Chaos in Electroencephalogram Signatures of Human Performance: A Systematic Review / S. Kargarnovin, C. Hernandez, F. V. Farahani, and W. Karwowski. - Brain Sci. - vol. 13, no. 5, May 2023. - p. 813.

2. X. Wang, J. Meng, G. Tan, and L. Zou Research on the relation of EEG signal chaos characteristics with high-level intelligence activity of human brain / X. Wang, J. Meng, G. Tan, and L. Zou. - *Nonlinear Biomedical Physics*. - vol. 4, no. 1, Apr. 2010. - p. 2.
3. J. Gallego-Rudolf, M. Corsi-Cabrera, L. Concha, J. Ricardo-Garcell, and E. Pasaye-Alcaraz Preservation of EEG spectral power features during simultaneous EEG-fMRI / J. Gallego-Rudolf, M. Corsi-Cabrera, L. Concha, J. Ricardo-Garcell, and E. Pasaye-Alcaraz. - *Front Neurosci*. - vol. 16, 2022. - p. 951321.
4. S. Makeig, A. Bell, T.-P. Jung, and T. J. Sejnowski Independent Component Analysis of Electroencephalographic Data / S. Makeig, A. Bell, T.-P. Jung, and T. J. Sejnowski. - in *Advances in Neural Information Processing Systems*. - vol. 8, 1995. -.
5. N. S. M. Nor, N. H. A. H. Malim, N. A. P. Rostam, J. J. Thomas, M. A. Effendy, and Z. Hassan Automated classification of eight different Electroencephalogram (EEG) bands using hybrid of Fast Fourier Transform (FFT) with machine learning methods / N. S. M. Nor, N. H. A. H. Malim, N. A. P. Rostam, J. J. Thomas, M. A. Effendy, and Z. Hassan. - *Neuroscience Research Notes*. - vol. 5, no. 1, 1, Mar. 2022. - pp. 116–116.
6. J. Hu, F. Jia, and W. Liu Application of Fast Fourier Transform / J. Hu, F. Jia, and W. Liu. - *Highlights in Science, Engineering and Technology*. - vol. 38, Mar. 2023. - pp. 590–597.
7. J. R. Winkler *Numerical recipes in C: The art of scientific computing, second edition* / J. R. Winkler. - *Endeavour*. - vol. 17, no. 4, Jan. 1993. - p. 554.
8. EEG Motor Movement/Imagery Dataset /. [Online]. Available: <https://archive.physionet.org/pn4/eegmmidb/>. [Accessed: 21-Feb-2024].
9. G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw BCI2000: A general-purpose brain-computer interface (BCI) system / G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw. - *IEEE Trans Biomed Eng*. - vol. 51, no. 6, Jun. 2004. - pp. 1034–1043.
10. A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals / A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley. - *Circulation*. - vol. 101, no. 23, Jun. 2000. - pp. E215–220.
11. T. Rawald, M. Sips, and N. Marwan PyRQA—Conducting recurrence quantification analysis on very long time series efficiently / T. Rawald, M. Sips, and N. Marwan. - *Computers & Geosciences*. - vol. 104, Jul. 2017. - pp. 101–108.
12. P. Welch The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms / P. Welch. - *IEEE Trans. Audio Electroacoust*. - vol. 15, no. 2, Jun. 1967. - pp. 70–73.

Received 17.09.2024.
Accepted 21.09.2024.

***The application of spectral analysis of EEG data
for the identification of open and closed eye states***

The article examines the analysis of electroencephalogram (EEG) data for the classification of open and closed eye states using the Fast Fourier Transform (FFT). It is shown that this method demonstrates stable recognition accuracy at the level of 70-80% in distinguishing between open- and closed-eye states, demonstrating its effectiveness in classifying biomedical signals. General information about EEG is described, points for their reading, in particular about the “10-10 system”, information about the main types of brain rhythms is given. Modern methods for analyzing EEG data were also reviewed, highlighting three main approaches: spectral analysis, recurrence analysis, and machine learning methods. Software was developed for classification of information presented in the form of EEG time series obtained in the state of open and closed eyes. The software was developed in Python utilizing the PyRQA library.

Панасенко Єгор Сергійович - аспірант кафедри комп'ютерних технологій Дніпровського національного університету імені Олеся Гончара, Факультет Прикладної Математики, спеціальність 113 “Прикладна математика”.

Білозьоров Василій Євгенєвич - доктор фізико-математичних наук, професор кафедри комп'ютерних технологій Дніпровського національного університету імені Олеся Гончара.

Yehor Serhiyovych Panasenko - PhD student, Department of Computer Technologies Oles Honchar Dnipro National University, Faculty of Applied Mathematics, specialty 113 “Applied Mathematics”.

Belozyorov Vasily Yevhenovych - Doctor of Physical and Mathematical Sciences, Professor, Department of computer Technologies, Oles Honchar Dnipro National University.

АЛГОРИТМИ ІМПУТУВАННЯ ПРОПУСКІВ У ДАНИХ НА ОСНОВІ ЕНТРОПІЇ

Анотація. Пропонується два алгоритми імпутування пропусків у даних (неітераційний та ітераційний) в задачах класифікації, оснований на мінімізації умовної ентропії. Розглядається рішення для кількісних та якісних ознак, у випадку кількісних – для дискретних та неперервних. Для аналізу алгоритмів пропонується три типи тестів. Перший тест працює з повним датасетом, в який штучно вносяться пропуски, проводиться імпутування різними методами, порівняння алгоритмів відбувається за середньоквадратичною похибкою та часом виконання алгоритмів. Другий тип тесту порівнює умовну ентропію до та після імпутування для різних методів. Третій тип тесту пов'язаний із задачею класифікації, коли моделі бінарної класифікації навчають на датасетах з імпутованими різними методами даними, та порівнюють точність класифікації на тестових вибірках. Розглядаються два відомих датасети про спостереження стосовно ішемічної хвороби серця.

Ключові слова: імпутування даних, пропуски у даних, умовна ентропія, теорія інформації, алгоритми обробки даних, мінімізація невизначеності, класифікація, якісні та кількісні ознаки, ітераційний метод, ентропійний підхід, машинне навчання, обробка відсутніх даних, взаємна інформація, інженерія програмного забезпечення, інтелектуальний аналіз даних, діаграма діяльності.

Вступ. Пропуски в даних можуть виникати з різних причин і мають різні характеристики. Розуміння типів пропусків у даних є важливим для вибору відповідних методів обробки та аналізу даних. Існують три основні типи пропусків [1]:

1. MCAR (Missing Completely at Random) — пропуски, що виникають повністю випадково. Пропуски вважаються повністю випадковими, якщо відсутність значення у змінній не пов'язана ні з якими іншими змінними або з самим значенням змінної. Наприклад, якщо при введенні даних деякі значення відсутні через випадкові помилки в процесі запису даних або спостереження не відбулось з випадкової причини, і ці пропуски не залежать від значень самих даних, то ці пропуски можна вважати MCAR. Аналіз даних за умов MCAR є найменш проблематичним, оскільки пропуски є випадковими і не упереджують результати аналізу.

2. MAR (Missing at Random) — пропуски, що виникають випадково. Пропуски вважаються випадковими, якщо ймовірність виникнення пропуску може залежати від інших спостережуваних даних, але не залежить від неспостережуваних значень, тобто пропуски можна пояснити іншими змінними в наборі даних. Наприклад, припустимо,

що люди з низьким доходом в опитуванні частіше не відповідають на питання про дохід, але це все одно можна передбачити на основі інших даних, таких як рівень освіти чи професія. У такому випадку пропуски можна вважати MAR. Пропуски типу MAR можуть бути оброблені такими методами, як множинна імпутація або регресійна імпутація, які враховують зв'язки між змінними для прогнозування відсутніх значень. Методи, які не враховують ці зв'язки, можуть призвести до зміщених результатів.

3. MNAR (Missing Not at Random) — пропуски, що не виникають випадково. Пропуски вважаються не випадковими, якщо ймовірність виникнення пропуску залежить від неспостережуваних значень або самого відсутнього значення. Це означає, що сам факт відсутності даних пов'язаний з конкретними характеристиками даних. Наприклад, якщо люди з високим рівнем доходу не хочуть відповідати на питання про свій дохід у соціальному опитуванні, і ця інформація про дохід недоступна, то пропуски не є випадковими (MNAR). У цьому випадку відсутні дані безпосередньо залежать від значень самих даних. Пропуски типу MNAR є найбільш складними для обробки, оскільки відсутність даних залежить від самих даних, які відсутні. Для коректного аналізу потрібно робити додаткові припущення або використовувати спеціалізовані методи моделювання пропусків.

Розуміння типів пропусків у даних є критично важливим для вибору правильних методів обробки даних. Правильне ідентифікування типу пропусків (MCAR, MAR або MNAR) дозволяє вибрати відповідні методи заповнення пропусків або коригування моделей, що забезпечує більш точні та надійні результати аналізу. Дана робота орієнтована на перші два типи пропусків MCAR та MAR.

Постановка проблеми в загальному вигляді. У задачі класифікації ми маємо набір даних, який містить декілька ознак $X = \{X_1, X_2, \dots, X_n\}$ та мітку класу Y , яка приймає значення з певної множини класів $\{C_1, C_2, \dots, C_k\}$. Метою є навчити класифікатор, який може передбачати мітку класу Y на основі вхідних ознак X .

Проте в реальних наборах даних деякі значення в матриці ознак X є відсутніми (позначається як NaN або інші спеціальні значення). Відсутні дані ускладнюють процес навчання класифікатора, оскільки багато алгоритмів машинного навчання не можуть обробляти пропуски безпосередньо. Пропуски даних можуть вплинути на класифікацію різними способами:

- зниження точності моделі, оскільки модель навчається на неповних даних і може пропустити важливі закономірності;
- зсув оцінок моделі, тобто якщо пропуски не є випадковими (випадок MNAR), модель може вивчити упереджені оцінки, що призведе до неправильних прогнозів;
- скорочення розміру вибірки для навчання, якщо обирати метод видалення рядків з пропусками, що може призвести до менш стабільних моделей.

Метою імпутування пропусків є заповнення відсутніх значень таким чином, щоб зменшити вплив пропусків на результати класифікації та забезпечити якомога точніше навчання моделі.

Аналіз останніх досліджень. В нашому попередньому дослідженні [2] описано популярні підходи та запропоновано декілька власних методів імпутування пропусків у даних на основі машинного навчання. Крім того, останнім часом набуває популярності глибинне навчання для імпутування пропусків, особливо у складних і великих наборах даних. В роботі [3] представлено метод з використанням генеративних змагальних мереж, в роботі [4] пропонується використовувати автоенкодера. Приділяється увага ансамблевим методам, наприклад, MissForest [5] та XGBoost [6, 7]. В роботі [8] представлено ідею методу імпутування пропущених даних на основі ентропійного підходу. Дана робота представляє алгоритм імпутування даних на основі мінімізації умовної ентропії.

Мета роботи. Метою роботи є розроблення алгоритму імпутування пропусків у даних, які представлені якісними та кількісними ознаками, в тому числі дискретними та неперервними даними.

Основна частина. Метод ентропійної інтерполяції відсутніх даних у задачах класифікації використовує поняття ентропії з теорії інформації для керування процесом заповнення відсутніх значень у наборах даних. У цьому контексті ентропія вимірює кількість невизначеності або випадковості в даних. Метою такого методу є зменшення невизначеності щодо міток класів шляхом вибору найбільш інформативних значень для відсутніх даних.

В теорії інформації ентропія є мірою невизначеності або випадковості. Для випадкової величини Y з розподілом ймовірностей $P(Y)$ ентропія $H(Y)$ визначається як:

$$H(Y) = - \sum_{y \in Y} P(y) \log P(y).$$

В контексті інтерполяції пропущених даних ентропія може бути використана для кількісної оцінки невизначеності, пов'язаної з інтерполяцією пропущених значень.

У задачах класифікації ентропія може бути використана для вимірювання невизначеності міток класів за даними спостережень. Якщо деякі ознаки відсутні, мета полягає в тому, щоб врахувати ці відсутні значення таким чином, щоб мінімізувати ентропію передбачуваних міток класів, тим самим покращуючи ефективність класифікації.

Для дискретної випадкової величини Y з можливими класами C_1, C_2, \dots, C_k і ймовірностями $P(C_1), P(C_2), \dots, P(C_k)$ ентропія $H(Y)$ визначається як:

$$H(Y) = - \sum_{i=1}^k P(C_i) \log_2 P(C_i) \quad (1)$$

Основа логарифма визначає одиницю ентропії. Зазвичай основа дорівнює 2 та одиниця ентропії – це біти.

На практиці, щоб обчислити ентропію для класів, потрібно виконати наступні кроки:

1. Порахувати кількість входжень кожного класу в наборі даних.
2. Обчислити ймовірності кожного класу: поділити кількість кожного класу на загальну кількість рядків даних.
3. Обчислити ентропію розподілу класів за формулою (1).

Для набору даних з кількома ознаками ентропія класів може змінюватися в залежності від значення конкретної ознаки. Умовна ентропія вимірює невизначеність класів за умови заданої ознаки.

Для ознаки X з можливими значеннями x_1, x_2, \dots, x_m умовна ентропія $H(Y|X)$ визначається за формулою

$$H(Y|X) = \sum_{j=1}^m P(x_j)H(Y|X = x_j), \quad (2)$$

де

$P(x_j)$ – ймовірність того, що ознака X приймає значення x_j ;

$H(Y|X = x_j)$ – ентропія Y за умови $X = x_j$, яка обчислюється наступним чином

$$H(Y|X = x_j) = - \sum_{i=1}^k P(C_i|X = x_j) \log_2 P(C_i|X = x_j). \quad (3)$$

З формул (2) та (3) отримаємо наступну загальну формулу для умовної ентропії

$$H(Y|X) = - \sum_{j=1}^m P(x_j) \sum_{i=1}^k P(C_i|X = x_j) \log_2 P(C_i|X = x_j) \quad (4)$$

На практиці, щоб обчислити умовну ентропію, потрібно виконати наступні кроки:

1. Обчислити $P(C_i|X = x_j)$ – умовні ймовірності для кожного значення класу C_i за умови значення ознаки x_j , тобто порахувати для кожного класу C_i , скільки в наборі даних рядків, для яких $Y = C_i$ та $X = x_j$.

2. Визначити ентропію для кожного значення ознаки за формулою (3).

3. Зважити ентропії на ймовірність кожного значення ознаки та підсумувати їх.

Взаємна інформація, або *інформаційний вигравш* показує, наскільки знання ознаки X зменшує невизначеність щодо Y , та обчислюється за формулою

$$G(Y, X) = H(Y) - H(Y|X). \quad (5)$$

Розглянемо *приклад обчислення*. Нехай в нас є набір даних, що має одну ознаку X з двома можливими значеннями x_1 та x_2 та бінарним класом Y з двома можливими значеннями C_1 та C_2 . В таблиці 1 представлено кількість рядків, що відповідають значенням x_j та C_i

Бачимо, що в нас є набір даних, що містить:

100 рядків, де 60 належать до класу C_1 і 40 - до класу C_2 ;

x_1 з'являється у 70 рядках: 50 з C_1 і 20 з C_2 ;

x_2 з'являється у 30 рядках: 10 з C_1 і 20 з C_2 .

Таблиця 1

Набір даних прикладу

	C_1	C_2	Разом
x_1	50	20	70
x_2	10	20	30
Разом	60	40	100

Обчислимо ентропію $H(Y)$ за формулою (1):

$$P(C_1) = \frac{60}{100} = 0.6, \quad P(C_2) = \frac{40}{100} = 0.4$$

$$H(Y) = -(0.6 \log_2 0.6 + 0.4 \log_2 0.4) \approx 0.97$$

Обчислимо ентропію для кожного значення ознаки за формулою (3):

$$P(C_1|X = x_1) = \frac{50}{70} = \frac{5}{7}, \quad P(C_2|X = x_1) = \frac{20}{70} = \frac{2}{7},$$

$$H(Y|X = x_1) = -\left(\frac{5}{7} \log_2 \frac{5}{7} + \frac{2}{7} \log_2 \frac{2}{7}\right) \approx 0.86,$$

$$P(C_1|X = x_2) = \frac{10}{30} = \frac{1}{3}, \quad P(C_2|X = x_2) = \frac{20}{30} = \frac{2}{3},$$

$$H(Y|X = x_2) = -\left(\frac{1}{3} \log_2 \frac{1}{3} + \frac{2}{3} \log_2 \frac{2}{3}\right) \approx 0.92,$$

Загальна умовна ентропія $H(Y|X)$ обчислюється за формулою (4):

$$H(Y|X) = \frac{70}{100} \times 0.86 + \frac{30}{100} \times 0.92 \approx 0.88$$

Обчислимо взаємну інформацію за формулою (5):

$$G(Y, X) = H(Y) - H(Y|X) = 0.97 - 0.88 = 0.09.$$

Головна ідея методу полягає в виборі для ознаки X серед всіх можливих значень x_1, x_2, \dots, x_m такого значення x_j для імпутування, яке буде максимізувати взаємну інформацію $G(Y, X) \rightarrow \max$. Оскільки $H(Y)$ не залежить від X , то нам достатньо мінімізувати умовну ентропію, тобто знайти такий x_j для якого $H(Y|X) \rightarrow \min$.

Пропонується виконувати імпутування кожної ознаки окремо за наступним алгоритмом. Спочатку розглянемо випадок дискретної випадкової величини (категоріальна ознака).

Алгоритм 1. Імпутування категоріальної ознаки

1. Розглянемо піднабір даних D , що складається з ознаки X та цільового стовпця Y . Розділимо дані на дві множини: повна D^f та неповна D^{nf} .
2. Беремо перший рядок з D^{nf} і виконуємо крок підбору значення, що імпутується.
3. Крок підбору значення: обчислюємо умовну ентропію для кожного можливого значення $x_j, j \in [1, \dots, m]$ на множині D^f , до якої доєднаний x_j .
4. Вибираємо в якості значення, що імпутується, таку x_j , яка мінімізує умовну ентропію.
5. Виконуємо оновлення набору даних, виконавши імпутацію в рядку, що розглядається.
6. Повторюємо кроки 1 – 5, поки множина D^{nf} не пуста.

Також пропонується модифікація алгоритму 1, де використовується декілька ітерацій підбору, поки загальна умовна ентропія продовжує зменшуватись.

Фактично виявляється, що для пропущеного значення класу C_i значення x_j максимізує умовну ймовірність $P(C_i|X = x_j)$ в заповненому наборі даних.

Алгоритм 2. Імпутування категоріальної ознаки ітераційно

1. Встановлюємо максимально припустиму кількість ітерацій.
2. Встановлюємо номер ітерації в 1.

3. Встановлюємо поточне значення умовної ентропії у заздалегідь задане велике число, наприклад, 10000.
4. Якщо номер ітерації не перевершує максимально припустиму кількість, переходимо до наступного кроку, інакше завершуємо алгоритм.
5. Запам'ятовуємо поточний вміст ознаки X .
6. Виконуємо алгоритм 1.
7. Обчислюємо умовну ентропію після імпутації.
8. Якщо значення умовної ентропії зменшилось у порівнянні з попереднім кроком ітерації, переходимо на крок 9, інакше – переходимо на крок 11.
9. Виконуємо оновлення набору даних, прийнявши всі імпутації.
10. Збільшуємо номер ітерації та переходимо на крок 4.
11. Якщо значення умовної ентропії не зменшилось у порівнянні з попереднім кроком ітерації, виконуємо відкат до попередньої версії імпутації та завершуємо алгоритм.

Для наочності представимо діаграми діяльності UML обох алгоритмів на рисунку 1.

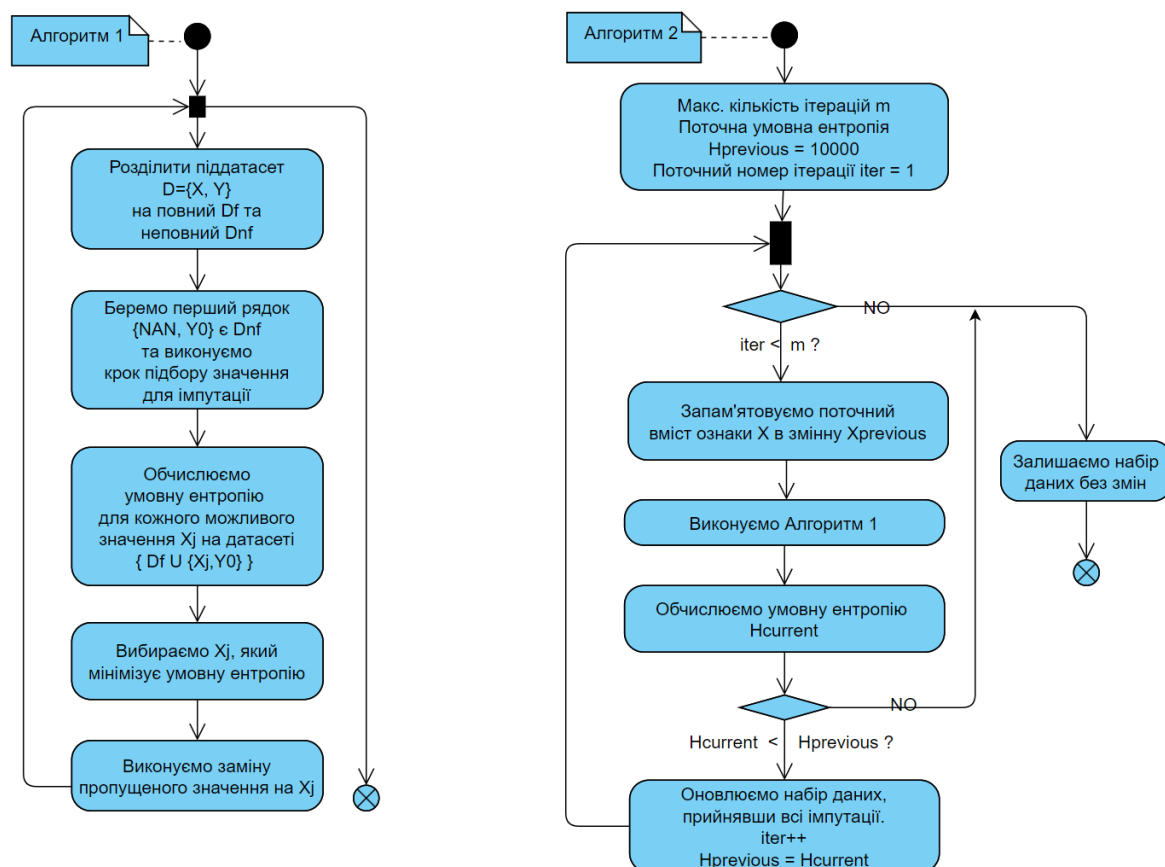


Рисунок 1 – Діаграми діяльності UML алгоритмів 1 та 2

В наших обчисленнях виявилось, що потрібно виконати максимум 3 – 4 ітерації, після чого умовна ентропія вже перестає зменшуватись.

Якщо наша ознака *якісна порядкова*, пропонується виконати крок перетворення у кількісну, описаний у в нашій попередній роботі [2]. Користь цих методів полягає в тому, що вони дають можливість перетворити всі якісні ознаки на кількісні без втрати інформації про пропуски. Після цього ми отримаємо випадок дискретної величини.

Якщо наша ознака відноситься до *неперервних* величин, ми можемо *звести* задачу до дискретного випадку, виконавши дискретизацію.

Пропонується *кількість* проміжків обчислювати за кількістю класів варіаційного ряду за формулою

$$M = 1 + 3.32 \cdot \lg(n), \quad (6)$$

де n – кількість рядків даних в множині повних даних D^f .

Таким чином, ми перейдемо знову до випадку дискретної величини. Для виконання зворотного перетворення, коли ми будемо з номеру імпутованого інтервалу отримувати значення, пропонується генерувати випадкову величину з нормального розподілу $N(m; \sigma)$, де m – середнє значення для ознаки X з D^f , σ – середньо-квадратичне відхилення цієї ознаки.

Цей алгоритм є жадібним до ресурсів. Продуктивність може бути покращена за рахунок використання інженерних технік, таких як багатопоточність, оскільки задача дуже гарно може бути розділена на підзадачі за рахунок того, що імпутування кожної ознаки може бути виконано незалежно від інших в окремому потоці. Крім того, можливі покращення за рахунок застосування технік, специфічних для обраної мови програмування, наприклад, векторизація обчислень в мові Python.

Аналіз результатів застосування методів. В якості наборів даних використовуються два набори даних UCI Heart Disease Data [9, 10] та Framingham Heart Study [11, 12], в яких представлено дані спостережень ішемічної хвороби серця. Обидва набори даних містять пропуски типу MAR або MCAR.

В наборі даних UCI Heart Disease Data 920 рядків, 13 інформативних ознак та один цільовий стовпець. Серед 13 ознак 5 ознак – кількісні неперервні, 8 ознак – якісні порядкові, серед яких 2 – бінарні. В ході первинного аналізу даних виявлено наявність пропусків. Тільки 299 рядків з 920 є повними (33%). Балансування відбулося за рахунок переходу від 4 класів до 2 класів, тобто до задачі бінарної класифікації.

В наборі даних Framingham Heart Study 4240 рядків, 14 інформативних ознак та один цільовий стовпець з двома класами. Серед 14 ознак 8 ознак – кількісні неперервні, 6 ознак – якісні порядкові, серед яких 5 – бінарні. В ході первинного аналізу даних виявлено наявність пропусків. 3658 рядків з 4240 є повними (86%). Датасет виявився дуже незбалансованим, причому кількість екземплярів одного класу перевищує кількість екземплярів іншого класу в 5 разів. Для балансування використовується метод SMOTE [13].

Для аналізу запропонованих підходів використовується три типи тестів. Перший та другий тести розглядають повний датасет, куди штучно вносяться пропуски, користуючись механізмом MCAR для генерування пропусків у даних. Для цього згенерованого датасету виконується імпутування за допомогою різних методів, оцінюється сере-

дньо-квадратична похибка імпутування та час виконання алгоритмів. Розглядаються 10% пропусків, 20%, 30%, 40%.

Другий тест вимірює умовну ентропію для кожного методу і виконує задачу тесту 1 ітераційно, поки зменшується ентропія для кожного методу. Ми виконуємо уточнення значень інтерполяції по кожній позиції, де був пропуск, при цьому вважаємо, що інші імпутовані значення належать датасету. Далі виконується порівняння значень умовної ентропії відносно початкової.

Третій тест стосується аналізу якості класифікації. Базовими моделями для порівняння якості класифікації виберемо дві моделі: отриману після навчання, використовуючи тільки повну частину датасету, та модель, де пропуски імпутовані найпростішим стандартним методом, наприклад, середнім значенням чи модою за допомогою *SimpleImputer* з бібліотеки *scikit-learn* [14] *python*. Виконуємо імпутування пропусків різними методами, навчаємо отримані датасети, отримуємо метрику асигасу (чи точність класифікації) на тестовій вибірці, порівнюємо значення цієї метрики для отриманих моделей. Для отримання стійких до зсуву результатів виконуємо процедуру розбиття датасетів на навчальну та тестові вибірки декілька разів, навчаємо моделі, отримуємо значення асигасу для поточної тестової вибірки, і наприкінці усереднюємо отримані значення. В третьому тесті для вихідних датасетів виконують імпутування пропущених даних різними методами, навчають моделі бінарної класифікації та порівнюють точність класифікації на тестовій вибірці. В якості методу класифікації використовується випадковий ліс (*Random Forest*).

В таблицях 2, 4 наведено результати тесту типу 1 для датасету *UCI Heart Disease Data* для алгоритмів 1 та 2 відповідно. В таблиці 3 та 5 представлено результати тесту типу 2 на тому ж датасеті для алгоритмів 1 та 2 відповідно. В таблиці 6 представлено результати запуску всіх методів ітераційно. В таблиці 7 ми наводимо результати тесту типу 3 для датасетів *UCI Heart Disease Data* та *Framingham Heart Study*.

Виконуємо тест першого типу для методів, аналогічно представленому переліку методів в роботі [2], куди додамо також наш новий метод на основі умовної ентропії. Візьмемо тільки повну частину датасету *UCI Heart Disease Data*. Згенеруємо 30% пропусків, додавши *NAN* у випадкові місця таблиці ознак. Оскільки методи на основі класифікації та регресії з роботи [2] перевіряють наявність патерну в пропусках, в умовах якого пропонується спеціальна реакція, що покращує результати методу, змодельуємо його наявність. Додамо патерн (рисунок 2), коли в нас є випадкова кількість повних стовпців (не більше половини), випадкова кількість повних рядків (не більше половини) та випадкова кількість рядків, в яких є тільки один пропуск.

Нам потрібно спочатку виконати перетворення якісних ознак на кількісні. Для цього застосовуємо три варіанти енкодерів, запропонованих в роботі [2]. Для енкодеру *le_non_missing* назвемо отриманий новий набір даних *data_1*, для енкодеру *le_non_missing_frequent(ascending=True)* - *data_2*, для енкодеру *le_non_missing_frequent(ascending=False)* - *data_3*.

Для набору даних *data_1* виконуємо імпутування наступними методами: *SimpleImputer* – стандартне заповнення значенням з найбільшою частотою, *IterativeIm-*

`puter(sample_posterior = False), kNNImputer, nonaImputer, fillna_k_columns, fillna_k_sorted_columns(ascending = True), fillna_k_sorted_columns(ascending = False), fillna_2steps_rg_class, fill_entropy_based_1step (алгоритм 1)`

Для набору даних `data_2` виконуємо такий метод імпутування: `fillna_2steps_rg_class`.

Для набору даних `data_3` виконуємо такі методи імпутування: `fillna_2steps_rg_class, fillna_2steps_rg(measure = 'value'), fillna_2steps_rg(measure = 'weight')`.

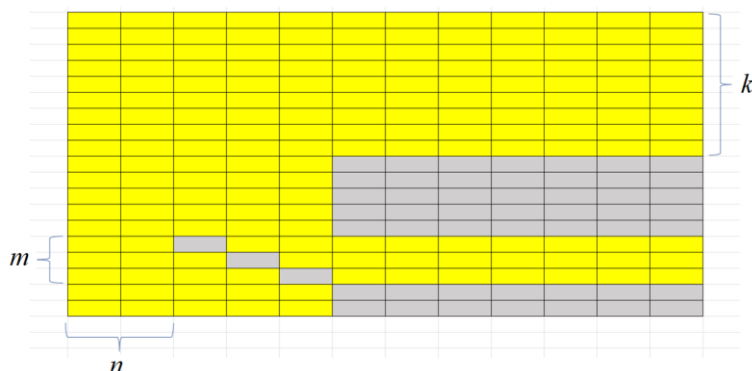


Рисунок 2 – Патерн для пропусків в даних (жовтий колір – наявні дані, сірий колір – пропуск у даних)

В таблиці 2 представлено результати тесту типу 1, де наводяться середньо-квадратичні похибки імпутування та час виконання алгоритмів. Нас цікавлять результати запропонованого методу на основі ентропійного підходу у порівнянні з вже проаналізованими в попередній роботі.

З аналізу цього прикладу бачимо, що метод працює довше, ніж інші методи, але не набагато, хоча з точки зору метрики середньо-квадратичного відхилення він значно програє всім іншим методам.

Таблиця 2

Результат тесту типу 1 в порівнянні з імпутуванням за допомогою алгоритму 1 (ентропійний неітераційний)

Назва методу	Імпутування 30% пропусків	Час виконання
<i>data_1</i>		
SimpleImputer	3.47	0.008939
IterativeImputer	2.46	0.162614
kNNImputer	2.75	0.012442
NonaImputer	2.54	1.082877
fillna_k_columns	2.55	1.083617
fillna_k_columns SortedAsc	2.62	1.080535
fillna_k_columns SortedDesc	2.51	1.071694

2steps	2.51	1.934818
entropyBased (алгоритм 1)	9.59	2.345332
<i>data_2</i>		
2stepsEncTrue	2.54	1.917507
<i>data_3</i>		
2stepsEncFalse	2.55	1.907776
2stepsRegValue	2.54	0.159464
2stepsRegWeight	2.55	0.154090

Порівняємо умовну ентропію для вказаних методів, тобто виконаємо тест типу 2. В таблиці 3 представлено значення умовної ентропії до та після імпутування. В якості прикладів візьмемо представників ознак різних типів (chol – це рівень холестерину, відноситься до неперервних величин, restecg – результат електрокардіографії у стані спокою, відноситься до якісних порядкових величин, sex – стать, відноситься до бінарних величин).

В таблиці 3 для набору даних data_1 жирним курсивом виділено базовий рядок зі значенням умовної ентропії до імпутування, далі жирним виділено рядок зі значеннями умовної ентропії для найкращого методу (на основі ентропії) та найгіршого (Simple). Стрілочками ввєрх чи вниз помічено яскраві приклади збільшення чи зменшення ентропії відносно базового значення.

З таблиці 3 можна побачити, що майже всі методи імпутування зменшують базову ентропію, але значний виграш дає саме метод, що оснований на ентропійному підході, причому для неперервної величини цей виграш найбільший і найменший – для бінарної величини.

Таблиця 3

Зміна значення умовної ентропії після застосування методів імпутування в порівнянні з алгоритмом 1

Назва методу	chol	restecg	sex
<i>data_1</i>	0.786876	0.968004	0.967503
SimpleImputer	0.795206 ↑	0.975178 ↑	0.96852 ↑
IterativeImputer	0.781841	0.838706	0.952813
kNNImputer	0.785193	0.783078	0.952813
NonaImputer	0.784556	0.957632	0.967657
fillna_k_columns	0.784556	0.957632	0.967657
fillna_k_columns SortedAsc	0.781794	0.958821	0.96852
fillna_k_columns SortedDesc	0.795206	0.96003	0.967657
2steps	0.783231	0.961332	0.96852
entropyBased	0.594163 ↓	0.870099 ↓	0.96741 ↓

(алгоритм 1)			
data_2	0.786876	0.968004	0.967503
2stepsEncTrue	0.78362 ↓	0.961307 ↓	0.96852 ↓
data_3	0.786876	0.968004	0.967503
2stepsEncFalse	0.78362	0.96013	0.96852
2stepsRegValue	0.783262 ↓	0.96133	0.967657 ↓
2stepsRegWeight	0.783262	0.961302 ↓	0.967657

Далі наведемо результати імпутування за допомогою алгоритму 2 (ентропійний ітераційний метод) у порівнянні з SimpleImputer та IterativeImputer (таблиця 4).

Таблиця 4

Результат тесту типу 1 в порівнянні з імпутуванням за допомогою алгоритму 2 (ентропійний ітераційний)

Назва методу	10% пропусків	20% пропусків	30% пропусків	40% пропусків	Час виконання
SimpleImputer	1.576	3.03	2.27	3.15	0.039279
IterativeImputer	1.238	2.30	1.79	2.76	0.491748
entropyBased (алгоритм 2)	5.634	8.9	6.77	9.66	16.97883

В таблиці 5 представлено кроки зменшення умовної ентропії на кожній ітерації алгоритму 2, час виконання алгоритму 11.73 мс.

Таблиця 5

Кроки зменшення умовної ентропії на кожній ітерації алгоритму 2

Номер ітерації	trestbps	chol	thalach	ca
0	0.6140	0.683	0.526	0.779
1	0.5764	0.654	0.494	0.709
2	0.5760	-	0.492	0.708
3	-	-	-	0.706
4	-	-	-	0.705

Бачимо, що вже на 4-й ітерації ентропія перестала зменшуватись для всіх ознак.

Далі проведемо тест другого типу, коли ми всі методи запускаємо ітераційно, поки зменшується умовна ентропія або не буде досягнуто максимальне значення кількості ітерацій. Наведемо результати для декількох представників різних типів ознак та для датасету data_1 (таблиця 6).

Зміна значення умовної ентропії
після застосування методів імпутування ітераційно

Назва методу	chol	restecg	sex
<i>data_1</i>	0.791235 0.790242 ↓	0.970871 0.972011 ↑	0.967796 0.96585 ↓
SimpleImputer	0.796228 ↑ 0.753609	0.975102 ↑ 0.972962	0.967657 ↑ 0.970045
IterativeImputer	0.791185 0.780171	0.938441 0.959821	0.949335 0.830418
kNNImputer	0.789338 0.783518	0.922918 0.959821	0.949519 0.820332
NonaImputer	0.789553 0.779655	0.968682 0.971386	0.967657 0.968787
fillna_k_columns	0.789553 0.779655	0.967621 0.971386	0.967657 0.968787
fillna_k_columns SortedAsc	0.787801 0.77828	0.964656 0.971386	0.967657 0.970045
fillna_k_columns SortedDesc	0.795379 0.782818	0.964656 0.971386	0.967657 0.968787
2steps	0.79057 0.778978	0.965685 0.971386	0.967657 0.969433
entropyBased	0.622744 ↓ 0.613428 ↓	0.875675 ↓ 0.961728 ↑	0.960461 ↓ 0.976096 ↑

З таблиці 6 бачимо результати, аналогічні наведеним в таблиці 3, тобто Simple-метод не зменшує ентропію, інші методи її зменшують, але більший вииграш з точки зору зменшення ентропії дає саме запропонований ентропійний метод. Також можна побачити, що ітерації уточнення майже не покращують результат, а навіть на другій ітерації можуть його погіршувати з точки зору умовної ентропії.

Далі розглянемо тест типу 3, який стосується аналізу якості класифікації. Базовими моделями для порівняння виступають дві моделі: отримана після навчання, використовуючи тільки повну частину датасету, та модель, де пропуски поповнені найпростішим стандартним методом, найчастішим значенням за допомогою *SimpleImputer* з бібліотеки *scikit-learn* [14] python. Будемо імпутувати пропуски різними методами, далі навчати класифікатор Random Forest та оцінювати метрику ассурасу (чи точність класифікації) на тестовій вибірці. Значення цієї метрики будемо використовувати для порівняння отриманих моделей. Виконуємо процедуру розбиття датасетів на навчальну та тестові вибірки декілька разів, навчаємо моделі, отримуємо значення ассурасу для поточної тестової вибірки, і наприкінці усереднюємо отримані значення. Така процедура

дозволить отримати стійкі до зсуву результати. В таблиці 7 жирним курсивом виділено базові результати та жирним шрифтом переможців за метрикою ассурасу. Також жирним виділено результати ентропійного методу.

Таблиця 7

Результати тесту типу 3

Назва методу	Датасет UCI HDD		Датасет Framingham FHS	
	Найкращі гіперпараметри моделі	Точність	Найкращі гіперпараметри моделі	Точність
<i>data_1</i>				
Baseline	{'max_depth': None, 'min_samples_leaf': 4, 'min_samples_split': 10, 'n_estimators': 150}	0.83	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 150}	0.88
SimpleImputer	{'max_depth': 10, 'min_samples_leaf': 4, 'min_samples_split': 10, 'n_estimators': 100}	0.82	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 100}	0.87
IterativeImputer(sample_posterior=False)	{'max_depth': 10, 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 50}	0.88	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 100}	0.90
IterativeImputer(sample_posterior=True)	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 10, 'n_estimators': 150}	0.82	{'max_depth': 20, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 100}	0.89
kNNImputer	{'max_depth': 10, 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 100}	0.84	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 100}	0.90
nonaImputer	{'max_depth': 10, 'min_samples_leaf': 2, 'min_samples_split': 5, 'n_estimators': 100}	0.89	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 150}	0.89
fillna_k_columns	{'max_depth': 10, 'min_samples_leaf': 2, 'min_samples_split': 5, 'n_estimators': 100}	0.89	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 150}	0.89
fillna_k_columns SortedAsc	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 100}	0.89	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 150}	0.89

fillna_k_columns SortedDesc	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 10, 'n_estimators': 100}	0.91	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 150}	0.89
2steps	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 150}	0.91	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 150}	0.88
entropyBased (алгоритм 2)	{'max_depth': None, 'min_samples_leaf': 2, 'min_samples_split': 10, 'n_estimators': 50}	0.90	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 150}	0.89
<i>data_2</i>				
2stepsEncTrue	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 100}	0.89	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 150}	0.88
<i>data_3</i>				
2stepsEncFalse	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 10, 'n_estimators': 100}	0.89	{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 150}	0.89
2stepsRegValue	{'max_depth': None, 'min_samples_leaf': 2, 'min_samples_split': 5, 'n_estimators': 150}	0.89	{'max_depth': 20, 'min_samples_leaf': 1,'min_samples_split': 2, 'n_estimators': 150}	0.88
2stepsRegWeight	{'max_depth': None, 'min_samples_leaf': 4, 'min_samples_split': 2, 'n_estimators': 100}	0.91	{'max_depth': 20, 'min_samples_leaf':1, min_samples_split':2, 'n_estimators': 150}	0.90

З аналізу результатів класифікації випливає, що запропонований ентропійний метод дозволяє покращити точність класифікації на рівні з найкращими методами, а на датасеті UCI Heart Disease Data результат класифікації покращено суттєво, порівняно з базовим результатом з 83% до 90%, порівняно з найпростішим імплементацією - з 82% до 90%.

Висновки та перспективи подальших досліджень. У даній роботі запропоновано два алгоритми імпутування пропусків даних на основі ентропійного підходу: неітераційний (алгоритм 1) та ітераційний (алгоритм 2), ідея яких полягає в мінімізації умовної ентропії по кожній ознаці окремо.

Проведено три типи тестів на двох наборах даних. Аналіз показав, що запропоновані алгоритми є достатньо повільними у порівнянні з іншими методами і можуть бути покращені, наприклад, за рахунок багатопроекторного виконання, як описано в нашій

роботі [15]. Тест типу 1 показав, що запропоновані алгоритми не дають вигреш за метрикою середньо-квадратичного відхилення, але вони суттєво зменшують ентропію (тест типу 2). В той же час ці методи показують покращення результатів класифікації порівняно з базовими моделями (тест типу 3).

Таким чином, запропоновані методи імпутування на основі ентропійного підходу показали гарні результати та можуть претендувати на те, щоб стати додатковим інструментарієм дослідника для підвищення точності прийняття рішень, але потрібно виконувати подальші дослідження щодо оптимізації обчислень для збільшення швидкодії цих методів.

ЛІТЕРАТУРА / REFERENCES

1. Roderick J. A. Little, Donald B. Rubin. *Statistical Analysis with Missing Data*, 3rd Edition. -Wiley, 2019. - 464 p. ISBN: 978-0-470-52679-8
2. Земляний О.Д., Байбуз О.Г. Методи імпутування пропусків у даних про ішемічну хворобу серця // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 2(151). – Дніпро, 2024. – С.33 – 49. DOI: <https://doi.org/10.34185/1562-9945-2-151-2024-04>
3. Yoon, J., Jordon, J., & Schaar, M.V. (2018). GAIN: Missing Data Imputation using Generative Adversarial Nets. ArXiv, abs/1806.02920. DOI: <https://doi.org/10.48550/arXiv.1806.02920>
4. Gondara, L., & Wang, K. (2017). Multiple Imputation Using Deep Denoising Autoencoders. ArXiv, abs/1705.02737. DOI: <https://doi.org/10.48550/arXiv.1705.02737>
5. Stekhoven, D. J., & Bühlmann, P. (2012). MissForest — non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112-118. DOI: <https://doi.org/10.1093/bioinformatics/btr597>
6. Rusdah, D.A., Murfi, H. XGBoost in handling missing values for life insurance risk prediction. *SN Appl. Sci.* 2, 1336 (2020). DOI: <https://doi.org/10.1007/s42452-020-3128-y>
7. Deng, Y., & Lumley, T. (2023). Multiple Imputation Through XGBoost. *Journal of Computational and Graphical Statistics*, 33(2), 352–363. DOI: <https://doi.org/10.1080/10618600.2023.2252501>
8. Delavallade, Thomas & Dang, Thanh. (2007). Using Entropy to Impute Missing Data in a Classification Task. *Proceedings of the IEEE International Conference on Fuzzy Systems, FUZZ-IEEE'07, London, UK.* 1 - 6. DOI: 10.1109/FUZZY.2007.4295430
9. Janosi, Andras, Steinbrunn, William, Pfisterer, Matthias, and Detrano, Robert. (1988). Heart Disease. UCI Machine Learning Repository. <https://doi.org/10.24432/C52P4X>
10. UCI Heart Disease Data. Heart Disease Data Set from UCI data repository. – [Електронний ресурс]. – Режим доступу: <https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data>
11. Framingham Heart Study-Cohort (FHS-Cohort). – [Електронний ресурс]. – Режим доступу: <https://biolincc.nhlbi.nih.gov/studies/framcohort/>
12. Framingham heart study dataset. – [Електронний ресурс]. – Режим доступу: <https://www.kaggle.com/datasets/aasheesh200/framingham-heart-study-dataset>
13. N. V. Chawla, K. W. Bowyer, L. O.Hall, W. P. Kegelmeyer, “SMOTE: synthetic minority over-sampling technique,” *Journal of artificial intelligence research*, 321-357, 2002
14. Imputation of missing values in scikit-learn. – [Електронний ресурс]. – Режим доступу: <https://scikit-learn.org/stable/modules/impute.html#impute>
15. Земляний О.Д., Байбуз О.Г. Порівняння багатопроекторної та багатопоточної реалізації ентропійного підходу для імпутування пропусків у даних на мові програмування Python // Виклики та

Algorithms for data imputation based on entropy

Recent advancements in data imputation have focused on various machine learning techniques, including methods like mean, median, and mode imputation, along with more complex approaches like k-nearest neighbors (KNN) and multiple imputation by chained equations (MICE). Research into entropy-based methods offers a promising direction. This method minimizes uncertainty by selecting imputation values that reduce the overall entropy of the dataset.

The goal of this work is to develop an algorithm that imputes missing data by minimizing conditional entropy, thus ensuring that the missing values are filled in a way that preserves the relationships between the variables. The method is designed for both qualitative and quantitative data, including discrete and continuous variables, aiming to reduce uncertainty in classification tasks and enhance the performance of machine learning models.

The proposed algorithm is based on conditional entropy minimization, using entropy as a measure of uncertainty in data. For each incomplete row, the algorithm computes the conditional entropy for possible imputation values. The value that minimizes conditional entropy is selected, as it reduces uncertainty in the target variable. This process is iterated for each missing value until all missing data is imputed.

Three types of tests were performed on two datasets. The analysis showed that the proposed algorithms are quite slow compared to other methods and can be improved, for example, by multiprocessing, as described in our work [15]. The type 1 test showed that the proposed algorithms do not give a gain on the RMS deviation metric, but significantly reduce entropy (type 2 test). At the same time, these methods show an improvement in classification performance over the baseline models (type 3 test).

Thus, the proposed entropy-based imputation methods have shown good results and can be considered by researchers as an additional tool to improve the accuracy of decision making, but further computational optimisation studies are needed to improve the performance of these methods.

The algorithm shows promise in improving classification accuracy by selecting imputation values that minimize conditional entropy. Future research will focus on optimizing the method for large datasets and expanding its application to various domains.

Земляний Олексій Дмитрович – аспірант 3-го року навчання Дніпровського національного університету імені Олеся Гончара; 380638802605; zemlyanoy.aleksey@gmail.com.

Байбуз Олег Григорович – доктор технічних наук, професор, завідувач кафедри математичного забезпечення ЕОМ Дніпровського національного університету імені Олеся Гончара; тел. +380672828822; olegbaybuz68@gmail.com.

Zemlianyi Oleksii Dmytrovych – 3rd year postgraduate student at Oles Honchar Dnipro National University; 380638802605; zemlyanoy.aleksey@gmail.com.

Baibuz Oleh Grigorovich - – Doctor of Technical Sciences, Professor, Head of the Department of Mathematical Support for Computers at Oles Honchar Dnipro National University; tel. +380672828822; olegbaybuz68@gmail.com.

**ВИКОРИСТАННЯ НІТРИД-АЛЮМІНІЄВИХ КЕРАМІЧНИХ ПЛАТ
ДЛЯ ЗАБЕЗПЕЧЕННЯ ТЕМПЕРАТУРНОЇ СТАБІЛЬНОСТІ
ВИМІРЮВАЛЬНИХ ПІДСИЛЮВАЧІВ**

Анотація. Актуальність теми дослідження. У статті розкрито принципи використання нітрид-алюмінієвих керамічних плат для забезпечення температурної стабільності вимірювальних підсилювачів. Зі збільшенням потужності електронних компонентів, що вимагають ефективного розсіювання тепла, виникає необхідність розробки матеріалів із високою теплопровідністю. Одним із таких матеріалів є нітрид алюмінію. Постановка проблеми. Попри широке застосування керамічних плат їх використання при розробці ширококутових високовольтних підсилювачів ще не достатньо вивчено, залишаються прогалини в методах одержання стабільного керамічного композиту на основі AlN. Мета статті полягає у одержанні температурно стабільних нітрид-алюмінієвих керамічних плат, які забезпечили б температурну стабільність вимірювальних підсилювачів. Методи дослідження. Аналіз, експеримент.

Результати. У статті представлені результати синтезу та технологія виготовлення кераміки із нітриду алюмінію. Встановлено, що введення у вихідну суміш різних добавок підвищує токсичність газоподібних викидів, ускладнює роботу і призводить до збільшення вартості нітриду алюмінію. Отримано вільним спіканням великогабаритні деталі (60×70×5 мм) теплопровідністю 160 Вт/(м·К) з температурно стабільного керамічного композиту на основі AlN для використання при розробці ширококутового високовольтного підсилювача. Доведено, що найбільш ефективно видалення кисню з кристалічної решітки нітриду алюмінію досягається при еквівалентному співвідношенні оксиду ітрію в кількості 5%. Вивчено вплив добавки оксиду ітрію на теплопровідність функціонального композиту з керамічною матрицею на основі нітриду алюмінію, одержаного вільним спіканням. Висновки. На підставі проведених досліджень зразків та аналізу отриманих даних можна стверджувати, що при використанні ітераційної корекції вплив статичної похибки може бути скомпенсований. Для використання ітераційної корекції система має бути стабільною і її параметри не повинні змінюватись у часі. Запропоновано раціональний метод стабілізації параметрів для забезпечення спільного та однорідного температурного поля що буде впливати однаково на всі компоненти шляхом використання високо теплопровідних матеріалів.

Ключові слова: нітридні алюмінати; керамічні плати; температурна стабільність; підсилювачі; оксид ітрію; оксид алюмінію; теплопровідність; високовольтний підсилювач.

Постановка проблеми. Порівняно з органічними ламінатами, кераміка AlN забезпечує чудову теплопровідність як ізолятор, зберігаючи при цьому коефіцієнт теплового розширення, близький до кремнію. Це дозволяє платам AlN без збоїв витримувати високі температури та рівні потужності. Силкові електронні модулі вітчизняного виробництва містять у конструкції елементи з'єднань метал – кераміка. Ці зв'язки потрібні для адгезійної міцності, стійкості до зміни температури. Типові сфери застосування: системи навігації літальних апаратів, елементи комутації силових схем та ін. На рис. 1 зображена типова комутаційна плата з провідними доріжками з нітриду алюмінію.

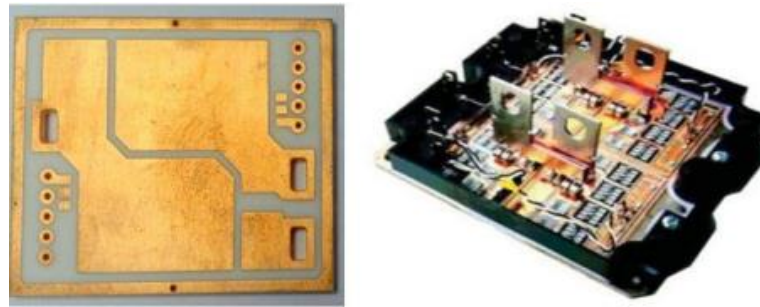


Рисунок 1 - Типова комутаційна плата з провідними доріжками та зовнішній вигляд силового модуля

При побудові високовольтних підсилювачів необхідно акцентувати увагу на їх температурному режимі та відводі тепла. Серед доступних нешкідливих керамічних плат найкращою є виготовлена на основі нітриду алюмінію кераміка «AlN», яка має теплопровідність 100-155 Вт/(м·К). Це на два порядки більше теплопровідності кераміки широкого застосування «СІТАЛЛ» [12].

Підкладки з нітриду алюмінію – найбільш перспективний кандидат для покращення розсіювання тепла в електронній техніці. Підкладки з AlN мають декілька відмінних характеристик, таких як висока теплопровідність (у 10 разів вища, ніж в оксиду алюмінію Al_2O_3), коефіцієнт теплового розширення близький до коефіцієнта теплового розширення кремнію, високі електроізоляційні властивості, низька діелектрична постійна, хороші механічні властивості, не токсичність, екологічна чистота та відносно низька вартість (питома вартість у 5-7 разів нижча питомої вартості кераміки з оксиду берилію) [10]. Нітрид-алюмінієві керамічні плати мають високу теплопровідність та за своїми параметрами придатні для використання.

Попри широке застосування керамічних план їх використання при розробці широкосмугових високовольтних підсилювачів ще не достатньо вивчено. Залишаються прогалини в методах одержання стабільного керамічного композиту на основі AlN.

Аналіз останніх досліджень та публікацій. Окреслена наукова проблематика з різних сторін представлена у роботах багатьох сучасних науковців. Так, наприклад, особливості виробництва керамічних плат з нітриду алюмінію та можливості їх застосування у різних пристроях представлені у статтях таких дослідників як Дж. Раухенкер, Т. Конеггер [12], Г. Окада, К. Фукуда, С. Касап, Т. Янагіда [10], М. Сігнор, Г. Решіо, К. ДеПаскалі, В. Яковаччі, П. Даріо, А. Леоне, Ф. Куаранта, Л. Франсіозо [14].

Основні динамічні фізико-хімічні властивості, зокрема параметри теплопровідності кристалічних ліній, тонких (товстих) плівок та керамічних плат типу AlN, а також можливості їх застосування у різних сферах виробництва електроніки висвітлено у роботах наступних науковців: Ю. Туз, О. Козир, А. Порхун [2], Ю. Чень, Х. Сон, Д. Лі, Х. Сунь, Х. Цзян, Г. Мяо, Ю. Чжоу [4], З. Ченг, Ю. Ко, А. Мамун, Т. Бай К. Хейн Л. Єйтс С. Грем [5], Н. Кім, М. Яралі, М. Морадня, М. Акіб, С. Ляо, Ф. Аль-Катар М. Нонг, Дж. Рьо [7], Ю. Ко, З. Ченг А. Мамун, З. Лю Т. Бай, К. Хусейн, П. Хопкінс [8], М. Негер, М. Герман, О. Фабрішна, Д. Павлючков, Х. Зайферт [9], С. Пандіт, М. Шнайдер, С. Бергер, С. Шварц, У. Шмід [11], Х. Ши, В. Лі, В. Као, Ю. Чуанг, Р. Лінь, Х. Лінь, М. Шюдзірі, М. Чен [13], М. Сігнор, Л. Веларді, К. Депаскалі, І. Кузнецова, Л. Бласі, Ф. Біскалія, Ф. Кваранта, Л. Франсіосо [15], Р. Сюй, М. Рохо, С. Іслам, А. Суд, Б. Варешкіч, А. Катре, Н. Мінго, Е. Поп [17].

Кераміка на основі нітриду алюмінію (AlN) все більше застосовується для відведення тепла у нових приладах електротехніки та електроніки [1; 2; 6]. Її отримують вільним спіканням з додаванням активаторів [5; 8; 16; 17]. Зазвичай використовують деталі розміром до 20 мм та товщиною 1 мм [11]. Сприяти підвищенню теплопровідності нітриду алюмінію можна за рахунок введення як активаторів спікання і розкислювання фтористих сполук рідкоземельних металів [7], вуглецю [14] або активних металів [4]. Однак ці методи вирішення задачі підвищення температури провідності призводять до погіршення діелектричних властивостей, зниження питомого електроопору та важливих експлуатаційних характеристик виробів із нітриду алюмінію. Вони недостатньо технологічні, неекономічні та не дозволяють отримувати високу теплопровідність кераміки. Крім того, введення у вихідну суміш різних добавок підвищує токсичність газоподібних викидів, ускладнює роботу і призводить до збільшення вартості нітриду алюмінію.

Мета статті полягає у одержанні температурно стабільних нітрид-алюмінієвих керамічних плат, які забезпечили б температурну стабільність вимірювальних підсилювачів

Виклад основного матеріалу дослідження. Проведено серію експериментів із дослідження нанесення покриттів у температурному режимі. При цьому відстань від підкладки до підкладки становила 75 мм. Температуру підкладки вимірювали за допомогою термопари через кожні 20 секунд при досягненні температури 220°C і вище, причому на початковій стадії швидкість нагрівання сягала температури 40° С / хв. У процесі відпрацювання режимів спостерігалися випадки руйнування зразків. Оскільки нанесення відбувалося з гарячої підкладки, нагрітої до температури 1000° С, то висловлено припущення, що руйнація відбувається через тепловий вплив. Виявлено, що завдяки дифузному розсіюванню матеріалу через високий тиск у зоні нанесення відбувається запилення зворотного боку підкладки. Висока швидкість нагрівання на початковій стадії призводить до руйнування плівок (рис. 2).



Рисунок 2 - Зовнішній вигляд зразків, що піддавалися перегріву та зруйнованих у процесі нанесення покриття

Пошук нових підходів і принципів побудови пристроїв, які мають зазначені параметри, є важливим і актуальним завданням. У табл. 1 порівнюються типові властивості трьох керамічних матеріалів, які в даний час використовуються в електронних приладах, підкладках.

Таблиця 1

Електрофізичні характеристики керамічних матеріалів з оксиду алюмінію (Al_2O_3), карбиду кремнію (SiC), нітриду алюмінію (AlN) та оксиду берилію (BeO)

Фізичні властивості	Al_2O_3 [9]	SiC [18]	AlN [15]	BeO [3]
Теплопровідність при температурі 25°C, Вт / мК	20...30	150	160...240	220...240
Міцність на вигин, МПа	300...350	320...350	250...300	170...230
Електрична міцність при температурі 25°C, кВ/мм	14...18	0,07	14...17	10
Питомий електроопір при температурі 25°C, Ом см	$>10^{14}$	$>10^{14}$	$>10^{14}$	$>10^{14}$
Діелектрична постійна при температурі 25°C на частоті 1 МГц	8,5...10	40...42	8,8...8,9	6,5...6,7
Тангенс кута діелектричних втрат ($\times 10^{-4}$) на частоті 1 МГц	2...3	500	1...5	1...5
Коефіцієнт температурного лінійного розширення ($\times 10^{-6}/^\circ C$) у діапазоні температур 25...400°C (для кремнію $3,5...4,0 \times 10^{-6}/^\circ C$)	6,7...7,3	3,7	4,3...4,6	7,5...8,0
Щільність, г/см ³	3,97	3,21	3,26	3,01
Модуль Юнга, ДПа	378	460	343	378

У зв'язку з поєднанням унікальних фізичних властивостей, наприклад, високої теплопровідності, високого електричного опору та коефіцієнта теплового розширення (КТР), близького до кремнію, нітрид алюмінію (AlN) є ключовим матеріалом, що використовується як підкладка для електронних приладів високої потужності. Хоча теоретична теплопровідність кристалічного AlN при кімнатній температурі, досягає 319 Вт/мК, виміряні значення для полікристалічної кераміки AlN знаходяться в діапазоні

від найнижчого показника 17 Вт/мК до найвищого 285 Вт/мК. Цей діапазон є типовим для багатьох видів кераміки з полікристалічною структурою.

Відносно висока вартість керамічного матеріалу з нітриду алюмінію означає, що AlN, як правило, використовують вибірково, а не як повну заміну звичайних плат FR-4. Зазвичай його використовують у силовій електроніці, світлодіодах високої яскравості та аерокосмічній електроніці [13].

У таких нітридах алюмінію розподіл тепловиділення забезпечується на основі методу послідовних наближень. Спочатку задається поле температур і розподіляється парозміст. Після визначення основних фізичних констант щільності матеріалів, перерізів взаємодії для цих умов, потоку нейтронів за величиною j_0 , знаходять щільність теплового потоку q . Після першої ітерації розрахунок повторюється. До ітераційних методів відносять: стаціонарний метод простих ітерацій, метод Зейделя або Гауса-Зейделя, метод спряжених градієнтів та ін.

Створення технологічного і економічного методу отримання кераміки на основі нітриду алюмінію з високою теплопровідністю та температурною стабільністю можливе при використанні порошку з мінімальним рівнем домішок і мінімальною кількістю добавок, які активують спікання. Це забезпечить високу щільність, температурну стабільність та теплопровідність кераміки на основі нітриду алюмінію.

Аналіз літературних джерел показав, що проблеми спікання щільних матеріалів з нітриду алюмінію пов'язані, з одного боку, із ковалентним характером зв'язку Al-N, що визначає необхідність високих температур спікання, а з іншого – з наявністю у вихідних порошках домішки кисню, що при спіканні проникає в решітку нітриду алюмінію, на яких і відбувається розсіювання фотонів, тобто знижується температурна стійкість матеріалу (рис. 3). Для вирішення цієї проблеми використано добавку Y_2O_3 від 1 до 5% (за масою).

При взаємодії Y_2O_3 з розташованими на поверхні частинками нітриду Al_2O_3 утворюється рідка фаза (рис. 4), що інтенсифікує спікання при температурі 1760°C, а очищення решітки нітриду алюмінію від кисню (чим і досягається висока теплопровідність матеріалу) відбувається при температурі 1800°C. При цьому утворюється алюмоітрієвий гранат, який осаджується на поверхні.



Рисунок 3 - Залежність коефіцієнта теплопровідності від кількості Y_2O_3 спікаючої добавки

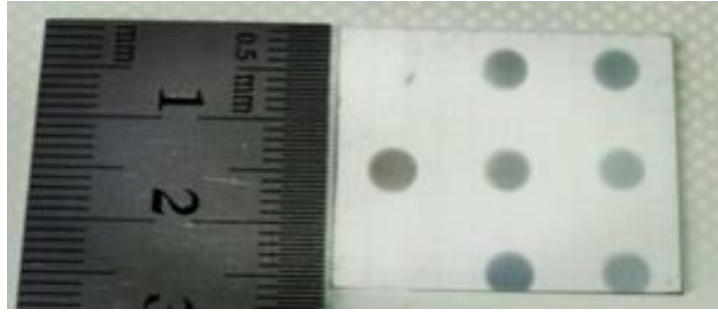


Рисунок 4 - Розпорошення задньої сторони підкладки під час покриття

На теплофізичні властивості алюмонітридної кераміки суттєво впливає метод формоутворення зразків. Метод пресування показав нерівномірну щільність, низькі електро- та теплофізичні характеристики взірців, а також низьке КПД і великі витрати при виготовленні підкладок розміром до 200×200 мм, тому надалі для проведення експериментальних робіт ми використали метод лиття на плівку. Мікроструктура одержаних зразків із вмістом 1 та 5 % (за масою) Y_2O_3 представлена на рис. 5.

З мікрофотографій видно, що вміст ітрій-алюмінієвого гранату (світла фаза в потрійних точках), який утворився при застиганні рідкої фази, приблизно однаковий. Вимірювалися електрофізичні параметри отриманих зразків та їх теплопровідність. Результати вимірювань коефіцієнта теплопровідності представлені на графіку (рис. 6).

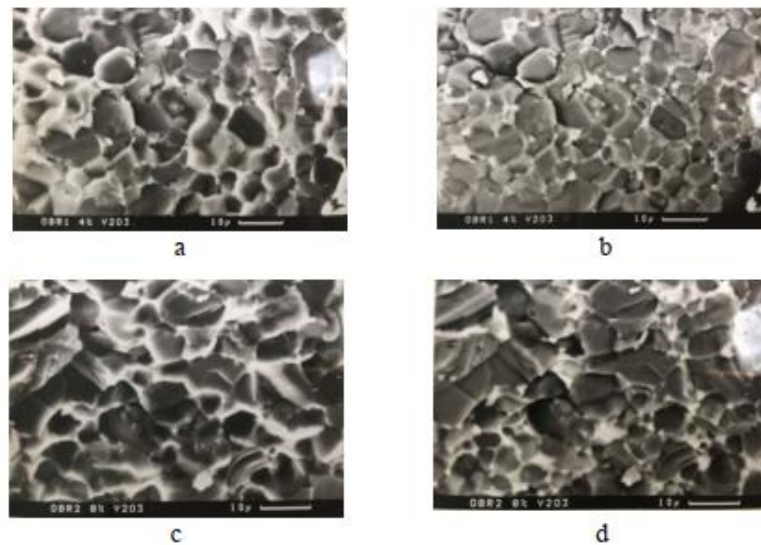


Рисунок 5 - Топологія (а, с) і фазовий контраст (b, d) зразків (злам) композиту на основі AlN, одержаних методом лиття на плівку із наступним спіканням, з різним вмістом у вихідній порошковій суміші Y_2O_3 :
1% (за масою) – а, b; 5% –с, d.

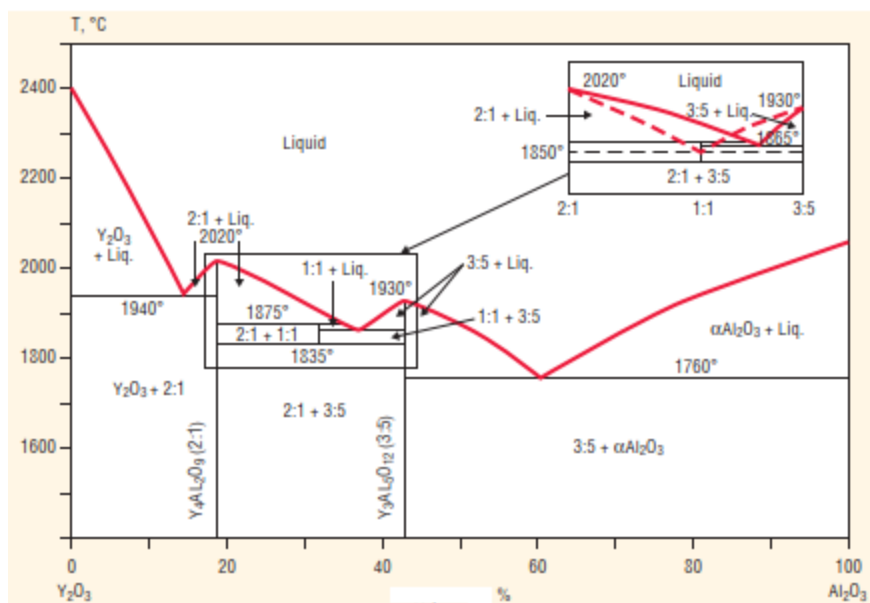


Рисунок 6 - Діаграма стану двофазної системи Y_2O_3 - Al_2O_3

У світловому теплообміні величина потоку, що падає на поверхню, залежить від кутового коефіцієнта (A), який залежить від відношення кута до нормалі поверхні. Розрахунки підтвердили, що при використовуваних режимах довжина вільного пробігу нанесеного матеріалу в 10 разів менша, ніж відстань до джерела підкладки:

$$L = \frac{0.63}{p} = \frac{0.63}{1} = 0.63 \text{ см, } 1/10 \text{ (1)}$$

Температура підкладки вимірювалась у різних її положеннях. Результати вимірювання температури для різних положень підкладки щодо мішені представлені на рис 7.

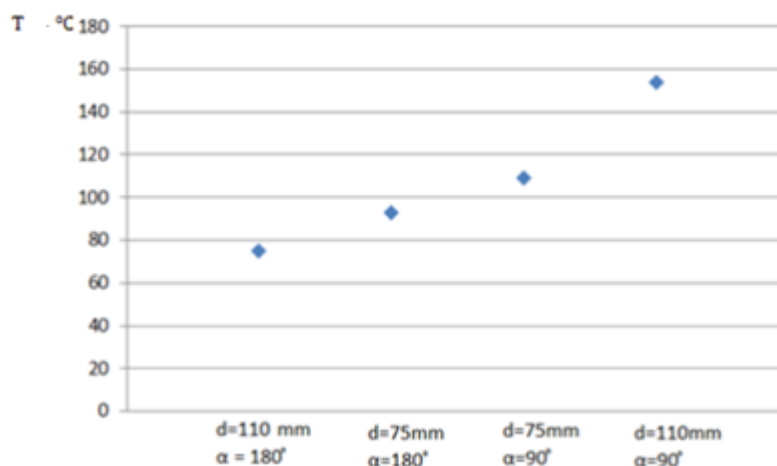


Рисунок 7 Залежність температури підкладки від кута та її розташування щодо мішені

Температура підкладки ($d = 110$ мм, $\alpha = 180^\circ$ становить 75° C), ($d = 75$ мм, $\alpha = 180^\circ$ становить 93° C), ($d = 75$ мм, $\alpha = 90^\circ$ становить 109° C) і ($d = 110$ мм, $\alpha = 90^\circ$

становить 153°C).

Похибка внаслідок температурного впливу має бути статичною, що спонукає використовувати методи температурної стабілізації. З точки зору особливостей конструкції та рівня потужності тепловиділення не доцільно використовувати термостати. Вольтметр з температурною корекцією і часовим розподілом каналів зображений на рис. 8.

Як показано в [13], похибки вольтметра включають адитивні і мультиплікативні похибки його блоків. Корекція похибок найбільш ефективна тоді, коли у вимірювальному каналі його мультиплікативні похибки переважають адитивні. У підсилювачів їх частотні похибки є мультиплікативними, в широкому діапазоні частот переважають адитивні. Саме тому в найбільш широкосмугових високочутливих вольтметрах застосовується мультиплікативна корекція.

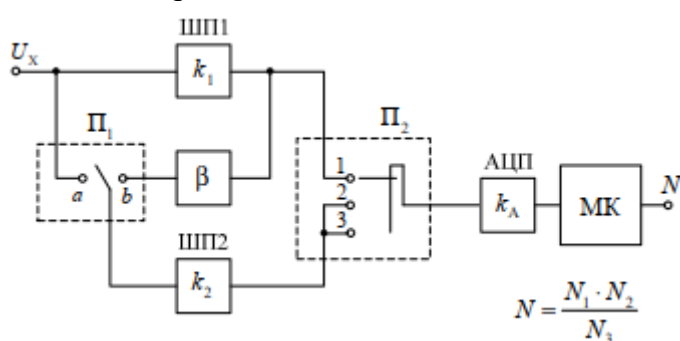


Рисунок 8 - Структурна схема вольтметра з температурною корекцією і часовим розподілом каналів

Рішення про тепловідведення з високочастотних підсилювачів з окремим віртуальним живленням не обмежується лише процесом розміщення компонентів на радіаторах. Через те, що окремі підсилювачі потребують значних за розмірами радіаторів для розсіювання тепла, однією з критичних проблем реалізації такого підсилювача є забезпечення тепловідведення з мінімальним ємнісним зв'язком між окремими силовими та сигнальними каскадами системи. Розроблений нами високовольтний підсилювач з керамічними ізоляційними деталями розміром $60 \times 70 \times 5$ мм складається із чотирьох мікросхем операційних підсилювачів Арех РА94 в корпусі DQ12 (рис. 9).



Рисунок 9 - Широкосмуговий високовольтний підсилювач з віртуальним живленням, який складається з чотирьох операційних підсилювачів, три з яких (ліворуч) розміщені на керамічних пластинах з нітриду алюмінію

Для усунення похибки підсилення можна використати методи ітераційної корекції. Для цього розміщуємо в кожному з трьох корпусів операційних мікросхем підсилювачі керамічної деталі з кераміки нітриду алюмінію розмірами $60 \times 70 \times 5$ мм і теплопровідністю $160 \text{ Вт}/(\text{м} \cdot \text{К})$. Кожна керамічна деталь з нітриду алюмінію, яка несе на собі операційний підсилювач, закріплена на загальному радіаторі з алюмінію. Загальний радіатор має розвинену задню поверхню і, крім того, відповідні отвори для примусового охолодження кожного операційного підсилювача окремим вентилятором. Отвір у загальному радіаторі і місце під деталь з нітриду алюмінію зображені на рис. 10.



Рисунок 10 - Вид радіатора, в якому розташоване місце під операційний підсилювач, та вид з боку охолодження

На підставі проведених досліджень зразків та аналізу отриманих даних можна стверджувати, що при використанні ітераційної корекції вплив статичної похибки може бути скомпенсований. Для використання ітераційної корекції система має бути стабільною і її параметри не повинні змінюватись у часі.

Висновки:

1. Керамічні плати AlN є особливим класом плат, які використовують унікальні переваги підкладок з нітриду алюмінію для керування температурним режимом, стійкістю та надійністю, особливо в складних умовах експлуатації.

2. Найбільш ефективно видалення кисню з кристалічної решітки нітриду алюмінію досягається при еквівалентному співвідношенні оксиду ітрію в кількості 5%. Для подальшого підвищення коефіцієнта теплопровідності та покращення теплофізичних характеристик зразків необхідне проведення додаткових досліджень у цьому напрямі.

3. Похибку підсилення усували методом ітераційної корекції, з розміщенням в трьох корпусах операційних мікросхем нітриду алюмінію розмірами $60 \times 70 \times 5$ мм і теплопровідністю $160 \text{ Вт}/(\text{м} \cdot \text{К})$.

4. Запропоновано раціональний метод стабілізації параметрів для забезпечення спільного та однорідного температурного поля що буде впливати однаково на всі компоненти шляхом використання високо теплопровідних матеріалів.

ЛІТЕРАТУРА

1. Теплопровідність надтвердих матеріалів / І.П. Фесенко, В.З. Туркевич, В.І. Часник та ін. – Корсунь-Шевченківський: ФОП Майдаченко І.В., 2018. – 68 с.

2. Туз Ю.М. Спосіб визначення динамічних характеристик термопар за допомогою радіоімпульсу струму / Туз Ю.М., Козир О.В., Порхун А.В. // Системи обробки інфор-

мації. – 2016. – №6. – С. 164-166.

3. Chen K. Ultrahigh Thermal Conductivity in Isotope-Enriched Cubic Boron Nitride / Chen K., Song B., Ravichandran N.K., Zheng Q., Chen X., Lee H., Sun H., Shi L. // *Science*. – 2020. – № 367. – PP. 555-559.
4. Chen Y. Influence of the Growth Temperature of AlN Nucleation Layer on AlN Template Grown by High-Temperature MOCVD / Chen Y., Song H., Li D., Sun X., Jiang H., Li Z., Miao G., Zhou Y. // *Mater. Lett.* – 2014. doi.org/10.1016/j.matlet.2013.09.096.
5. Cheng Z. Experimental Observation of High Intrinsic Thermal Conductivity of AlN / Cheng Z., Koh Y.R., Mamun A., Shi J., Bai T., Huynh K., Yates L., Graham S. // *Phys. Rev. Mater.* – 2020. – № 4. – PP. 044602.
6. Handbook of deposition technologies for films and coatings: science, applications and technology / ed. by P.M. Martin. – Burlington; Oxford: William Andrew / Elsevier, 2010. – 936 p.
7. Kim N.-I. Piezoelectric Sensors Operating at Very High Temperatures and in Extreme Environments Made of Flexible Ultrawide-Bandgap Single-Crystalline AlN Thin Films / Kim N.-I., Yarali M., Moradnia M., Aqib M., Liao C.-H., AlQatari F., Nong M., Ryou J.-H. // *Adv. Funct. Mater.* – 2023. – № 33. – P. 2212538.
8. Koh Y.R. Bulk-like Intrinsic Phonon Thermal Conductivity of Micrometer Thick AlN Films / Koh Y.R., Cheng Z., Mamun A., BinHoque M.S., Liu Z., Bai T., Hussain K., Hopkins P.E. // *ACS Appl. Mater. Interfaces*. – 2020. – № 12. – P. 29443–29450.
9. Neher M. Liquid Phase Formation in the System AlN-Al₂O₃-Y₂O₃. Part I. Experimental Investigations / Neher M., Hermann M., Fabrichnaya O., Pavlyuchkov D., Seifert H.J. // *J. Eur. Ceram. Soc.* – 2013. – vol. 33. – № 13-14. – PP. 2447-2455.
10. Okada G. Aluminum Nitride Ceramic as an Optically Stimulable Luminescence Dosimeter Plate / Okada G., Fukuda K., Kasap S., Yanagida T. // *Photonics*. – 2016. – vol. 3. – № 23. doi:10.3390/photonics3020023.
11. Pandit S. Impact of AlN Seed Layer on Microstructure and Piezoelectric Properties of Y_xAl_{1-x}N (x=15%) Thin Films / Pandit S., Schneider M., Berger C., Schwarz S., Schmid U. // *Adv. Electron. Mater.* – 2023. – № 9. – PP. 2200789.
12. Rauchenecker J. Additive Manufacturing of Aluminum Nitride Ceramics with High Thermal Conductivity Via Digital Light Processing / Rauchenecker J., Konegger T. // *Open Ceramics*. – 2022. – № 9. – P. 100-215. doi:10.1016/j.oceram.2021.100215.
13. Shih H.Y. Low-Temperature Atomic Layer Epitaxy of AlN Ultrathin Films by Layer-by-Layer *Situ* Atomic Layer Annealing / Shih H.Y., Lee W.H., Kao W.C., Chuang Y.C., Lin R.M., Lin H.C., Shiojiri M., Chen M.J. // *Sci. Rep.* – 2017. doi.org/10.1038/srep39717.
14. Signore M.A. Fabrication and Characterization of AlN-based Flexible Piezoelectric Pressure Sensor Integrated into an Implantable Artificial Pancreas / Signore M.A., Rescio G., DePascali C., Iacovacci V., Dario P., Leone A., Quaranta F., Francioso L. // *Sci. Rep.* – 2019. – № 9. – P. 17130.
15. Signore M.A. Effect of Silicon-based Substrates and Deposition Type on Sputtered AlN Thin Films: Physical & Chemical Properties and Suitability for Piezoelectric Device Integration / Signore M.A., Velardi L., DePascali C., Kuznetsova I., Blasi L., Biscaglia F., Quaranta

F., Francioso L. // *Appl. Surf. Sci.* – 2022. – № 599. – PP. 154017.

16. Tuz Y.M. Power Measurement at Wireless Transmission of Electricity in the Frequency Range from 0.01 to 10 MHz / Tuz Y.M., Vdovychenko A.V. // *Proc. International Conference on Precision Electromagnetic Measurements (CPEM)*. – Paris France, 2018. – PP. 8-13.

17. Xu R.L. Thermal Conductivity of Crystal line AlN and the Influence of Atomic-Scale Defects / Xu R.L., Muñoz Rojo M., Islam S., Sood A., Vareskic B., Katre A., Mingo N., Pop E. // *J. Appl. Phys.* – 2019. – № 126. – P. 185105.

18. Zheng Q. Thermal Conductivity of GaN, 71GaN, and SiC from 150 to 850 K / Zheng Q., Li C., Rai A., H. Leach J., Broido D.A., Cahill D.G. // *Phys. Rev. Mater.* – 2019. – № 3. – P. 014601.

REFERENCES

1. Thermal conductivity of superhard materials / I.P. Fesenko, V.Z. Turkevich, V.I. Chasnyk and others. – Korsun-Shevchenkivskiy: FOP Maidachenko I.V., 2018. – 68 p. [In Ukrainian].

2. Tuz Yu.M. The method of determining the dynamic characteristics of thermocouples using a radio current pulse / Yu.M. Tuz, O.V. Kozyr, A.V. Porhun. // *Information processing systems*. – 2016. – No. 6. – P. 164-166. [In Ukrainian].

3. Chen K. Ultrahigh Thermal Conductivity in Isotope-Enriched Cubic Boron Nitride / Chen K., Song B., Ravichandran N.K., Zheng Q., Chen X., Lee H., Sun H., Shi L. // *Science*. – 2020. – № 367. – PP. 555-559. [In English].

4. Chen Y. Influence of the Growth Temperature of AlN Nucleation Layer on AlN Template Grown by High-Temperature MOCVD / Chen Y., Song H., Li D., Sun X., Jiang H., Li Z., Miao G., Zhou Y. // *Mater. Lett.* – 2014. DOI: <https://doi.org/10.1016/j.matlet.2013.09.096>. [In English].

5. Cheng Z. Experimental Observation of High Intrinsic Thermal Conductivity of AlN / Cheng Z., Koh Y.R., Mamun A., Shi J., Bai T., Huynh K., Yates L., Graham S. // *Phys. Rev. Mater.* – 2020. – № 4. – PP. 044602. [In English].

6. Handbook of deposition technologies for films and coatings: science, applications and technology / ed. by P.M. Martin. – Burlington; Oxford: William Andrew / Elsevier, 2010. – 936 p. [In English].

7. Kim N.-I. Piezoelectric Sensors Operating at Very High Temperatures and in Extreme Environments Made of Flexible Ultrawide-Bandgap Single-Crystalline AlN Thin Films / Kim N.-I., Yarali M., Moradnia M., Aqib M., Liao C.-H., AlQatari F., Nong M., Ryou J.-H. // *Adv. Funct. Mater.* – 2023. – № 33. – P. 2212538. [In English].

8. Koh Y.R. Bulk-like Intrinsic Phonon Thermal Conductivity of Micrometer Thick AlN Films / Koh Y.R., Cheng Z., Mamun A., BinHoque M.S., Liu Z., Bai T., Hussain K., Hopkins P.E. // *ACS Appl. Mater. Interfaces*. – 2020. – № 12. – P. 29443–29450. [In English].

9. Neher M. Liquid Phase Formation in the System AlN-Al₂O₃-Y₂O₃. Part I. Experimental Investigations / Neher M., Hermann M., Fabrichnaya O., Pavlyuchkov D., Seifert H.J. // *J. Eur. Ceram. Soc.* – 2013. – vol. 33. – № 13-14. – PP. 2447-2455. [In English].

10. Okada G. Aluminum Nitride Ceramic as an Optically Stimulable Luminescence Dosimeter Plate / Okada G., Fukuda K., Kasap S., Yanagida T. // *Photonics*. – 2016. – vol. 3. – № 23.

DOI: 10.3390/photronics3020023. [In English].

11. Pandit S. Impact of AlN Seed Layer on Microstructure and Piezoelectric Properties of $YxAl_{1-x}N$ ($x=15\%$) Thin Films / Pandit S., Schneider M., Berger C., Schwarz S., Schmid U. // Adv. Electron. Mater. – 2023. – № 9. – PP. 2200789. [In English].

12. Rauchenecker J. Additive Manufacturing of Aluminum Nitride Ceramics with High Thermal Conductivity Via Digital Light Processing / Rauchenecker J., Konegger T. // Open Ceramics. – 2022. – № 9. – P. 100-215. DOI: 10.1016/j.oceram.2021.100215. [In English].

13. Shih H.Y. Low-Temperature Atomic Layer Epitaxy of AlN Ultrathin Films by Layer-by-Layer Situ Atomic Layer Annealing / Shih H.Y., Lee W.H., Kao W.C., Chuang Y.C., Lin R.M., Lin H.C., Shiojiri M., Chen M.J. // Sci. Rep. – 2017. DOI: <https://doi.org/10.1038/srep39717>. [In English].

14. Signore M.A. Fabrication and Characterization of AlN-based Flexible Piezoelectric Pressure Sensor Integrated Into an Implantable Artificial Pancreas / Signore M.A., Rescio G., DePascali C., Iacovacci V., Dario P., Leone A., Quaranta F., Francioso L. // Sci. Rep. – 2019. – № 9. – P. 17130. [In English].

15. Signore M.A. Effect of Silicon-based Substrates and Deposition Type on Sputtered AlN Thin Films: Physical & Chemical Properties and Suitability for Piezoelectric Device Integration / Signore M.A., Velardi L., DePascali C., Kuznetsova I., Blasi L., Biscaglia F., Quaranta F., Francioso L. // Appl. Surf. Sci. – 2022. – № 599. – PP. 154017. [In English].

16. Tuz Y.M. Power Measurement at Wireless Transmission of Electricity in the Frequency Range from 0.01 to 10 MHz / Tuz Y.M., Vdovychenko A.V. // Proc. International Conference on Precision Electromagnetic Measurements (CPEM). – Paris France, 2018. – PP. 8-13. [In English].

17. Xu R.L. Thermal Conductivity of Crystal Line AlN and the Influence of Atomic-Scale Defects / Xu R.L., Muñoz Rojo M., Islam S., Sood A., Vareskic B., Katre A., Mingo N., Pop E. // J. Appl. Phys. – 2019. – № 126. – P. 185105. [In English].

18. Zheng Q. Thermal Conductivity of GaN, 71GaN , and SiC from 150 to 850 K / Zheng Q., Li C., Rai A., H. Leach J., Broido D.A., Cahill D.G. // Phys. Rev. Mater. – 2019. – № 3. – P. 014601. [In English].

Received 30.09.2024.
Accepted 05.10.2024.

Use of aluminum nitride ceramic plates to ensure the temperature stability of measurement amplifiers

Analysis of recent research and publications. The outlined scientific issues are presented from different angles in the works of many modern scholars. For example, the peculiarities of the production of ceramic boards made of aluminum nitride and the possibilities of their use in various devices are presented in the articles of such researchers as J. Raueneker, G. Okada, Rauchenecker, T. Konegger, H. Okada, K. Fukuda, S. Kasap, T. Yanagida, M. Signor, G. Reschio, C. DePasquale, V. Iacovacci, P. Dario, A. Leone, F. Quaranta, L. Francioso.

The main dynamic physical and chemical properties, in particular, the thermal conductivity parameters of crystal lines, thin (thick) films and ceramic boards of the AlN type, as well as the possibilities of their application in various fields of electronics production, are

covered in the works of the following scientists: Y. Tuz, O. Kozyr, A. Porhun, Y. Chen, H. Song, D. Li, X. Sun, H. Jiang, G. Miao, Y. Zhou, Z. Cheng, Y. Ko, A. Mamun, T. Bai, K. Hein, L. Yates, S. Graham, N. Kim, M. Yarali, M. Moradnya, M. Aqib, S. Liao, F. Al-Qatar, M. Nong, J. Rhee, Y. Ko, Z. Cheng, A. Mamun, Z. Liu, T. Bai, K. Hussein, P. Hopkins, M. Neger, M. Herman, O. Fabrishna, D. Pavlyuchkov, H. Seifert, S. Pandit, M. Schneider, S. Berger, S. Schwartz, U. Schmid, H. Shi, W. Li, W. Kao, Y. Chuang, R. Lin, H. Lin, M. Shiojiri, M. Chen, M. Signor, L. Velardi, C. Depascali, I. Kuznetsova, L. Blasi, F. Biscaglia, F. Quaranta, L. Francioso, R. Xu, M. Rojo, S. Islam, A. Sud, B. Vareskic, A. Catre, N. Mingo, E. Pop.

The aim of the study is to obtain temperature-stable nitride-aluminum ceramic boards that would ensure the temperature stability of measuring amplifiers

Summary of the main material. The article presents the results of synthesis and manufacturing technology of aluminum nitride ceramics. It has been established that the introduction of various additives into the initial mixture increases the toxicity of gaseous emissions, complicates the work, and increases the cost of aluminum nitride. Large-sized parts (60×70×5 mm) with a thermal conductivity of 160 W/(m·K) from a temperature-stable ceramic composite based on AlN were obtained by free sintering for use in the development of a broadband high-voltage amplifier. It is proved that the most effective removal of oxygen from the crystal lattice of aluminum nitride is achieved at an equivalent ratio of yttrium oxide in the amount of 5%. The effect of yttrium oxide additive on the thermal conductivity of a functional composite with a ceramic matrix based on aluminum nitride obtained by free sintering was studied.

Conclusions. Based on the studies of the samples and the analysis of the data obtained, it can be argued that the effect of static error can be compensated for by using iterative correction. To use iterative correction, the system must be stable and its parameters must not change over time. A rational method of parameter stabilization is proposed to ensure a common and homogeneous temperature field that will affect all components equally by using highly thermally conductive materials.

Key words: nitride aluminates; ceramic boards; temperature stability; amplifiers; yttrium oxide; aluminum oxide; thermal conductivity; high voltage amplifier.

Довгаль Артем Володимирович – аспірант, кафедра інформаційно-вимірювальних технологій, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

Dovgal Artem Volodymyrovych – postgraduate student, Department of Information and Measurement Technologies, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

КОНСТРУКТИВНО-ПРОДУКЦІЙНЕ МОДЕЛЮВАННЯ СИСТЕМИ ЕЛЕКТРОПОСТАЧАННЯ ТЯГИ ПОСТІЙНОГО СТРУМУ

Анотація. Розроблена загальна конструктивно-продукційна модель ділянки електропостачання тяги постійного струму. Модель може використовуватись для рішення низки задач, пов'язаних зі зменшенням електроспоживання як на залізничному так і міському громадському електротранспорті. Розроблена модель орієнтована на визначення наявності і номенклатури обладнання тягових підстанцій та раціонального використання енергії рекуперації. Однак може бути застосована і для вирішення інших задач, пов'язаних з конструкціями системи тягового електропостачання. Наведено приклад сформованої схеми лінійної ділянки електропостачання з трьома підстанціями.

Ключові слова: конструктивно-продукційне моделювання, тяга постійного струму, формальні граматики, конструктор, програмне забезпечення, інформаційні технології.

Вступ. Вимоги сучасності щодо економічності та екологічності транспортних систем вимагають від країн розширювати свої залізничні системи, водночас використовуючи сучасні методи і технології енергозбереження. Основними технологіями енергозбереження в системах електричної тяги на сьогодні є впровадження енергоефективних режимів руху транспортних засобів і оптимізація типу розмірного ряду потужностей рухомого складу і тягових підстанцій, упровадження відновлювальних джерел, багатофункціональних перетворювачів і пристроїв зберігання електричної енергії в системи електропостачання та ін. При цьому одним з важливих резервів зниження енергоємності і покращення екологічних показників транспорту з електричною тягою є використання можливостей часткового відновлення витраченої на тягу енергії шляхом її рекуперації, що дозволяє зменшити споживання енергії на перевезення за різними оцінками на 10...40 % [1].

Аналіз останніх досліджень і публікацій. Поява сучасних залізничних технологій забезпечує рекуперативну гальмівну систему, яка може відновлювати значну частину енергії від роботи гальм. Управління рекуперативною енергією стає ключем до покращення роботи залізниць. Завдяки інтеграції графіка руху поїздів та системи зберігання енергії управління енергією може бути більш гнучким та ефективним, що призведе до покращення енергозберігаючої роботи [2].

Згідно з останніми дослідженнями, впровадження сучасних систем регенеративного гальмування значно підвищує ефективність роботи електричних залізничних транспортних систем [3]. Відновлення та зберігання значної частини енергії, втраченої

під час гальмування, набувають особливої важливості. Розвиток технологій зберігання енергії, таких як батареї та суперконденсатори, дозволяє ефективніше використовувати регенеративну енергію, що в свою чергу покращує екологічну стійкість залізничного транспорту [4].

Зміна графіка руху дозволяє краще синхронізувати споживання та зберігання енергії, що значно підвищує ефективність роботи залізничних систем. Такий підхід також допомагає зменшити пікове електроспоживання, що знижує загальні витрати на електроенергію [3, 5].

Розробка системи зберігання енергії для рекуперативного гальмування високошвидкісних залізничних систем є важливим кроком до підвищення енергоефективності та зниження витрат на електроенергію. Використання таких систем дозволяє зберігати відновлену енергію для подальшого використання під час руху поїздів [6, 7].

Рекуперативна енергія є критичною для підвищення енергоефективності та екологічної стійкості високошвидкісних залізничних систем. Інноваційні підходи до управління енергією дозволяють значно зменшити витрати на електроенергію та знизити викиди парникових газів, роблячи залізничний транспорт більш екологічно чистим та енергоефективним [8, 9].

Новий метод розподілу та повторного використання енергії рекуперативного гальмування в електропоїздах передбачає використання сучасного обладнання. Запропонована система інтегрує перетворювачі енергії, тягову мережу, електропоїзди та пристрої зберігання енергії на основі використання механізмів гнучкого розподілу для ефективного зберігання та використання кінетичної та потенційної енергії поїздів. Цей підхід підвищує енергоефективність та забезпечує стійке рішення для повторного використання гальмівної енергії поїздів [10].

1. Викладення основного матеріалу

1. Постановка задачі. На основі викладеного виникає задача розробки загальної моделі ділянки електропостачання постійного струму засобами конструктивно-продукційного моделювання, яка в свою чергу базується на засадах формальних грамастик. Модель має передбачати можливість енергозбереження за рахунок раціонального використання енергії рекуперації.

2. Основні положення конструктивно-продукційного моделювання. Наведемо основні поняття конструктивно-продукційного моделювання.

Узагальненим конструктором у парадигмі конструктивно-продукційного моделювання є трійка [11]:

$$C_G = \langle M, \Sigma, \Lambda \rangle, \quad (1)$$

де M – неоднорідний розширюваний носій, Σ – сигнатура операцій та відношень, яка включає операції зв'язування, підстановки та виведення, операції над атрибутами та відносини підстановки, Λ – множина тверджень, що складає інформаційне забезпечення конструювання (ІЗК). Загальні положення Λ викладені у [11].

Призначення конструкційно-продукційного моделювання (КПМ) полягає у формуванні множин потенційно можливих конструкцій та/або конструктивних процесів за допомогою операцій зв'язування, підстановки, виведення та інших, що задаються правилами у Λ .

У даній роботі конструкціями є структурні схеми системи тягового електропостачання постійного струму.

Для формування конструкцій необхідно виконувати низку уточнюючих перетворень узагальненого конструктора [11]:

– спеціалізація визначає предметну область: семантичну природу носія, скінчену множину операцій та їх семантику, атрибутику операцій, порядок їх виконання та обмеження на правила підстановки;

– інтерпретація полягає у зв'язуванні операцій сигнатури з алгоритмами виконання деякого алгоритмічного конструктора [12]. При інтерпретації виконується зв'язування інформаційної моделі засобів формування конструкцій та моделі виконавця, утворюючи конструктивну систему здатну автономно формувати конструкції;

– конкретизація полягає у розширенні аксіоматики множиною правил продукції, завданні конкретних множин нетермінальних та термінальних символів з їх атрибутами та, за необхідності, значень атрибутів;

– реалізація полягає у формуванні конструкції з елементів носія шляхом виконання алгоритмів, пов'язаних з операціями сигнатури. Реалізація можлива тільки для попередньо спеціалізованого, інтерпретованого та конкретизованого конструктора [11].

Відповідно до Λ узагальненого конструктора [11] формою ${}_w l$ з атрибутом w називається набір терміналів та нетерміналів, що поєднуються операціями зв'язування. Конструкцією називається форма, що містить лише термінали [11]. Терміналами є елементи для формування конструкцій, а нетермінали – поміжні абстрактні елементи, необхідні для завдання правил підстановки.

Правила підстановки мають вигляд $\psi_r : \langle s_r, g_r \rangle \in \Psi$, де s_r – набір відносини підстановки, g_r – набір операцій над атрибутами. Відношення підстановки – двомісне відношення з атрибутами. ${}_{w_i} l_i \rightarrow {}_{w_j} l_j$ [11]. Для форми

${}_{w_l} l_l = {}_{w_0} \oplus ({}_{w_1} l_1, {}_{w_2} l_2, \dots, {}_{w_h} l_h, \dots, {}_{w_k} l_k)$ та доступного відношення підстановки

${}_{w_h} l_h \rightarrow {}_{w_q} l_q$ такого, що ${}_{w_h} l_h \prec {}_{w_l} l_l$ (${}_{w_h} l_h$ є частиною ${}_{w_l} l_l$), результатом

тримісної операції підстановки ${}_{w_p} \Rightarrow ({}_{w_h} l_h, {}_{w_q} l_q, {}_{w_l} l_l)$ буде форма

${}_{w_l}^* l_l = {}_{w_0} \oplus ({}_{w_1} l_1, {}_{w_2} l_2, \dots, {}_{w_q} l_q, \dots, {}_{w_k} l_k)$ [11], де \oplus – будь-яка операція зв'язування з

Σ .

Операція часткового виводу полягає у:

– виборі одного з доступних правил підстановки $\psi_r : \langle s_r, g_r \rangle \in \Psi$, з відносинами підстановки s_r та виконання на його основі операції підстановки;

– виконання операцій над атрибутами g_r .

Форми, що отримуються в результаті операцій часткового виводу, починаючи з початкового нетерміналу, є сентенціальними.

Операція повного виводу (або просто виводу) полягає у послідовному виконанні операції часткового виводу, починаючи з початкового нетерміналу, і закінчуючи остаточною формуванням множини конструкцій, інколи єдиної.

3. Конструктивно-продукційна модель ділянки електропостачання. Визначимо спеціалізацію конструктору – моделі (структурної схеми) системи тягового електропостачання:

$$C = \langle M, \Sigma, \Lambda \rangle_s \mapsto C_{ZE} \langle M_{ZE}, \Sigma_{ZE}, \Lambda_{ZE} \rangle, \quad (2)$$

де $\Lambda_{ZE} = \Lambda \cup \Lambda_1 \cup \Lambda_2$, $\Lambda_1 = \{M_{ZE} \supset T_1 \cup N_1\}$, $\Sigma_{ZE} = \{\Xi, \Theta, \Phi\}$, $\Xi = \{^2, :^2, \times^2\}$, $\Phi = \{:=^1, +^2\}$, $\Theta = \{\Rightarrow, |\Rightarrow, ||\Rightarrow\}$, множини: Ξ – відносин (і відповідних операцій) зв'язування елементів конструкцій, Φ – операцій над атрибутами, Θ – операцій виведення, T_1, N_1 – терміналів та нетерміналів.

ІЗК Λ_2 містить визначення, доповнення та обмеження, які уточнюють алфавіт, атрибути носія, відносини підстановки, задають особливості виконання операцій підстановки та виведення.


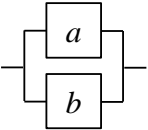
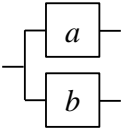
Термінальним алфавітом T_1 є множина елементів системи тягового енергопостачання, споживання та перетворення електроенергії в мережі постійного струму залізниць, трамвайних та тролейбусних ліній та метрополітенів з їх суттєвими для цього завдання властивостями – атрибутами. Елементи будемо позначати малими латинськими символами, наприклад, ${}_x a$ термінал a з атрибутом x , $x \dashv a$ атрибут x терміналу a .

Нетермінальний алфавіт $N_1 = \{\phi_i\}$ складається з багатьох допоміжних символів з атрибутами, що позначаються грецькими літерами.

Відношення та операції зв'язування Ξ задають послідовне і паралельне з'єднання елементів електричного кола (табл. 1)

Таблиця 1

Відповідність позначок відношень та структурних схем

Відношення	$a \cdot b$	$a : b$	$a \times b$
Структурна схема			

У загальному випадку відношення табл. 1 не мають властивостей асоціативності і комутативності, тому для однозначного визначення порядку виконання операцій зв'язування будуть застосовуватися дужки. У тих окремих випадках, де виконується асоціативність дужки, можуть бути опущені.

Операції над атрибутами: $:=(a,b)$ – присвоєння значення b змінній a ; $+(a,b,c)$ – присвоєння суми значень b і c змінній a .

Інтерпретація КПМ полягає у розширенні аксіоматики: $\forall \circ \in \Sigma_{ZE}$ задається $(A \stackrel{Y}{\leftarrow} \circ)$, тобто операції \circ присвоюється значення атрибуту як $A \stackrel{Y}{\leftarrow}$ – алгоритму деякого алгоритмічного конструктора C_A [12], в такий спосіб операції з сигнатури Σ пов'язуються з алгоритмами їх виконання

$$\begin{aligned} \langle C_{ZE} = \langle M_{ZE}, \Sigma_{ZE}, \Lambda_{ZE} \rangle, C_A = \langle M_A, \Sigma_A, \Lambda_A \rangle \rangle \quad I \mapsto \\ I \mapsto C_{ZEI} = \langle M_{ZE}, \Sigma_{ZE}, \Lambda_3, Z \rangle, \end{aligned} \quad (3)$$

де $M_A \supset \bigcup_{A_i^0 \in V_A} (X(A_i^0) \cup Y(A_i^0))$ – носій алгоритмічної структури, Σ_A – множина операцій зв'язування алгоритмів (у [12] визначено операції композиції та умовного виконання), Λ_A – ІЗК алгоритмічного конструктора (типова аксіоматика наведена в [12]), $V_A = \{A_i^0 \stackrel{Y_i}{\leftarrow} X_i\}$ – множина утворюючих алгоритмів алгоритмічного конструктора (множина елементарних алгоритмів, що реалізуються деяким виконавцем) $V_A \subset M_A$; X_i, Y_i – множина вхідних та вихідних даних алгоритму $A_i^0 \stackrel{Y_i}{\leftarrow} X_i$, відповідно. $\Lambda_3 = \Lambda_{ZE} \cup \Lambda_A$.

При інтерпретації виконується зв'язування інформаційної моделі способу побудови конструкцій та моделі виконавця.

Реалізація КПМ C_{ZEI} – схема ділянки системи електропостачання, тому термінали – елементи системи електропостачання зі своїми технічними характеристиками.

Конкретизація конструктивно-продукційної структури полягає у розширенні аксіоматики множиною конкретних правил продукцій та завданні конкретних множин терміналів та нетерміналів:

$$\begin{aligned} C_{ZEI} = \langle M_{ZE}, \Sigma_{ZE}, \Lambda_3, Z \rangle \quad K \mapsto \\ K \mapsto C_{ZEIC} = \langle M_{ZE}, \Sigma_{ZE}, \Lambda_4, Z \rangle, \end{aligned} \quad (4)$$

де $\Lambda_4 = \Lambda_3 \cup \Lambda_5$.

ІЗК Λ_5 містить наступні доповнення.

Термінали алфавіту T_1 :

– $_{U_{ES}} epss$ - зовнішня система електропостачання з атрибутом U_{ES} - рівень вхідної напруги системи тягового електропостачання;

– $\eta, k, U_{in}, U_{out} tr$ – понижуючий трансформатор з атрибутами η – к.к.д., k – коефіцієнт трансформації, U_{in}, U_{out} – вхідна та вихідна напруга;

– $\eta, U_{in}, U_{out} cr$ – випрямляч з відповідними атрибутами;

– $\eta, U_{in}, U_{out} aci$ – інвертор;

– $\eta, W, W_{max}, V ESD$ – накопичувач електроенергії з атрибутами: W – поточний рівень заряду накопичувача, W_{max} – максимальна електрична ємність, V – максимальна швидкість (час) заряду/розряду);

– $\rho, L, \bar{X}, n cl$ – тягова мережа з атрибутами: L – Довжина ділянки, ρ – питомий опір мережі, \bar{X} – Вектор розмірності n , визначає положення перемичок на багатокільцевих ділянках;

– $v, l, P tr$ – електрорухомий склад з атрибутами v – режим руху, l – розташування на міжпідстанційній зоні, P – потужність яка споживається або виділяється;

– $\eta, U_{in}, U_{out}, \bar{U}, n rn$ – регулятор напруги випрямляча, \bar{U} – вектор розмірності n , кожен елемент якого u_i – регульовані рівні вихідної напруги;

– sw – перемикач.

Крім зазначених вище атрибутів, всі термінали мають ідентифікуючий атрибут id , який позначається індексом терміналу або не терміналу і дозволяє однозначно визначити конкретне обладнання системи електропостачання.

Нетермінальний алфавіт N_1 складається із допоміжних елементів $N_1 = \{\sigma, \alpha, \beta, \delta, \gamma, \nu, \theta, \lambda\}$. Початковий нетермінал – σ .

ІЗК Λ_5 містить такі правила продукції.

Перше правило містить відношення, що зв'яже ділянку тягової мережі з двома підстанціями живлення:

$$\psi_1 = \left\langle \begin{array}{l} s_1 = \langle \sigma \rightarrow \alpha \cdot \beta \cdot \nu \rangle, \\ g_1 = \left\langle \begin{array}{l} = (index \downarrow \alpha = i), +(i, i, 1), \\ = (index \downarrow \beta = i), +(i, i, 1), \\ = (index \downarrow \nu = i), +(i, i, 1), \end{array} \right\rangle \end{array} \right\rangle, \quad (5)$$

Правило ψ_2 дозволяє рекурсивно доповнити модель наступною (сусідньою) ділянкою (ітеративно кількома ділянками):

$$\psi_2 = \langle s_2 = \langle \nu \rightarrow \alpha \times (\beta \cdot \nu) \rangle, g_2 = \langle \dots \rangle \rangle. \quad (6)$$

Позначка « \dots » у g_2 і у подальшому означає формування відповідних індексів id , як у g_1 .

Правило Ψ_3 призначене для завершення рекурсії.

$$\Psi_3 = \langle s_3 = \langle v \rightarrow \alpha \rangle, g_3 = \langle \dots \rangle \rangle. \quad (7)$$

Наступне правило дозволяє моделювати колійне розгалуження з відповідною тяговою мережею та підстанціями (без підстанції в місці розгалуження):

$$\Psi_4 = \langle s_4 = \langle v \rightarrow (\beta \cdot v) \times (\beta \cdot v) \rangle, g_4 = \langle \dots \rangle \rangle \quad (8)$$

або з підстанцією:

$$\Psi_5 = \langle s_5 = \langle v \rightarrow \alpha \cdot (\beta \cdot v) \times (\beta \cdot v) \rangle, g_5 = \langle \dots \rangle \rangle. \quad (9)$$

Ψ_6 - допоміжне правило для обладнання тягової підстанції

$$\Psi_6 = \langle s_6 = \langle \alpha \rightarrow (\delta \times \gamma) \rangle, g_6 = \langle \dots \rangle \rangle. \quad (10)$$

Правило s_7 призначене для включення накопичувача енергії з перемикачем у модель тягової підстанції, Ψ_8 – альтернатива правилу Ψ_7 : підстанція накопичувачем не обладнана:

$$\Psi_7 = \langle s_7 = \langle \delta \rightarrow_{\eta, W, W_{\max}} esd \cdot sw \rangle, g_7 = \langle \dots \rangle \rangle; \quad (11)$$

$$\Psi_8 = \langle s_8 = \langle \delta \rightarrow \varepsilon \rangle, g_8 = \langle \rangle \rangle. \quad (12)$$

Символ ε – порожньо, означає відсутність елемента на структурній схемі.

Правило Ψ_9 дозволяє включити в схему тягової підстанції живлення від зовнішньої енергосистеми та понижуючий трансформатор:

$$\Psi_9 = \langle s_9 = \langle \gamma \rightarrow (\theta : \lambda) \cdot_{\eta, U_{in}, U_{out}} tr \cdot_{U_{ES}} epss \rangle, g_9 = \langle \dots \rangle \rangle. \quad (13)$$

Два альтернативні правила Ψ_{10} і Ψ_{11} дозволяють включити в структурну схему випрямляч з регулятором напруги (Ψ_{10}) або без останнього (Ψ_{11}):

$$\Psi_{10} = \langle s_{10} = \langle \theta \rightarrow_{\eta, U_{in}, U_{out}} cr \rangle, g_{10} = \langle \dots \rangle \rangle; \quad (14)$$

$$\Psi_{11} = \langle s_{11} = \langle \theta \rightarrow_{\eta, U_{in}, U_{out}} cr \cdot_{\eta, U_{in}, U_{out}, \bar{U}, n} rn \rangle, g_{11} = \langle \dots \rangle \rangle. \quad (15)$$

Альтернативні правила Ψ_{12} та Ψ_{13} дозволяють включити в структурну схему інвертор з перемикачем (Ψ_{12}) або без них обох (Ψ_{13}):

$$\Psi_{12} = \langle s_{12} = \langle \lambda \rightarrow br \cdot_{\eta, U_{in}, U_{out}} aci \rangle, g_{12} = \langle \dots \rangle \rangle; \quad (16)$$

$$\Psi_{13} = \langle s_{13} = \langle \lambda \rightarrow \varepsilon \rangle, g_{13} = \langle \rangle \rangle. \quad (17)$$

Правило, що дозволяє моделювати електроживлення багатоколійних ділянок

$$\Psi_{14} = \langle s_{14} = \langle \beta \rightarrow \beta : \beta \rangle, g_{14} = \langle \dots \rangle \rangle \quad (18)$$

та з можливими перемичками на ділянці

$$\Psi_{15} = \langle s_{15} = \langle \beta \rightarrow \beta \cdot \beta \rangle, g_{15} = \langle \dots \rangle \rangle. \quad (19)$$

Правило Ψ_{15} дозволяє замінити нетермінал β на ділянку тягової мережі без потяга:

$$\Psi_{16} = \langle s_{16} = \langle \beta \rightarrow_{\eta, \rho, L, \bar{x}, n} cl \cdot \beta \rangle, g_{16} = \langle \dots \rangle \rangle \quad (20)$$

та з потягом у системі

$$\Psi_{17} = \langle s_{17} = \langle \beta \rightarrow_{v, l, P} tr \cdot \beta \rangle, g_{17} = \langle \dots \rangle \rangle; \quad (21)$$

Вивід завершується правилом Ψ_{18} яке позбавляє останнього не терміналу у сентенціальній формі, перетворюючи її на конструкцію:

$$\Psi_{18} = \langle s_{18} = \langle \beta \rightarrow \varepsilon \rangle, g_{18} = \langle \rangle \rangle. \quad (22)$$

4. Приклад моделі лінійної ділянки електропостачання з трьома тяговими підстанціями. В результаті операції часткового виведення ($\sigma \mid\Rightarrow l_1$), з вибором правила підстановки s_1 буде сформована сентенційна форма $l_1 = \alpha_1 \cdot \beta_2 \cdot \alpha_3$ і відповідна їй структурна схема, представлена на рис. 1.



Рисунок 1 – Структурне представлення сентенційної форми l_1

Продовжуючи підстановку $l_1 \mid\Rightarrow l_2 \mid\Rightarrow l_3$, із застосуванням правила s_2 , послідовно отримаємо сентенційні форми $l_2 = \alpha_1 \cdot \beta_2 \cdot (\alpha_4 \times (\beta_5 \cdot \nu_6))$ та $l_3 = \alpha_1 \cdot \beta_2 \cdot (\alpha_4 \times (\beta_5 \cdot \alpha_7))$ та структурну схему, із залагодженими заготовками двох ділянок тягової мережі (β) та трьох тягових підстанцій (α) (рис. 2).

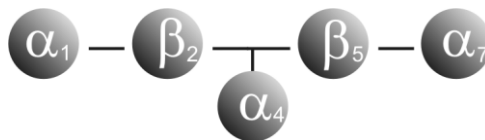


Рисунок 2 – Структурне представлення сентенційної форми l_3

Трохи відхилимось від послідовного викладення прикладу і продемонструємо яким чином можна моделювати розгалужену мережу електрифікованих ділянок колії.

Застосування правила s_4 до форми l_2 і двічі s_3 за операцією часткового виводу $l_2 \mid\Rightarrow l_3^* \mid\Rightarrow l_4^* \mid\Rightarrow l_5^*$ де $l_3^* = \alpha_1 \cdot \beta_2 \cdot (\alpha_4 \times (\beta_5 \cdot (\beta_7 \cdot \nu_8) \times (\beta_9 \cdot \nu_{10})))$, $l_4^* = \alpha_1 \cdot \beta_2 \cdot (\alpha_4 \times (\beta_5 \cdot (\beta_7 \cdot \alpha_{11}) \times (\beta_9 \cdot \nu_{10})))$, $l_5^* = \alpha_1 \cdot \beta_2 \cdot (\alpha_4 \times (\beta_5 \cdot (\beta_7 \cdot \alpha_{11}) \times (\beta_9 \cdot \alpha_{12})))$, дозволяє сформувати структурну схему, в якій закладено одне розгалуження колії (з контактною мережею), що живиться окремими тяговими підстанціями (рис. 3). Рекурсивне застосування правил s_4 (або s_5) та s_3 дозволяє моделювати будь яку розгалужену мережу електропостачання.

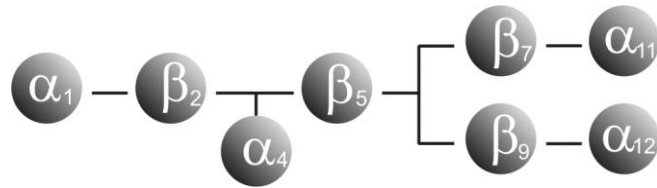


Рисунок 3 – Структурне представлення сентенційної форми l_5^*

Повернемося до нашого прикладу лінійної ділянки мережі електропостачання.

Продовжуючи підстановки ($l_3 \Rightarrow l_4$) за правилом s_6 отримаємо $l_4 = (\delta_8 \times \gamma_9) \cdot \beta_2 \cdot (\alpha_4 \times (\beta_5 \cdot \alpha_7))$ і структурну схему з основами конфігурації електрообладнання однією з тягових підстанцій (рис. 4).

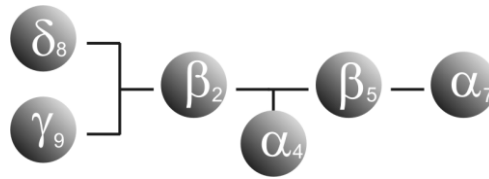


Рисунок 4 – Структурне представлення сентенційної форми l_4

Подальша підстановка $l_4 \Rightarrow l_5$ за правилом s_7 дозволяє отримати $l_5 = (esd_{10} \cdot sw_{11} \times \gamma_9) \cdot \beta_2 \cdot (\alpha_4 \times (\beta_5 \cdot \alpha_7))$ з додаванням до структурної схеми на рис. 4 накопичувачі електроенергії на відповідній тяговій підстанції (рис. 5).

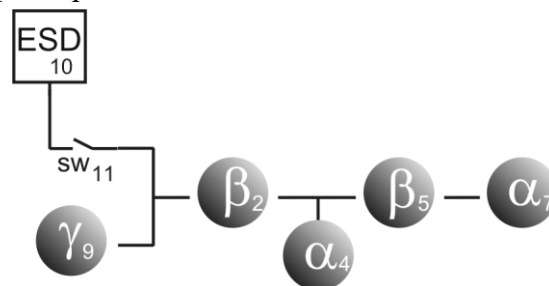


Рисунок 5 – Структурне представлення сентенційної форми l_5

Застосування правила s_9 під час встановлення $l_5 \Rightarrow l_6$ дозволяє отримати $l_6 = (esd_{10} \cdot sw_{11} \times (\theta_{12} : \lambda_{13}) \cdot tr_{14} \cdot epss_{15}) \cdot \beta_2 \cdot (\alpha_4 \times (\beta_5 \cdot \alpha_7))$. Структурна схема на рис. 5 доповнена обладнанням для передачі надлишкової енергії рекуперації у зовнішню мережу живлення від відповідної тягової підстанції (рис. 6).

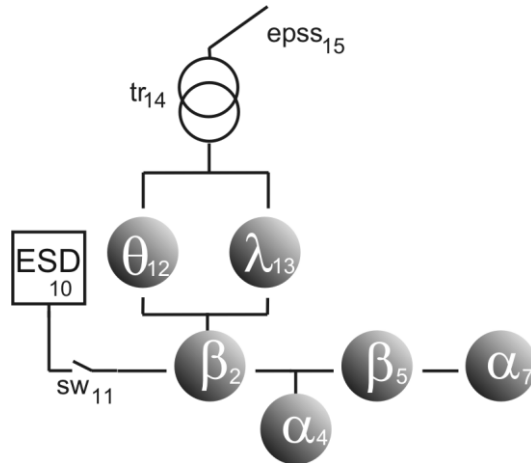


Рисунок 6 – Структурне представлення сентенційної форми l_6

Послідовне застосування правил s_{10} та s_{12} при підстановках $l_6 \Rightarrow l_7 \Rightarrow l_8$ дозволяють отримати $l_7 = (esd_{10} \cdot sw_{11} \times (cr_{16} : \lambda_{13}) \cdot tr_{14} \cdot epss_{15}) \cdot \beta_2 \cdot (\alpha_4 \times (\beta_5 \cdot \alpha_7))$, $l_8 = (esd_{10} \cdot sw_{11} \times (cr_{16} : sw_{17} \cdot aci_{18}) \cdot tr_{14} \cdot epss_{15}) \cdot \beta_2 \cdot (\alpha_4 \times (\beta_5 \cdot \alpha_7))$. За структурною схемою рис. 7 завершена комплектація обладнання відповідної підстанції інвертором та регулятором напруги. Структурну схему цієї тягової підстанції повністю сформовано (рис. 7).

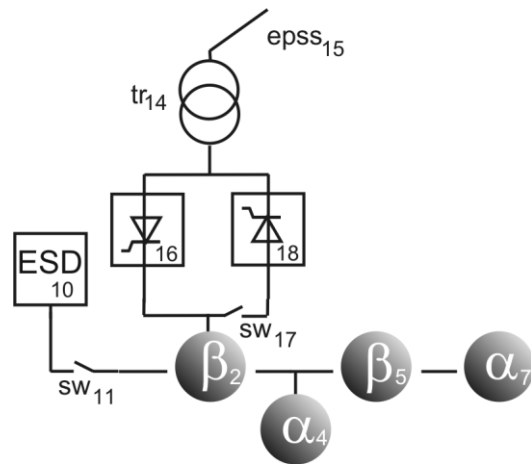


Рисунок 7 – Структурне представлення сентенційної форми l_8

Наступна послідовність підстановок $l_8 \Rightarrow l_9 \Rightarrow \dots \Rightarrow l_{18}$, із подвійним застосуванням правил $s_6, s_7, s_9, s_{10}, s_{13}$, дозволяє отримати $l_{18} = ((esd_{10} \cdot sw_{11} \times (cr_{16} : (sw_{17} \cdot aci_{18}))) \cdot tr_{14} \cdot epss_{15}) \cdot \beta_2 \cdot (((esd_{19} \cdot sw_{20} \times (cr_{25} : (sw_{26} \cdot aci_{27}))) \cdot tr_{23} \cdot epss_{24})) \times (\beta_5 \cdot (((esd_{28} \cdot sw_{29} \times (cr_{34} : (sw_{35} \cdot aci_{36}))) \cdot tr_{32} \cdot epss_{33}))))$. В результаті операцій часткового виведення повністю сформовано структурну схему всіх тягових підстанцій системи електропостачання (рис. 8).

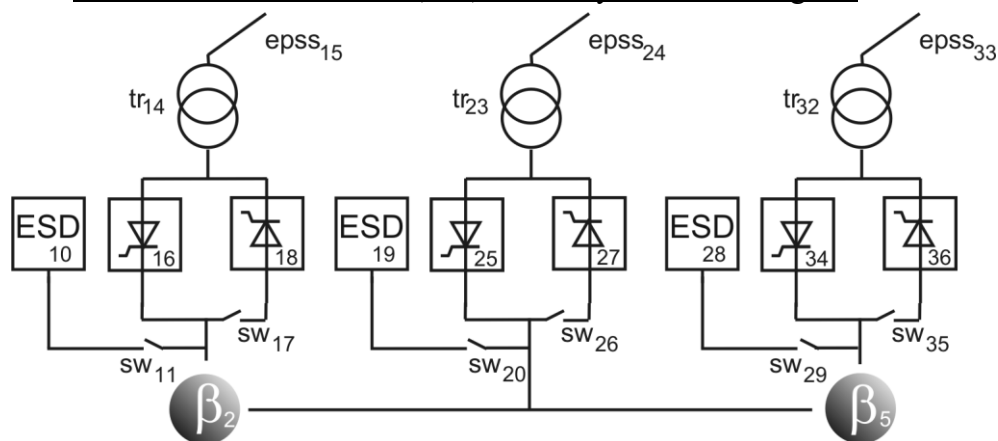


Рисунок 8 – Структурне представлення сентенційної форми l_{18}

Заміна нетерміналів β_2 та β_5 у проміжній формі l_{18} послідовно за правилами $s_{14}, s_{16}, s_{16}, s_{17}, s_{16}, s_{18}$ та $s_{14}, s_{16}, s_{18}, s_{16}, s_{17}, s_{16}, s_{18}$ відповідно, (тобто $\beta_2 \mid \Rightarrow \beta_{37} : \beta_{38} \mid \Rightarrow cl_{39} \cdot \beta_{40} : \beta_{38} \mid \Rightarrow cl_{39} \cdot tr_{41} \cdot \beta_{42} : \beta_{38} \mid \Rightarrow (cl_{39} \cdot tr_{41} \cdot cl_{43} \cdot \beta_{44}) : \beta_{38} \mid \Rightarrow (cl_{39} \cdot tr_{41} \cdot cl_{43}) : \beta_{38} \mid \Rightarrow (cl_{39} \cdot tr_{41} \cdot cl_{43}) : (cl_{45} \cdot \beta_{46}) \mid \Rightarrow (cl_{39} \cdot tr_{41} \cdot cl_{43}) : (cl_{45})$ та $\beta_5 \mid \Rightarrow \beta_{47} : \beta_{48} \mid \Rightarrow cl_{49} \cdot \beta_{50} : \beta_{48} \mid \Rightarrow cl_{49} : \beta_{48} \mid \Rightarrow cl_{49} : (cl_{51} \beta_{52}) \mid \Rightarrow cl_{49} : (cl_{51} \cdot tr_{53} \cdot \beta_{54}) \mid \Rightarrow cl_{49} : (cl_{51} \cdot tr_{53} \cdot cl_{55} \cdot \beta_{56}) \mid \Rightarrow cl_{49} : (cl_{51} \cdot tr_{53} \cdot cl_{55})$) формує варіант двоколісної ділянки з двома потягами як на структурній схемі рис. 9.

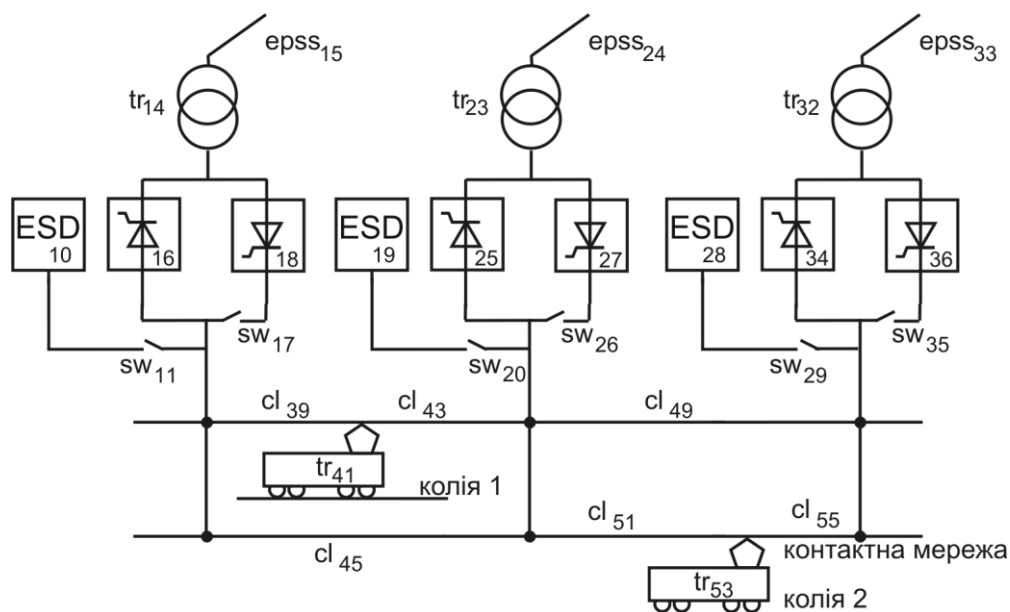


Рисунок 9 – Структурна схема ділянки електропостачання

Сформована структурна схема ділянки споживання енергії рекуперації моделює структуру задіяних тягових підстанцій з максимальною комплектацією та деяку поїзну ситуацію.

Атрибутика елементів схеми формується пристроями вимірювання та передачі інформації та експертними висновками.

2. Обговорення результатів

1. Наведена узагальнена модель дозволяє формувати множину потенційно можливих схем ділянок системи електропостачання, які можуть слугувати основою для подальшого конструктивного моделювання: режимів тягового електропостачання та стану електромережі, варіантів розташування обладнання і поїзної ситуації, визначення на цих моделях експертних висновків щодо використання енергії рекуперації, формування нейрон-фазі або нейронної мережі прийняття рішень, керуючого на їх основі конструктора.

2. Розроблена модель може бути корисною при рішенні низки задач, а саме:

–підвищення енергоефективності системи електричної тяги і, як наслідок, зниження викидів парникових газів у зовнішнє середовище;

–вибору раціональної структури системи електропостачання (розміщення та номенклатури обладнання);

–максимальне використання потенціалу рекуперації енергії та підвищення ресурсу (строку служби) обладнання на основі раціонального керування обладнанням.

3. Вивід схеми може бути як у автоматичному режимі (внутрішнім виконавцем конструктора), так і в керованому (спільно внутрішнім і зовнішнім виконавцями). В останньому випадку можуть бути сформовані схеми електропостачання близькі до реальних з варіативністю обладнання, енергетичної та поїзної ситуації.

3. Висновки

Одним із основних напрямків розвитку економіки України є енергозбереження, що пов'язано зі складною ситуацією на ринку енергоносіїв.

Раціональне використання енергії рекуперації на підставі використання моделей на сучасних принципах дозволяє значно заощаджувати енергоресурси.

Розроблений метод моделювання заснований на можливостях КППМ у новій предметній галузі. Термінальний алфавіт семантично наповнений образами електрообладнання, тягової мережі та споживачами електроенергії, що мають відповідну атрибутику.

Виконано конкретизацію КППМ, яка дозволяє врахувати значну кількість можливостей та особливостей сучасного обладнання систем тягового електропостачання, ділянок тягової мережі та поїзної ситуації. Наведений окремий випадок формування структурної схеми демонструє можливості КППМ стосовно низки задач.

ЛІТЕРАТУРА / REFERENCES

1. Sablin O. Rational distribution of excess regenerative energy in electric transport systems on the basis of fuzzy logic application / O. Sablin, V. Kuznetsov, V. Shinkarenko, A. Ivanov // The archives of transport, 42(2), 2017, 7-17. DOI: 10.5604/01.3001.0010.0527
2. Kampeerawat W. A strategy for utilization of regenerative energy in urban railway system by application of smart train scheduling and wayside energy storage system / W.Kampeerawat, T. Koseki // Energy Procedia, 138, 2017, 795–800. DOI: 10.1016/j.egypro.2017.10.070.

3. Khodaparastan M., Mohamed A.A., Brandauer W. Recuperation of regenerative braking energy in electric rail transit systems. IEEE Trans. Intell. Transp. Syst. 20(8), 2019, 2831–2847. DOI: 10.1109/TITS.2018.2886809
 4. Chen J., Hu H., Ge Y., Wang K., Huang W., He Z. An Energy storage system for recycling regenerative braking energy in high-speed railway. IEEE Trans. Power Deliv. 36(1), 2021, 320–330. DOI: 10.1109/TPWRD.2020.2980018
 5. Chen J., Ge Y., Wang K., Hu H., He Z. Integrated Regenerative Braking Energy Utilization System for Multi-Substations in Electrified Railways, in IEEE Transactions on Industrial Electronics, 70(1), 2023, 298-310, DOI: 10.1109/TIE.2022.3146563
 6. Yu Y., Mi Z., Guo X., Niu X., Zheng X. and Sun C., Control design and implementation of a spiral spring energy storage system connected to a grid via PMSG, in CSEE Journal of Power and Energy Systems, 4(3), 2018, 339-351. DOI: 10.17775/CSEEJPES.2016.00860
 7. Yan N., Li X., Ma S., Zhao H., Zhang B., "Capacity Configuration of Energy Storage Systems for Echelon Utilization Based on Accelerated Life Test in Microgrids", CSEE Journal of Power and Energy Systems, 9(3), 2023, 1139-1150,. DOI: 10.17775/CSEEJPES.2020.01670
 8. Deng W. et al. Back-to-back hybrid energy storage system of electric railway and its control method considering regenerative braking energy recovery and power quality improvement //Proceedings of the CSEE, –. 39(10), 2019, 2914-2924. DOI: 10.1016/j.energy.2022.123263.
 9. Hu H., Liu Y., Li Y. Traction power systems for electrified railways: evolution, state of the art, and future trends. Rail. Eng. Science 32, 2024, 1–19,. DOI: 10.1007/s40534-023-00320-6
 10. Lu J., Zhu C. & Li X. Research on the Recovery and Reuse Method of Train Regenerative Braking Energy Based on the Decommissioned Equipment of EMU Trains. J. Electr. Eng. Technol. 18, 2023, 3941–3949. DOI:10.1007/s42835-023-01433-y
 11. Shynkarenko V.I. Constructive-Synthesizing Structures and Their Grammatical Interpretations.I. Generalized Formal Constructive-Synthesizing Structure / V.I. Shynkarenko , V. M. Ilman // Cybernetics and Systems Analysis, 50(3), 2014, 655-662. DOI: 10.1007/s10559-014-9655-z
- Shynkarenko V.I. Structural Models of Algorithms in Problems of Applied Programming. I. Formal Algorithmic Structures / V.I. Shynkarenko, V.M. Ilman, V.V. Skalozub // Cybernetics and Systems Analysis, 45(3), 2009, 329-339. DOI : 10.1007/s10559-009-9118-0

Received 30.09.2024.

Accepted 05.10.2024.

Constructive-synthesizing modeling of the direct current traction power supply system

A general constructive-synthesizing model of the DC traction power supply section has been developed. The model can be used to solve a number of problems related to reducing electricity consumption in both railway and urban public electric transport. The developed model is focused on determining the availability and nomenclature of traction substation equipment and rational use of recuperation energy. However, it can also be applied to solve other problems related to the designs of the traction power supply system. An example of the formed scheme of a linear power supply section with three substations is given.

The presented generalized model allows to form a set of potentially possible schemes of sections of the power supply system, which can serve as the basis for further constructive modeling: traction power supply modes and the state of the power grid, equipment location options and train situation, determination of expert conclusions on the use of recuperation energy on these models, formation of a neuro-fuzzy or neural network of decision-making and, based on them control constructor.

The developed modeling method is based on the capabilities of the constructive-synthesizing modeling in a new subject area. The terminal alphabet is semantically filled with images of electrical equipment, traction network and electricity consumers with appropriate attributes.

The constructive-synthesizing modeling specification has been performed, which allows taking into account a significant number of capabilities and features of modern equipment of traction power supply systems, traction network sections and train situation. The given individual case of forming a structural diagram demonstrates the capabilities of the constructive-synthesizing modeling in relation to a number of tasks.

Keywords: constructive-synthesizing modeling, direct current traction, formal grammars, constructor, software, information technologies.

Шинкаренко Віктор Іванович – доктор технічних наук, професор Дніпровського інституту інфраструктури і транспорту Українського державного університету науки і технологій.

Гуда Антон Ігорович – доктор технічних наук, професор Дніпровського металургійного інституту Українського державного університету науки і технологій.

Саблін Олег Ігорович – доктор технічних наук, професор Дніпровського інституту інфраструктури і транспорту Українського державного університету науки і технологій.

Іванов Олександр Петрович – кандидат технічних наук, доцент Дніпровського інституту інфраструктури і транспорту Українського державного університету науки і технологій.

Shynkarenko Viktor – doctor of engineering's sciences, professor SEI DIIT Ukrainian State University of Science and Technology.

Guda Anton – doctor of engineering's sciences, professor SEI DMI Ukrainian State University of Science and Technology/

Sablin Oleg – doctor of engineering's sciences, professor SEI DIIT Ukrainian State University of Science and Technology.

Ivanov Oleksandr candidate of technical sciences, associate professor, SEI DIIT Ukrainian State University of Science and Technology.

МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ТЕПЛОВОГО РЕЖИМУ В СОНЯЧНІЙ ТЕРМОФОТОЕЛЕКТРИЧНІЙ ПАНЕЛІ

Анотація. В роботі запропоновано дизайн комбінованої сонячної термофотоелектричної панелі, в якій одночасно генерується електрична та тепла енергія. Тепловий режим в шарах термофотоелектричної панелі досліджувався на основі розробленої нестационарної нелінійної математичної моделі та створеного програмного додатку. В результаті комп'ютерного моделювання доведені перспективи використання термофотоелектричної панелі відповідного дизайну для термічної стабілізації сонячних елементів та генерації низькопотенційної теплоти.

Ключові слова: нестационарна математична модель, система нелінійних диференціальних рівнянь, числові методи, сонячні елементи, розподіл температур, термофотоелектрична панель, низькопотенційна теплота.

Постановка проблеми. Пряме фотоелектричне перетворення енергії Сонця є найбільш розвинутою та адаптованою технологією, яка широко використовується в наземній і космічній енергетиці. Фотоелектричні сонячні панелі забезпечують споживачів електричною енергією, яка відповідає світовим екологічним та економічним стандартам. В той же час фотоелектричні системи мають низькі питомі енергетичні показники, що обмежує їх застосування. По-перше, ці низькі енергетичні показники обумовлені фізичною природою фотоелектричного процесу та сучасним рівнем технологій створення відповідних сонячних елементів [1]. По-друге, для отримання промислових значень електричної потужності, сонячні панелі повинні займати значні земельні ділянки, виділення яких пов'язано із об'єктивними труднощами. Збільшення площі сонячних фотоелектричних станцій призводить до проблем з комутацією між сонячними панелями, що, в свою чергу, зменшує загальну енергетичну потужність станції. Не меншою проблемою, що призводить до падіння енергетичної потужності фотоелектричних станцій, є температурні коливання в місці їх розташування. Зниження ефективності при збільшенні температури характерно для більшості типів сучасних промислових сонячних елементів [2]. Тому важливою науковою проблемою є мінімізація впливу температур на енергетичні показники сонячних панелей.

Аналіз останніх досліджень і публікацій. Існує два основних шляхи розв'язання цієї проблеми – створення нових типів сонячних елементів з малою залежністю ефективності від температури або зміна дизайну сонячних панелей. В світі проводяться інтенсивні наукові дослідження і отримані зразки нових високоефективних сонячних

елементів, в яких ККД мало залежить від температури або навіть зростає при її підвищенні [3]. Але висока вартість таких сонячних елементів не дозволяє використовувати їх в промислових масштабах.

Зміна дизайну сонячних панелей може призводити до підвищенню їх енергетичної ефективності навіть при використанні традиційних сонячних елементів типу моно c-Si (кремнієві монокристалічні) або poly-Si (кремнієві полікристалічні) та ін. Зміна дизайну сонячних панелей проводиться пасивними та активними методами. До пасивних методів відноситься відповідне розташування сонячних панелей з точки зору врахування географічних та кліматичних факторів [4,5], а також особливостей монтажу [6]. До сучасних пасивних методів зміни дизайну сонячних панелей відноситься також введення в їх конструкцію фазоперехідного матеріалу, в якому накопичується зайве тепло, але на даний момент такі системи ще не набули широкого промислового впровадження.

На відміну від комбінованих фазоперехідних сонячних панелей, все більш розповсюдженими стають комбіновані термофотоелектричні (PV/T) панелі [7]. Дизайн таких панелей формується шляхом введення в їх конструкцію додаткових каналів з теплоносієм, який знімає залишкове тепло з сонячних елементів. Це може призводити до термостабілізації або навіть охолодження сонячних елементів, що підвищують їх ефективність. Крім того, додаткова теплота, яка передається конвекцією теплоносію, може використовуватись за призначенням.

Мета дослідження. Метою даної роботи є доведення перспектив удосконалення традиційної poly-Si сонячної панелі шляхом додавання шару теплоносія з фронтальної сторони. Така комбінована PV/T панель досліджувалась методами математичного та комп'ютерного моделювання на основі створеного власного програмного коду. В ході числових досліджень визначено розподіл температур в кожному шарі запропонованої PV/T панелі, вплив на цей розподіл зовнішніх (географічних, кліматичних) та внутрішніх (особливості дизайну та динаміки теплоносія) факторів.

Викладення основного матеріалу дослідження. В якості об'єкта дослідження обиралась модель сонячної панелі типу Photowatt PWX500 з poly-Si сонячними елементами. В цій панелі сонячні елементи розташовані між двома скляними поверхнями з проміжними шарами із герметіку типу EVA та скловолокна. З фронтальної сторони сонячної панелі додатково додається плоский канал з теплоносієм (водою), який може бути розділений перегородками. Вибір саме плоского каналу для теплоносія базується на його енергетичних перевагах, які доведені в [8]. Вся конструкція PV/T панелі розміщується в металевому або полімерному корпусі і з фронтальної сторони закрита захисним склом. Таким чином, запропонована модель PV/T панелі фактично є плоским сонячним колектором особливого типу, в якому генерується як теплова, так і електрична енергія. На рис. 1 схематично показана позамасштабна фізична модель запропонованої PV/T панелі.

Математична модель для визначення розподілу температур в шарах PV/T панелі запропонованого дизайну будувалась на основі нестационарної математичної моделі

для класичної сонячної панелі, яку розроблено в [9]. В математичну модель [9] додатково вводились рівняння, що описують теплообмін в захисному склі та в шарі теплоносія, а також функції апроксимації кліматичних даних в місці функціонування PV/Т панелі. Таким чином, нестационарна математична модель розподілу температур в шарах PV/Т панелі має вигляд:

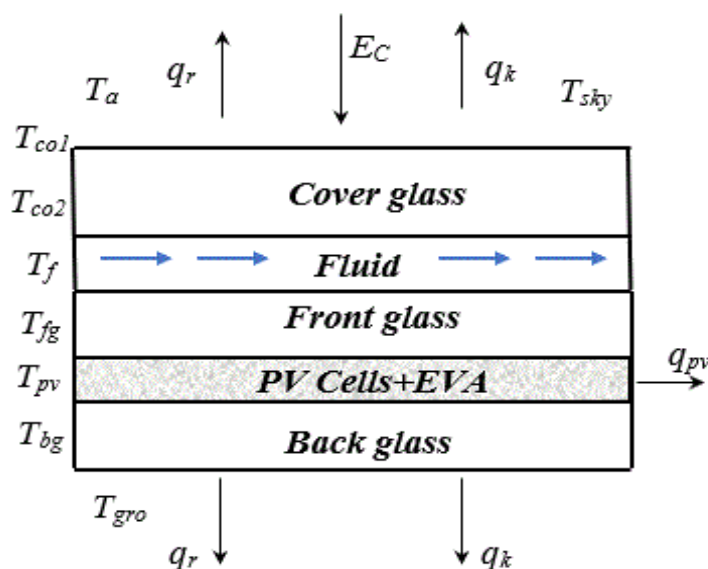


Рисунок 1 – Позамаштабна схема PV/Т панелі

- для фронтальної поверхні захисного скла:

$$\rho_{col} h_{col} C_{p_{col}} \frac{dT_{col}(t)}{dt} = \alpha_a(t)(T_a(t) - T_{col}(t)) + \varepsilon_{col} \sigma F_{colsky} (T_{sky}^4(t) - T_{fg}^4(t)) + \varepsilon_{col} \sigma F_{colgro} (T_{gro}^4(t) - T_{col}^4(t)) + \frac{\lambda_{col}}{h_{col}} (T_{co2}(t) - T_{col}(t)) + A_{col} E_C(t); \quad (1)$$

- для тильної поверхні захисного скла:

$$\rho_{co2} h_{co2} C_{p_{co2}} \frac{dT_{co2}(t)}{dt} = \alpha_f(t)(T_f(t) - T_{co2}(t)) + \varepsilon_{fg}^{co2} \sigma (T_{fg}^4(t) - T_{co2}^4(t)) + \frac{\lambda_{co2}}{h_{co2}} (T_{col}(t) - T_{co2}(t)); \quad (2)$$

- для теплоносія:

$$\rho_f h_f C_{p_f} \frac{dT_f(t)}{dt} = \alpha_f(t)(T_{co2}(t) - T_f(t)) + \alpha_f(t)(T_{fg}(t) - T_f(t)) + \dot{m}_f C_{p_f} (T_f(t) - T_{in}); \quad (3)$$

- для фронтальної поверхні скла сонячної панелі:

$$\rho_{fg} h_{fg} C_{p_{fg}} \frac{dT_{fg}(t)}{dt} = \alpha_f(t)(T_f(t) - T_{fg}(t)) + \varepsilon_{fg}^{co2} \sigma (T_{co2}^4(t) - T_{fg}^4(t)) + \bar{\lambda}_{pv}^{fg} (T_{pv}(t) - T_{fg}(t)); \quad (4)$$

- для шару сонячних елементів:

$$\rho_{pv} h_{pv} C_p \frac{dT_{pv}(t)}{dt} = \bar{\lambda}_{pv}^{fg} (T_{bg}(t) - T_{pv}(t)) + \bar{\lambda}_{bg}^{pv} (T_{fg}(t) - T_{pv}(t)) + q_{pv}(t); \quad (5)$$

- тильної поверхні скла сонячної панелі:

$$\rho_{bg} h_{bg} C_p \frac{dT_{bg}(t)}{dt} = \alpha_a(t) (T_a(t) - T_{bg}(t)) + \varepsilon_{bg} \sigma F_{bgsky} (T_{sky}^4(t) - T_{bg}^4(t)) + \varepsilon_{bg} \sigma F_{bggro} (T_{gro}^4(t) - T_{bg}^4(t)) + \bar{\lambda}_{bg}^{pv} (T_{pv}(t) - T_{bg}(t)). \quad (6)$$

В математичній моделі (1) – (6) $T, h, \rho, C_p, \lambda, \varepsilon$ – температура, товщина, густина, коефіцієнт теплоємності, коефіцієнт теплопровідності та ступень чорноти кожного шару PV/Т панелі, відповідно; T_a, T_{sky}, T_{gro} – температура навколишнього повітря, неба та землі, відповідно [10]; $\bar{\lambda}$ – коефіцієнт ефективної теплопровідності відповідних шарів PV/Т панелі [11]; α_a, α_f – коефіцієнти тепловіддачі між PV/Т панеллю та навколишнім середовищем та між теплоносієм та стінками каналу, відповідно; \dot{m}_f – питома масова витрата теплоносія; A_{col} – коефіцієнт поглинання поверхні захисного скла; σ – стала Стефана – Больцмана; F_i – коефіцієнти, в яких враховується кут розташування панелі відносно горизонту; $E_c(t), q_{pv}(t)$ – щільність теплового потоку від Сонця та від сонячних елементів, відповідно.

Розроблена математична модель (1)–(6) була розв’язана з використанням власного програмного додатку, який реалізовано на мови програмування системи комп’ютерної алгебри Mathematica [12].

Значення щільності теплового потоку від Сонця, температури навколишнього середовища та швидкості вітру, яке необхідно для визначення коефіцієнту тепловіддачі між поверхнями PV/Т панелі, знаходилось за даними відкритих весвітніх кліматичних баз [13]. Значення $q_{pv}(t)$ розраховувалось у відповідності до підходу, представленого в [2].

Моделювання проходило в режимі реального часу в обраній географічній локації для визначеного дня року. Геометричні параметри PV/Т панелі відповідали геометричним параметрам сонячної панелі типу Photowatt PWX500 з товщиною шарів $h_{pv} = 0.22 \cdot 10^{-3} \text{ м}$, $h_{fg} = h_{bg} = 0.0195 \text{ м}$. Товщина шару теплоносія обиралась $h_f = 10 \cdot 10^{-3} \text{ м}$, а товщина захисного скла $h_{col} = h_{co2} = 0.004 \text{ м}$. Теплові втрати з бокових поверхонь панелі не враховувались. Заявлене виробником ККД poly-Si сонячних елементів при температурі $T_{\eta=0.13} = 25^{\circ} \text{ C}$ дорівнювало $\eta = 0.13$. PV/Т панель встановлювалась під кутом $\beta = 45^{\circ}$ до горизонту, початкова температура всіх її шарів складала $T_{ноч} = 10^{\circ} \text{ C}$, а температура теплоносія на вході в канал $T_{in} = 12.5^{\circ} \text{ C}$.

На рис. 2 представлено результати розрахунків по визначенню розподілу температур в шарах PV/Т панелі при питомій масовій витраті теплоносія

$$\dot{m}_f = 1.325 \cdot 10^{-3} \frac{\text{кг}}{\text{с} \cdot \text{м}^2}.$$

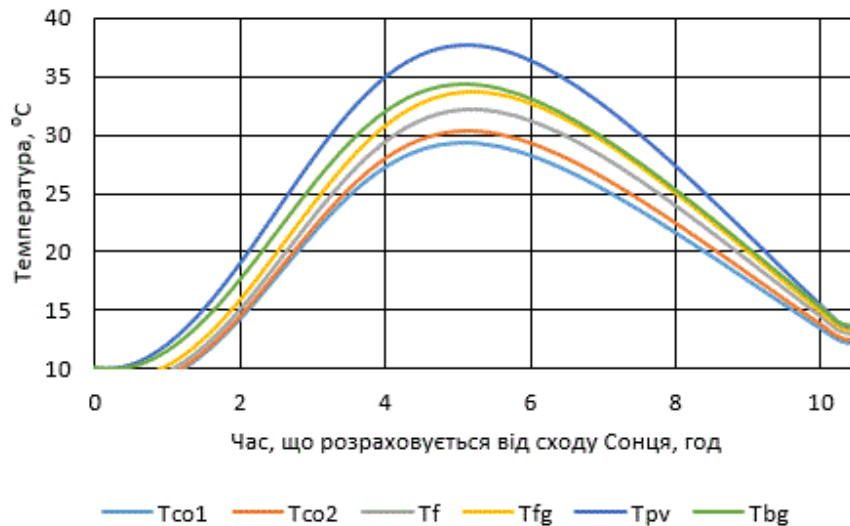


Рисунок 2 – Розподіл температур протягом світового дня в шарах PV/Т панелі

Аналіз графіків показує, що найвищу температуру в PV/Т панелі має шар сонячних елементів, при чому максимальна їх температура перевищує майже на 13°C температуру, яка відповідає максимальній ефективності. Таку ж температуру сонячні елементи мають і в сонячній панелі традиційного дизайну [5], тобто теплоносії, що циркулює в каналі PV/Т панелі, не сприяє охолодженню сонячних елементів. В той же час в PV/Т панелі температура теплоносія в середньому підвищується більш ніж на 20°C і це стає достатнім для генерації низькопотенційної теплоти. Підвищення масової витрати теплоносія призводить до ще більшого зростання температури як сонячних елементів, так і теплоносія. На рис. 3 показані розподіли температур в шарі сонячних елементів та в теплоносії при питомій масовій витраті теплоносія $\dot{m}_f = 1.325 \cdot 10^{-3} \frac{\text{кг}}{\text{с} \cdot \text{м}^2}$ та

$$\dot{m}_f = 2.65 \cdot 10^{-3} \frac{\text{кг}}{\text{с} \cdot \text{м}^2}, \text{ відповідно.}$$

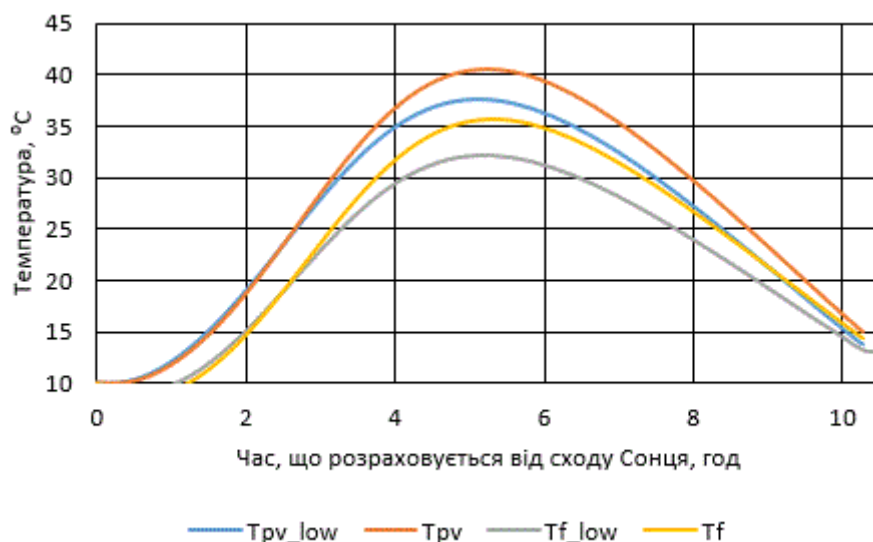


Рисунок 3 – Розподіл температур протягом світового дня в шарі сонячних елементів та в теплоносії PV/Т панелі

Висновки. В роботі запропоновано дизайн сонячної PV/Т панелі для одночасної генерації електричної та теплової енергії. Розроблена математична модель, створений програмний додаток та проведені числові дослідження довели, що обраний дизайн PV/Т панелі забезпечує лише термостабілізацію сонячних елементів, не охолоджуючи їх. В той же час, отриманий рівень середніх температур теплоносія в PV/Т панелі є достатнім для генерацію низькопотенційної теплоти.

ЛІТЕРАТУРА

1. <https://www.nrel.gov/pv/cell-efficiency.html>
2. Захаров Д.В., Книш Л.І. Математичне моделювання впливу температурного режиму на ефективність сонячної панелі, Проблеми обчислювальної механіки і міцності конструкцій, 2022, Т.34, С. 48-58.
3. Fazal M.A., Rubaiee Saeed, Progress of PV cell technology: Feasibility of building materials, cost, performance, and stability, Solar Energy, 2023, Vol.258, P.203 – 219.
4. Rakhimov E.Yu., Avezova N.R., Emamgholizadeh Samad, Ziaii Mansour, Assessment of the Technical Potential of PV Stations on the Example of the Fergana Valley. Part II: Analysis of Sunny, Partly Cloudy and Cloudy Days, Applied solar energy, 2024, Vol. 60, P. 346 – 356.
5. Книш Л.І., Захаров Д.В. Нестационарна математична модель розподілу температур в шарах сонячної панелі при реальних умовах її функціонування, Матеріали XXV Міжнародної науково-практичної конференції “Відновлювана енергетика та енергоефективність у XXI столітті”, м. Київ, 22–24 травня 2024р., С.202.
6. Delia D'Agostino, Danny Parker, Paco Melià, Giovanni Dotelli, Optimizing photovoltaic electric generation and roof insulation in existing residential buildings, Energy and Building, 2022, Vol.255, 111652.

7. Adnan Ibrahim, Mohd Yusof Othman, Mohd Hafidz Ruslan, Sohif Mat, Kamaruzzaman Sopian, Recent advances in flat plate photovoltaic/thermal (PV/T) solar collectors, *Renewable and Sustainable Energy Reviews*, 2011, Vol.15, Is.1, P.352 – 365.
8. Cristofari C., Notton G., Canaletti J.L. Thermal behavior of a copolymer PV/Th solar system in low flow rate conditions, *Solar Energy*, 2009, Vol. 83, P. 1123–1138.
9. Захаров Д. В., Книш Л. І. Нестационарна математична модель розподілу температур в шарах сонячної панелі. *Технічна механіка*, 2023 (3), С.79 – 87.
10. Kudish A.I., Evseev E.G., Walter G., Leukefeld T., Simulation study of a solar collector with a selectively coated polymeric double walled absorber plate, *Energy Conversion and Management*, 2002, Vol.43, P. 651 – 671.
11. Cristofari C., Notton G., Poggi P., Louche A., Modelling and performance of a copolymer solar water heating collector, *Solar Energy*, 2002, Vol. 72 (2), P. 99–112.
12. <https://www.wolframcloud.com>
13. <https://neo.gsfc.nasa.gov>

REFERENCES

1. <https://www.nrel.gov/pv/cell-efficiency.html>
2. Zakharov D.V., Knysh L.I., Mathematical modeling of the influence of temperature on the efficiency of the solar panel, *Problems of computational mechanics and strength of structures*, 2022, Vol.34, P.48-58.
3. Fazal M.A., Rubaiee Saeed, Progress of PV cell technology: Feasibility of building materials, cost, performance, and stability, *Solar Energy*, 2023, Vol.258, P.203 – 219.
4. Rakhimov E.Yu., Avezova N.R., Emamgholizadeh Samad, Ziaii Mansour, Assessment of the Technical Potential of PV Stations on the Example of the Fergana Valley. Part II: Analysis of Sunny, Partly Cloudy and Cloudy Days, *Applied solar energy*, 2024, Vol. 60, P. 346 – 356.
5. Knysh L., Zakharov D., Non-stationary mathematical model of temperature distributions in solar panel layers during real operating conditions, *Materials of the XXV International scientific and practical conference «Renewable energy and energy efficiency in the XXI century»*, Kyiv, May 22-24, 2024, P.202.
6. Delia D'Agostino, Danny Parker, Paco Melià, Giovanni Dotelli, Optimizing photovoltaic electric generation and roof insulation in existing residential buildings, *Energy and Building*, 2022, Vol.255, 111652.
7. Adnan Ibrahim, Mohd Yusof Othman, Mohd Hafidz Ruslan, Sohif Mat, Kamaruzzaman Sopian, Recent advances in flat plate photovoltaic/thermal (PV/T) solar collectors, *Renewable and Sustainable Energy Reviews*, 2011, Vol.15, Is.1, P.352 – 365.
8. Cristofari C., Notton G., Canaletti J.L. Thermal behavior of a copolymer PV/Th solar system in low flow rate conditions, *Solar Energy*, 2009, Vol. 83, P. 1123–1138.
9. Zakharov D.V., Knysh L.I., Non-stationary mathematical model of the temperature distribution in solar panel layers, *Technical mechanics*, 2023, №3, P. 79 -87.
10. Kudish A.I., Evseev E.G., Walter G., Leukefeld T., Simulation study of a solar collector with a selectively coated polymeric double walled absorber plate, *Energy Conversion and Management*, 2002, Vol.43, P. 651 – 671.

11. Cristofari C., Notton G., Poggi P., Louche A., Modelling and performance of a copolymer solar water heating collector, *Solar Energy*, 2002, Vol. 72(2), P. 99–112.

12. <https://www.wolframcloud.com>

13. <https://neo.gsfc.nasa.gov>

Received 02.10.2024.

Accepted 15.10.2024.

Mathematical modeling of the thermal regime in solar photovoltaic thermal panel

Design of a combined solar photovoltaic thermal panel (PV/T) for the simultaneous generation of electrical and thermal energy was proposed in this study. The basis of the new design is a traditional solar panel with poly-Si solar cells. A flat channel with a heat transfer fluid is added to the front size of such a panel. This channel is bounded by cover glass. A non-stationary mathematical model was developed for determination of temperature regime in the PV/T panel. This model consists to the system of nonlinear ordinary differential equations, which describes mutual influence of external and internal heat flows and temperatures. A Math-software was created based on the developed mathematical model. The numerical studies were conducted in in real-time mode for selected geographical location of the PV/T panel. Heat flux density from the Sun, wind speed and ambient temperature were determined based on data from open worldwide climate databases. As result of computer modeling, the typical temperature distributions in each layer of the PV/T panel during daylight hours were founded. It was determined that the heat transfer fluid moving in a transparent channel from the front side of the solar panel does not cool the solar cell. This heat transfer fluid ensures only their thermal stability at the corresponding value of the specific mass flow rate. With an increase of the specific mass flow rate of the heat transfer fluid, the growth of solar cells temperature is observed under unchanged environmental conditions. An the same time, the proposed design of the PV/T panel ensures a significant increase of the heat transfer fluid temperature. This makes it possible to use it in low-potential heat generation systems. This leads to an increase in the economic efficiency of solar panels, economy of occupied areas, optimization of system of production, consumption and storge of thermal and electrical energy.

Книш Людмила Іванівна – д.т.н., професор, професор кафедри комп'ютерних технологій Дніпровського національного університету імені Олеся Гончара (м. Дніпро).

Захаров Денис Віталійович – аспірант кафедри комп'ютерних технологій Дніпровського національного університету імені Олеся Гончара (м. Дніпро).

Knysh Liudmyla – doctor of technical science, professor, professor of department of computer technologies of Oles Honchar Dnipro National University (Dnipro).

Zakharov Denis – PhD student of department of computer technologies of Oles Honchar Dnipro National University (Dnipro).

ANALYSIS OF THE IMPACT OF TASK PRIORITIZATION LISTS ON THE POTENTIAL FOR AVOIDING ANOMALIES IN TASK SCHEDULING

Annotation. This paper addresses relevant issues related to the anomalous deterioration in objective function values when attempting to improve the initial parameters in one of the discrete optimization problems. The primary focus is on investigating the conditions under which it is possible to prevent the occurrence of such anomalies. Contemporary scientific works devoted to schedule optimization and task prioritization management, particularly for location-allocation problems arising in the fields of computer science, engineering, and operations research, are reviewed. A priority dynamic redistribution algorithm is proposed, which allows minimizing delays and ensuring efficient resource utilization during parallel task execution. An example of applying the algorithm is provided, and its effectiveness in preventing anomalies.

Keywords: schedule theory, optimal orderings, location-allocation problems, discrete optimization, anomalies, priority dynamic redistribution, scheduling algorithms, interruptions, mathematical modeling.

In the modern world of technology, which is continuously evolving and integrating into all areas of activity, effective task management is a key aspect of achieving high levels of productivity. However, alongside the development of systems, numerous challenges arise in their improvement and optimization, which require innovative mathematical approaches and solutions. One of the most complex problems in this area is the anomalous deterioration of objective function values when attempting logical improvements to the initial system parameters.

Problem statement. The study of anomalous cases in discrete optimization problems, particularly in constructing parallel optimal orderings, is often associated with determining the conditions under which it is possible to prevent the occurrence of these anomalies [1]. Considering the generalization of the problem of minimizing the total task completion time, an interesting question arises regarding the impact of interruptions on the value of the objective function [2].

Analysis of the latest research and publications. In recent decades, a large number of scientific works have been devoted to schedule optimization and task prioritization management in various fields such as computer science, engineering, and operations research. These works emphasize the importance of ensuring efficient resource allocation and minimizing delays in task execution. In particular, significant attention is paid to algorithms that allow dy-

dynamic resource redistribution and account for changes in the execution environment in real time.

One of the key studies is the work of Avinab Marahatta and co-authors [3], which proposes a classification scheme for dynamic task scheduling in virtualized cloud data centers, improving energy efficiency and task execution by dynamically redistributing resources and adapting to changes in system load.

In [4], a review of 67 different task scheduling algorithms is conducted, particularly those optimizing energy consumption in cloud environments. The advantages and disadvantages of various approaches, including dynamic task redistribution algorithms that consider priorities and variable execution conditions, are discussed.

The purpose of the study. The purpose of this research is to develop and justify the effectiveness of a priority dynamic redistribution algorithm that prevents the occurrence of anomalies in parallel task execution. By allowing interruptions and dynamic redistribution, the algorithm aims to optimize task execution order, minimize delays, and ensure efficient resource utilization. The task is to establish the conditions under which the proposed algorithm guarantees the execution of more important tasks first, allows more prioritized tasks to be placed in the leftmost permissible positions, and considers changes in the availability of executors and the state of task execution in real time.

Presentation of the main material of the research. When studying anomalous cases, questions often arise about determining the conditions under which it is possible to prevent the occurrence of these anomalies. In particular, the study of the impact of interruptions on the increase in the objective function value becomes interesting.

Let a directed graph $G = (V, U)$ be given, along with a priority list $L = (i_1, \dots, i_n)$ and vertex weights $T = (\tau_1, \dots, \tau_n)$. We introduce additional initial conditions: constraints on the ordering width with the set of values h_i and the allowance for interruptions during work execution [2]. The constructed parallel ordering should satisfy the priority list and have a minimal length.

Statement. Allowing interruptions for some tasks in cases where anomalies exist helps avoid their occurrence under the following conditions:

Priority: The priority list must be built according to the optimal priority list construction algorithm; high-priority tasks can interrupt the execution of lower-priority tasks.

Interruption: An interrupted task can be resumed at any step, but according to the priority list.

Dynamic Actualization: At each step of task execution, work can be interrupted and tasks can be redistributed.

Proof.

Priority. Let a priority list be given, built according to the optimal priority list construction algorithm. We will prove that this condition guarantees the execution of more important tasks first. Suppose that a high-priority task (for example, task i) can interrupt a lower-priority task (for example, task j). According to this condition, task i is always executed before task j

if they are simultaneously available for execution. This means that a high-priority task will never be blocked by a lower-priority task, which prevents delays in the execution of critically important tasks.

Interruption. Consider the condition that an interrupted task can be resumed at any step, but according to the priority list. We will prove that this allows more prioritized tasks to be placed in the leftmost permissible positions. Let task i be interrupted by task j with a higher priority. Since task i can be resumed later, but only after all higher-priority tasks have been completed, this ensures that task i does not occupy the executor at a moment when it is possible to execute task j . Thus, interruptions allow more prioritized tasks to be placed in the leftmost permissible positions in the schedule, reducing the overall execution time and preventing anomalies.

Dynamic Actualization. Consider the condition that at each step of task execution, work can be interrupted and tasks can be redistributed. We will prove that this condition accounts for changes in the availability of executors and the state of task execution in real time. Suppose task i is being executed, but at the current step, a higher priority task j appears. Dynamic actualization allows the immediate suspension of task i and the assignment of resources to execute task j . Thus, interruptions and task redistributions account for changes in the availability of executors and the state of task execution in real time, allowing for a prompt response to changing conditions and avoiding delays. This also helps prevent anomalies by ensuring efficient resource use and task execution in order of importance.

Condition 1 guarantees the execution of more important tasks first.

Condition 2 allows more prioritized tasks to be executed in the leftmost permissible positions [5].

Condition 3 accounts for changes in the availability of executors and the state of task execution in real time. We will illustrate the fulfillment of these conditions with the following example.

Example. Graph G is given (fig. 1),

$$L = (1,2,3,4,5,6,7,8,9,10), T = (4,4,2,2,1,3,4,2,1,7), h = 3.$$

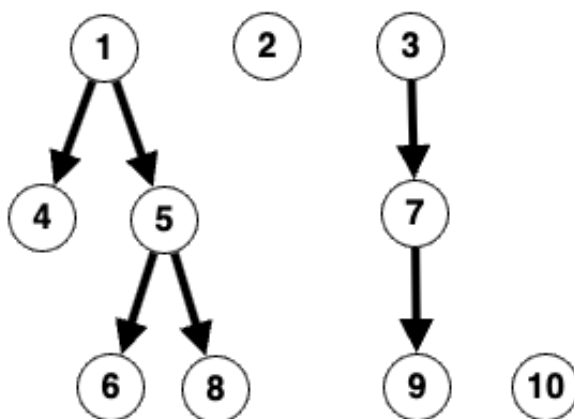


Figure 1 - Graph G

Table 1

Optimal ordering

1	1	1	1	4	4	8	8						
2	2	2	2	5	6	6	6						
3	3	7	7	7	7	9	10	10	10	10	10	10	10

The length is $l = 14$.

Table 2

Optimal ordering with interruptions

1	1	1	1	4	4	7	9						
2	2	2	2	5	7	8	8						
3	3	7	7	6	6	6	10	10	10	10	10	10	10

The length of the ordering did not change, $l = 14$.

Apply the dynamic resource redistribution algorithm to solve this problem. Modify L , using the optimal priority list construction algorithm.

Find the total task completion time for each of the graph paths:

$$l_{1-4} = 5, l_{1-6} = 8, l_{1-8} = 7, l_2 = 4, l_{3-9} = 7, l_{10} = 7.$$

Update the priority list: $L' = (1, 5, 6, 8, 3, 7, 9, 10, 4, 2)$.

Table 3

Optimal ordering using list L' and the possibility of interruptions

1	1	1	1	5	6	6	6	10	2	2	2		
3	3	7	7	7	8	8	10	4					
10	10	10	10	10	7	9	4	2					

The length of the ordering is $l = 12$, which is 14.3% shorter than the initial result.

Let's take a closer look at the fulfillment of conditions 1-3 on the constructed ordering from table 3.

Condition 1: task i_{10} is interrupted by task i_7 , and task i_7 is interrupted by task i_8 .

Condition 2. Find \underline{S} (table 4), \bar{S} (table 5) and the sets of allowable places for vertices i_5 and i_7 .

Ordering \underline{S}

1	1	1	1	4	4		
2	2	2	2	5	6	6	6
3	3	7	7	7	7	9	
10	10	10	10	10	10	10	
					8	8	

Table 5

Ordering \bar{S}

	10	10	10	10	10	10	10
	3	3	7	7	7	7	9
						8	8
1	1	1	1	5	6	6	6
						4	4
				2	2	2	2

$\mu_5 = [5], \mu_7 = [3,7]$. Task i_5 is performed in the only allowable place. Task i_7 starts execution at the leftmost allowable place due to the impossibility of executing higher priority tasks at time t_3 , and after the interruption, it starts execution at the leftmost allowable place according to the priority list.

Condition 3: task i_7 is interrupted by task i_8 , while on the next step, task i_7 interrupts task i_{10} .

The pseudocode of the algorithm for solving this problem in Python:

```

#Priority Dynamic Redistribution Algorithm
#Initialization: Mark all executors as free
executors = [free for _ in range(total_executors)]
tasks = priority_queue #Assume this is a list of tasks sorted
by priority

while tasks:
    #Assign work: Assign available task i to a free executor
    for task in tasks:
        if task.available:
            for executor in executors:
                if executor == free:
                    executor = task
                    task.assigned = True
                    break

    #Redistribution: Check for higher priority tasks
    for executor in executors:

```

```
if executor != free:
    current_task = executor
    for task in tasks:
        if task.priority < current_task.priority and
task.available:
            #Interrupt current task and assign higher
priority task

            executor = task
            task.assigned = True
            current_task.interrupted = True
            break

    #Update the list of tasks: remove completed tasks, update
availability
    tasks = [task for task in tasks if not task.completed]
    #Mark all free executors
    for i in range(len(executors)):
        if executors[i].completed:
            executors[i] = free
#End of Algorithm
```

Let's consider the operation of the algorithm step by step.

Initialization: All executors are initially marked as free.

Job Assignment: Iterate through the list of tasks (assumed to be sorted by priority) and assign available tasks to free executors.

Redistribution: For each executor who is currently working on a task, check if a higher priority task is available. If a higher priority task is found, interrupt the current task and assign the higher priority task to the executor.

Task Update: Remove completed tasks from the list and update the availability of tasks and executors.

Free Executors: Mark executors as free if they have completed their assigned tasks.

This pseudocode provides a clear and structured way to manage tasks based on their priority and dynamically reassign executors to higher priority tasks as they become available.

Conclusions. The results obtained during the research demonstrate that using the priority dynamic redistribution of tasks significantly reduces the risk of anomalies in parallel ordering problems. The proposed algorithm ensures the execution of the most important tasks first and allows them to be placed in the leftmost allowable positions. This ensures optimal resource utilization and minimizing total execution time. Dynamic task distribution updating allows for real-time adjustments to executor availability and task execution possibilities, enabling timely responses to changes in conditions to avoid delays. Thus, applying the proposed approach provides an effective solution for parallel ordering problems, particularly in real-time conditions.

Moreover, it is worth noting several important questions that require further research.

Clarification of algorithm efficiency criteria. It is necessary to explore in more detail which efficiency criteria should be considered when evaluating the performance of the priority dynamic redistribution in real-world systems. For example, it is essential to analyze how the algorithm handles different types of anomalies and which parameters have the most significant impact on its effectiveness.

Adaptation of the algorithm to other problem types. It is worth considering the possibility of adapting the proposed algorithm to other types of discrete optimization problems where anomalies arise. Specifically, this includes location-allocation problems with complex constraints, which occur in various applied fields such as logistics, medicine and finance.

Investigation of the impact of unforeseen events. It is important to study how the proposed algorithm responds to sudden changes in task execution conditions, such as the emergence of new tasks or executor failures. Such studies could enhance the algorithm and make it more adaptable to unforeseen events.

Integration with other optimization algorithms. Such as genetic algorithms, machine learning, or deep analysis methods. This would enable the creation of hybrid algorithms that could more effectively handle complex tasks.

Analysis of implementation costs. It is crucial to explore the implementation costs of the proposed approach in real-world systems. This includes computational resources needed to run the algorithm and the time required to adapt the algorithm to specific operating conditions.

These further research directions will contribute both to new theoretical results for solving some classical discrete optimization problems and to the potential for effective application to important practical tasks.

REFERENCES

1. Graham R. (1969) Bounds on multiprocessing timing anomalies. *SIAM Journal on Applied Mathematics*. Vol. 17. PP. 416–429. DOI: 10.1137/0117039
2. Малієнко О.О., Коваленко Є.О. (2024) Дослідження впливу переривань на виникнення аномалій у задачах паралельного упорядкування. *Комбінаторні конфігурації та їхні застосування*. Вип. 26. С. 103-107.
3. Marahatta A., Pirbhulal S., Zhang F., Parizi R. M., Choo K.-K. R., Liu Z. (2021) Classification-Based and Energy-Efficient Dynamic Task Scheduling Scheme for Virtualized Cloud Data Center. *IEEE Transactions on Cloud Computing*. Vol. 9. PP. 1376-1390 DOI: 10.1109/TCC.2019.2918226
4. Ghafari, R., Kabutarkhani, F. H., Mansouri, N. (2022) Task scheduling algorithms for energy optimization in cloud environment: a comprehensive review. *Cluster Comput.* Vol. 25. PP. 1035-1093 DOI: 10.1007/s10586-021-03512-z
5. Турчина В.А., Федоренко Н.К. (2011) Алгоритми побудов усіх паралельних упорядкувань заданої довжини. *Питання прикладної математики і математичного моделювання*. С. 268-274.

Received 19.10.2024.
Accepted 25.10.2024.

***Аналіз впливу списку пріоритетів виконання завдань
на можливість уникнення аномалій у задачах упорядкування***

У даній роботі розглянуто актуальні проблеми, пов'язані з аномальним погіршенням значень цільової функції при спробах покращення початкових параметрів в одній із задач дискретної оптимізації. Основна увага приділена дослідженню умов, за яких можливе запобігання виникненню таких аномалій. Розглядаються сучасні наукові роботи, присвячені оптимізації розкладів і управлінню пріоритетами завдань, зокрема для задач розподілу та розміщення, що виникають в галузях комп'ютерних наук, інженерії та операційних досліджень. Запропоновано алгоритм пріоритетного динамічного перерозподілу, що дозволяє мінімізувати затримки та забезпечити ефективно використання ресурсів при паралельному виконанні завдань. Наведено приклад застосування алгоритму та доведено його ефективність у запобіганні виникненню аномалій.

Малієнко Ольга Олександрівна – аспірант Дніпровського національного університету імені Олеся Гончара.

Турчина Валентина Андріївна – кандидат фізико-математичних наук, доцент, завідувач кафедри обчислювальної математики та математичної кібернетики Дніпровського національного університету імені Олеся Гончара.

Olha Maliienko – Postgraduate Student of Oles Honchar Dnipro National University.

Valentyna Turchyna – Candidate of Physical and Mathematical Sciences, Associate Professor, Head of the Department of Computational Mathematics and Mathematical Cybernetics of Oles Honchar Dnipro National University.

ПРОГНОЗУВАННЯ ЕФЕКТИВНОСТІ AGILE-ПРОЕКТІВ ЗА ДОПОМОГОЮ МАШИННОГО НАВЧАННЯ

Анотація. Дослідження спрямоване на вирішення актуальних завдань у розвитку інформаційних систем і інтеграцію штучного інтелекту для оптимізації методологій Agile. Мета полягає у моделюванні та прогнозуванні ефективності Agile-процесів, враховуючи змінні вимоги та випадкові відхилення. Прогнозування ефективності розглядається як процес, де параметри адаптуються залежно від системи та даних. Розглядається вибір оптимальних моделей для прогнозування, порівнюючи SVM, Random Forest та нейронні мережі (CNN, LSTM). Нейронні мережі виявилися більш ефективними для прогнозування успішності проектів. Практична цінність дослідження полягає в ідентифікації LSTM та інших архітектур як ефективних інструментів. Інтеграція моделей машинного навчання, зокрема LSTM та CNN, покращує точність прогнозування та управління Agile-проектами, підвищуючи ефективність проектного управління.

Ключові слова: Agile, прогнозування, нейронні мережі, штучний інтелект, інформаційні системи, машинне навчання, LSTM.

Постановка проблеми. У сучасних умовах розвитку інформаційних технологій та швидких змін вимог до програмного забезпечення методології Agile стають все популярнішими. За даними Zenhub 2024 року [1], впровадження Agile-методологій скорочує час розробки продуктів і підвищує їх якість, покращуючи конкурентоспроможність організацій. Інтеграція машинного навчання додатково підвищує ефективність та точність прогнозів, дозволяючи швидко реагувати на зміни та ефективно управляти проектами. Інтеграція штучного інтелекту (ШІ) відкриває нові можливості для підвищення ефективності Agile-процесів.

Актуальність дослідження обумовлена необхідністю оптимізації процесів розробки програмного забезпечення в умовах швидко змінюваних вимог і обмежених ресурсів. Використання технологій штучного інтелекту для підтримки Agile-методологій дозволяє підвищити точність прогнозування, автоматизувати планування та оцінку ризиків, що, в свою чергу, сприяє більш ефективному управлінню проектами. Це підтверджується дослідженням [2], яке показує, що моделі машинного навчання, такі як LSTM, CNN та Random Forest, демонструють значні переваги для прогнозування успішності Agile-проектів, забезпечуючи високу точність та адаптивність. Поточне дослідження

зосереджене на вивченні впливу ШІ на Agile-процеси, а також на розробці нових підходів до інтеграції інноваційних технологій у методології управління проектами.

Інтеграція інструментів штучного інтелекту в Agile-процеси є новим напрямком, який вимагає детального дослідження та аналізу. Використання технологій обчислювального інтелекту, таких як нейронні мережі та генеративні змагальні мережі (GAN), стало ефективним підходом для прогнозування часових рядів та управління проектами [3], [4]. Використання машинного навчання для прогнозування дозволяє значно покращити управління проектами, зокрема точність прогнозування та ефективність управління ризиками [5].

Дослідження [6] застосування сучасних методів машинного навчання, таких як AutoML та глибоке навчання, для прогнозування ефективності проектів, показує значне покращення результатів порівняно з традиційними методами. Це дозволяє не тільки підвищити продуктивність команд розробників, але й значно зменшити ризики та покращити якість кінцевого продукту. У цьому контексті, актуальним є дослідження використання різних моделей машинного навчання, таких як Support Vector Machines (SVM), Random Forest та нейронних мереж (NN), зокрема Long Short-Term Memory (LSTM), для прогнозування ефективності проектів і оптимізації Agile-процесів.

Аналіз останніх досліджень і публікацій. У процесі проведення даного дослідження було проаналізовано ряд останніх публікацій, що стосуються використання штучного інтелекту та інструментів моделювання в рамках методологій Agile. Інтеграція штучного інтелекту в процеси Agile розробки значно підвищує ефективність, швидкість та якість розробки програмного забезпечення, автоматизуючи рутинні завдання, такі як генерація коду, тестування та виправлення помилок. Це дозволяє командам зосередитися на інноваціях та вирішенні складних проблем, підвищуючи загальну продуктивність та якість продукту [7].

Наприклад, у роботі [8] досліджували використання генеративного штучного інтелекту для автоматизації планування та оцінки ризиків у Agile-проектах, підкреслюючи позитивний вплив штучного інтелекту на їх ефективність та адаптивність. Автори провели порівняльний аналіз різних моделей машинного навчання, зокрема генеративних змагальних мереж (GAN), для прогнозування ефективності проектів, виявляючи їх значний вплив на результати.

Інші дослідження, такі як [9], показали, що використання методів машинного навчання, таких як генеративні змагальні мережі (GAN) та автоматизовані інструменти машинного навчання (AutoML), суттєво покращує точність прогнозів та ефективність управління Agile-проектами. Дослідження, таке як [10], аналізувало впровадження технологій штучного інтелекту в методології Agile, висвітлюючи переваги у підвищенні продуктивності команд та якості кінцевого продукту, а також виклики, пов'язані з безпекою даних та потребою в спеціалізованих технічних знаннях. Загальний аналіз свідчить, що інтеграція штучного інтелекту в методології Agile є перспективним напрямком, який значно підвищує адаптивність та ефективність проектів в інформаційних системах.

Метою цього дослідження є створення та інтеграція передових моделей машинного навчання для підвищення ефективності методологій Agile. Дослідження спрямоване на розробку інноваційних підходів, що використовують технології штучного інтелекту для оптимізації управління проектами, зокрема через поліпшення прогнозування ефективності Agile-процесів. Особлива увага приділяється адаптації моделей до змінних умов та вимог, що забезпечує гнучкість і оперативність інформаційних систем у динамічному середовищі.

Викладення основного матеріалу дослідження. Для моделювання та прогнозування ефективності Agile-процесів у дослідженні були використані різноманітні інструменти штучного інтелекту, що дозволяють глибше аналізувати великі обсяги даних та виявляти складні залежності. Усі графіки та візуалізації створені за допомогою мови програмування Python та спеціалізованих бібліотек, таких як Matplotlib, Seaborn та TensorFlow, що забезпечують високу точність та наочність представлених даних.

Основні етапи дослідження включали ретельний збір даних, що охоплює метрики ефективності та історичні дані проєктів, вибір відповідних моделей машинного навчання, їх тренування на зібраних даних, тестування моделей для оцінки їхньої точності та надійності, а також детальний аналіз отриманих результатів для виявлення оптимальних рішень у контексті управління Agile-процесами.

1. Для проведення дослідження використовувалися такі дані:

Метрики ефективності, обсяги робіт, часові рамки та ресурси проєктів Agile, історичні дані про успішність виконання Agile-проєктів, включаючи часові ряди завдань, етапи виконання та результати, дані про вимоги та їхні зміни під час реалізації проєктів для врахування їхнього впливу на кінцевий результат.

2. Для проведення даного дослідження було обрано наступні моделі:

Лінійна регресія, Random Forest, Support Vector Machines (SVM), нейронні мережі, включаючи Convolutional Neural Networks (CNN) та Long Short-Term Memory (LSTM). Вибір цих моделей обумовлений їхньою здатністю до аналізу великих обсягів даних та обробки складних залежностей між змінними. Параметри моделей налаштовувалися на основі гіперпараметрів, таких як кількість шарів та нейронів у випадку нейронних мереж.

3. Тренування та тестування моделей:

– Моделі тренувалися на історичних даних проєктів. Машинне навчання, особливо LSTM, демонструє значні переваги для прогнозування в умовах Agile. Практична реалізація Agile-моделей потребує відповідних інструментів управління часом проєкту, включаючи автоматизовані інструменти та платформи для ефективного планування, відстеження та коригування часових рамок [11].

– Для тестування використовувалися різні набори даних для оцінки точності та надійності моделей. Використання методів, таких як Random Forest та Support Vector Machines, значно покращує точність прогнозів та ефективність управління проєктами Agile.

– Формули для обчислення метрик:

1. MAE (Mean Absolute Error):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \tilde{y}_i| \quad (1)$$

2. RMSE (Root Mean Squared Error):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2} \quad (2)$$

3. R² Coefficient of Determination)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

4. MAPE (Mean Absolute Percentage Error):

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \tilde{y}_i}{y_i} \right| \quad (4)$$

5. EVS (Explained Variance Score):

$$EVS = 1 - \frac{Var(y - \tilde{y})}{Var(y)} \quad (5)$$

де y_i – фактичне значення, \tilde{y}_i – прогнозоване значення, \bar{y} – середнє значення, n – кількість спостережень.

4. Оцінка та аналіз результатів:

– Проведено порівняння моделей за метриками точності: середня абсолютна похибка (MAE), середньоквадратична похибка (RMSE) та коефіцієнт детермінації (R²). Дослідження [12] підтверджує, що інтеграція технологій штучного інтелекту в методології Agile підвищує продуктивність команд та покращує якість кінцевого продукту, що відображається в метриках точності моделей.

– Для глибшого розгляду ефективності моделей включено аналіз додаткових метрик, таких як MAPE та EVS. Це дозволяє отримати комплексне уявлення про ефективність моделей. У результаті проведено порівняльний аналіз прогнозовної якості використаних моделей (табл. 1).

Таблиця 1

Порівняння точності моделей машинного навчання за основними метриками (MAE, RMSE, R², MAPE, EVS)

Модель	MAE	RMSE	R ²	MAPE	EVS
Лінійна регресія	1.513	1.945	0.802	10.5	0.8
Random Forest	1.265	1.688	0.862	0.2	0.86
Support Vector Machines	1.425	1.817	0.832	9.3	0.83
CNN	1.200	1.580	0.880	7.8	0.88
LSTM	1.150	1.500	0.890	7.5	0.89
Bidirectional LSTM	1.100	1.450	0.900	7.2	0.90

Графік на рисунку 1 демонструє порівняльний аналіз точності прогнозів різних моделей машинного навчання. Лінійна регресія демонструє середній рівень точності прогнозування, але її простота та швидкість роблять її корисною для базових прогнозів.

Random Forest продемонстрував високу точність та надійність у прогнозуванні ефективності проектів завдяки своїй здатності враховувати нелінійні взаємозв'язки між змінними. Support Vector Machines (SVM) забезпечили високу точність прогнозів, але їхня складність вимагає значних обчислювальних ресурсів. Нейронні мережі, зокрема Convolutional Neural Networks (CNN), показали добрі результати у прогнозуванні ефективності проектів, особливо при роботі з великими обсягами даних. Long Short-Term Memory (LSTM) та Bidirectional LSTM виявилися найбільш ефективними моделями для прогнозування, демонструючи високу точність та мінімальні прогнознi похибки.

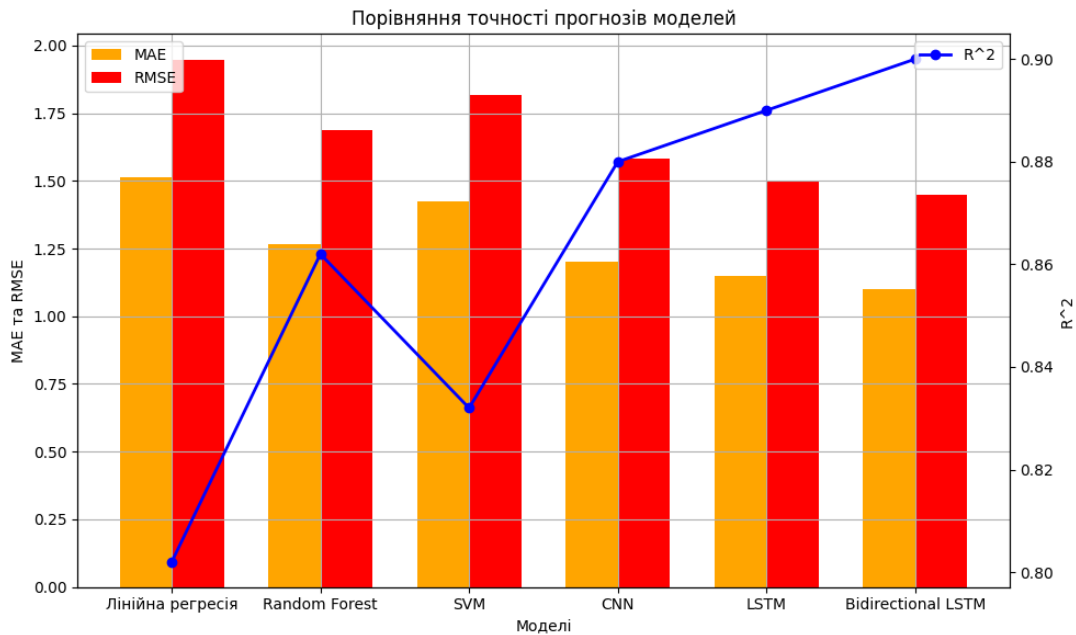


Рисунок 1 - Порівняння точності прогнозів моделей машинного навчання за показниками MAE, RMSE та R² (створено у середовищі Python)

Фактори, що впливають на точність прогнозування:

1. Архітектура моделі: Нейронні мережі, такі як LSTM та Bidirectional LSTM, демонструють найвищу точність завдяки здатності враховувати залежності в часових рядах та навчанню на довгих послідовностях даних. Ці моделі ефективно розпізнають патерни в даних, що дозволяє їм точно прогнозувати майбутні значення.

2. Обсяг та якість даних: Висока точність моделей забезпечується завдяки великому обсягу історичних даних та їх високій якості. Чим більше даних доступно для навчання моделі, тим точніші будуть її прогнози.

3. Гіперпараметри моделі: Правильний вибір гіперпараметрів, таких як кількість шарів і нейронів у моделі, значно впливає на її продуктивність. Оптимізація гіперпараметрів може суттєво покращити точність прогнозів.

4. Методи передобробки даних: Використання методів фільтрації та нормалізації даних підвищує точність прогнозів. Передобробка даних допомагає усунути шум і аномалії, забезпечуючи більш чисті та якісні вхідні дані для навчання моделі.

Графік на Рисунку 2 нижче демонструє порівняння прогнозів моделей машинного навчання з фактичними значеннями цільової змінної. Чорні точки представляють ре-

льні дані, тоді як лінії синього, зеленого та червоного кольорів показують прогнози моделей лінійної регресії, Random Forest та Support Vector Machines (SVM) відповідно.

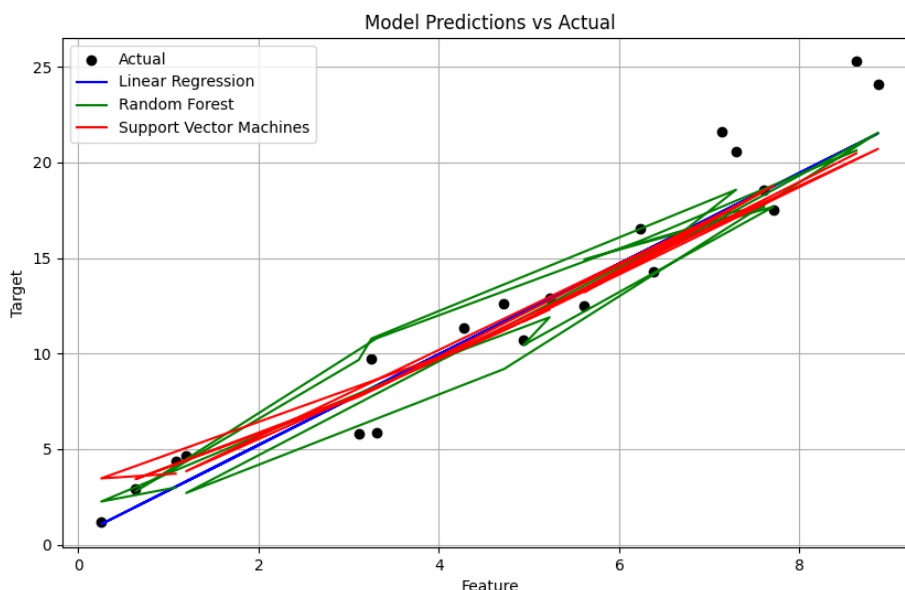


Рисунок 2 - Порівняння прогнозів моделей Лінійної регресії, Random та Support Vector Machines з фактичними даними (створено у середовищі Python)

Модель Random Forest демонструє найкращу відповідність фактичним даним, показуючи високу точність, хоча інколи і виявляє схильність до перенавчання. Лінійна регресія, хоча і намагається збігатися з середніми значеннями, часто недооцінює високі та переоцінює низькі значення. SVM показує проміжні результати, пропонуючи збалансовану точність, але все ще поступається Random Forest у точності відображення фактичних даних. Загалом, аналіз показує, що Random Forest є найбільш надійною моделлю для точного прогнозування у цьому контексті.

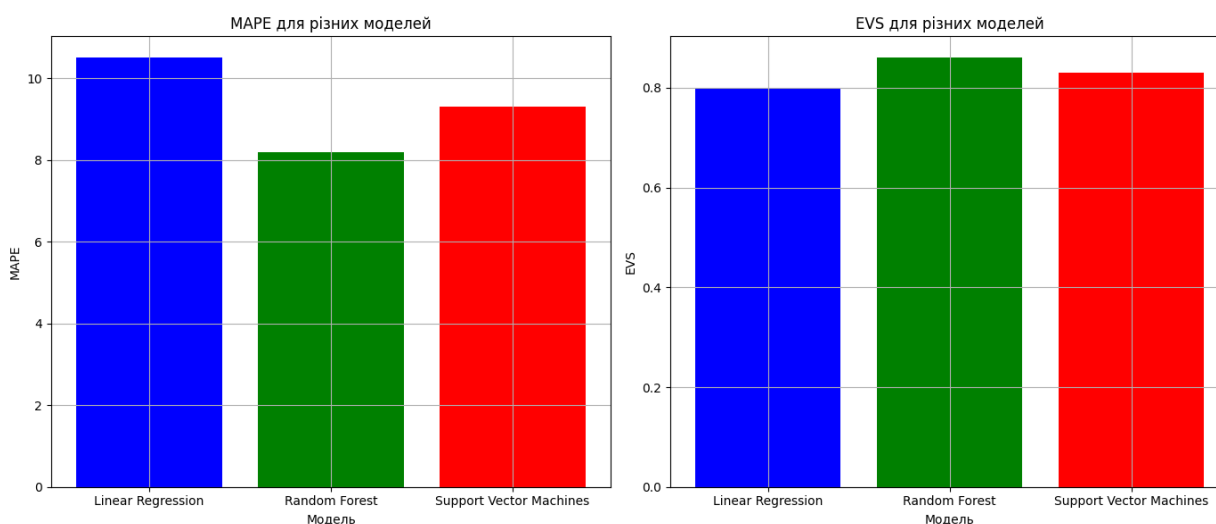


Рисунок 3 - Показники MAPE та EVS для різних моделей (створено у Python)

Графік на рисунку 3 демонструє показники MAPE та EVS для різних моделей. На основі аналізу показників MAPE та EVS, можна зробити висновок, що модель Random

Forest є найефективнішою серед представлених моделей, демонструючи найнижчі значення MAPE і найвищі значення EVS. Модель лінійної регресії показує найгірші результати за обома метриками, тоді як SVM знаходиться між ними, демонструючи середню ефективність. Це підтверджує важливість вибору відповідної моделі машинного навчання для досягнення максимальної точності та ефективності прогнозування.

Висновки. Результати цього дослідження підтверджують, що моделі машинного навчання, такі як Random Forest та LSTM, мають значний потенціал для покращення ефективності управління проектами за методологією Agile. Виявлена висока точність прогнозів та здатність цих моделей враховувати різноманітні нелінійні взаємозв'язки між змінними, що робить їх незамінними інструментами в сучасному динамічному середовищі розробки програмного забезпечення.

З огляду на результати дослідження, рекомендується впроваджувати ці моделі для прогнозування та управління ризиками у великих і складних проектах. Їх використання може суттєво підвищити ефективність та надійність планування, дозволяючи командам оперативно адаптуватися до змін вимог та зовнішніх умов. Це, в свою чергу, сприятиме зниженню ймовірності зривів термінів і перевитрати ресурсів. Важливим аспектом є інтеграція моделей машинного навчання у вже існуючі інструменти та платформи, які використовуються для управління проектами за методологією Agile. Це забезпечить безшовний перехід до нових технологій і мінімізує опір змінам з боку користувачів. Основні обмеження цього дослідження включають обмежену кількість аналізованих даних та можливі відмінності у результатах залежно від специфіки проектів. Майбутні дослідження можуть зосередитися на інтеграції інших моделей глибокого навчання, таких як GAN (Generative Adversarial Networks) або Transformer-based моделі, що можуть надати додаткові переваги у специфічних умовах різних галузей.

Загалом, інтеграція передових моделей машинного навчання у методології Agile є перспективним напрямком, який здатен суттєво покращити управління проектами, знизити ризики та підвищити якість кінцевих продуктів. Рекомендується активно продовжувати дослідження в цій галузі та впроваджувати отримані результати у практику для досягнення найкращих результатів.

ЛІТЕРАТУРА

1. ZenHub Blog. What does 2024 hold for agile? [Електронний ресурс]. Режим доступу: <https://blog.zenhub.com/what-does-2024-hold-for-agile/> (дата звернення: 02.08.2024).
2. Veido, D., Misnevs, B., Plotkin, A. The Method of Agile Projects Success Evaluation Using Machine Learning // Kabashkin, I., Yatskiv, I., Prentkovskis, O. (eds). Reliability and Statistics in Transportation and Communication. RelStat 2019. Lecture Notes in Networks and Systems. Springer, Cham. 2020. Vol. 117. https://doi.org/10.1007/978-3-030-44610-9_47
3. Westergaard, G.; Erden, U.; Mateo, O.A.; Lampo, S.M.; Akinci, T.C.; Topsakal, O. Time Series Forecasting Utilizing Automated Machine Learning (AutoML): A Comparative Analysis Study on Diverse Datasets. Information. 2024. Vol. 15. Article 39. <https://doi.org/10.3390/info15010039>

4. Casolaro, A.; Capone, V.; Iannuzzo, G.; Camastra, F. Deep Learning for Time Series Forecasting: Advances and Open Problems. *Information*. 2023. Vol. 14. Article 598. <https://doi.org/10.3390/info14110598>
5. Wasserbacher, H., Spindler, M. Machine learning for financial forecasting, planning and analysis: recent developments and pitfalls. *Digit Finance*. 2022. Vol. 4, No. 2. P. 63–88. <https://doi.org/10.1007/s42521-021-00046-2>
6. Potgieter, P.H. Machine Learning and Forecasting: A Review // Alleman, J., Rappoport, P., Hamoudia, M. (eds). *Applied Economics in the Digital Era*. Palgrave Macmillan, Cham. 2020. https://doi.org/10.1007/978-3-030-40601-1_8
7. Haidabrus, B. Generative AI in Agile, Project, and Delivery Management // Ivanov, V., Trojanowska, J., Pavlenko, I., Rauch, E., Pitel', J. (eds). *Advances in Design, Simulation and Manufacturing VII. DSMIE 2024. Lecture Notes in Mechanical Engineering*. Springer, Cham. 2024. https://doi.org/10.1007/978-3-031-61797-3_9
8. Cabrero-Daniel, Beatriz. AI for Agile Development: a Meta-Analysis [Електронний ресурс]. ResearchGate, 2023. Режим доступу: https://www.researchgate.net/publication/370776199_AI_for_Agile_development_a_Meta-Analysis (дата звернення: 02.08.2024).
9. Anas Bahi, Jihane Gharib, Youssef Gahi. Integrating Generative AI for Advancing Agile Software Development and Mitigating Project Management Challenges // *International Journal of Advanced Computer Science and Applications (IJACSA)*. 2024. Vol. 15, No. 3. <http://dx.doi.org/10.14569/IJACSA.2024.0150306>
10. Savio, R., Ali, J. Artificial Intelligence in Project Management & Its Future // *Saudi Journal of Engineering and Technology*. 2023. <https://doi.org/10.36348/sjet.2023.v08i10.002>
11. Taboada, I., Daneshpajouh, A., Toledo, N., de Vass, T. Artificial Intelligence Enabled Project Management: A Systematic Literature Review. *Appl. Sci*. 2023. Vol. 13. Article 5014. <https://doi.org/10.3390/app13085014>
12. Nenni, M.E., De Felice, F., De Luca, C. et al. How artificial intelligence will transform project management in the age of digitization: a systematic literature review. *Manag Rev Q*. 2024. <https://doi.org/10.1007/s11301-024-00418-z>

REFERENCES

1. ZenHub Blog. (2024). What does 2024 hold for agile? [Electronic resource]. Available at: <https://blog.zenhub.com/what-does-2024-hold-for-agile/> (Accessed: 2 August 2024).
2. Veido, D., Misnevs, B., Plotkin, A. The Method of Agile Projects Success Evaluation Using Machine Learning // Kabashkin, I., Yatskiv, I., Prentkovskis, O. (eds). *Reliability and Statistics in Transportation and Communication. RelStat 2019. Lecture Notes in Networks and Systems*. Springer, Cham. 2020. Vol. 117. https://doi.org/10.1007/978-3-030-44610-9_47
3. Westergaard, G.; Erden, U.; Mateo, O.A.; Lampo, S.M.; Akinci, T.C.; Topsakal, O. Time Series Forecasting Utilizing Automated Machine Learning (AutoML): A Comparative Analysis Study on Diverse Datasets. *Information*. 2024. Vol. 15. Article 39. <https://doi.org/10.3390/info15010039>

4. Casolaro, A.; Capone, V.; Iannuzzo, G.; Camastra, F. Deep Learning for Time Series Forecasting: Advances and Open Problems. *Information*. 2023. Vol. 14. Article 598. <https://doi.org/10.3390/info14110598>
5. Wasserbacher, H., Spindler, M. Machine learning for financial forecasting, planning and analysis: recent developments and pitfalls. *Digit Finance*. 2022. Vol. 4, No. 2. P. 63–88. <https://doi.org/10.1007/s42521-021-00046-2>
6. Potgieter, P.H. Machine Learning and Forecasting: A Review // Alleman, J., Rappoport, P., Hamoudia, M. (eds). *Applied Economics in the Digital Era*. Palgrave Macmillan, Cham. 2020. https://doi.org/10.1007/978-3-030-40601-1_8
7. Haidabrus, B. Generative AI in Agile, Project, and Delivery Management // Ivanov, V., Trojanowska, J., Pavlenko, I., Rauch, E., Pitel, J. (eds). *Advances in Design, Simulation and Manufacturing VII. DSMIE 2024. Lecture Notes in Mechanical Engineering*. Springer, Cham. 2024. https://doi.org/10.1007/978-3-031-61797-3_9
8. Cabrero-Daniel, B. (2023). AI for Agile Development: A Meta-Analysis [Electronic resource]. ResearchGate. Available at: https://www.researchgate.net/publication/370776199_AI_for_Agile_development_a_Meta-Analysis (Accessed: 2 August 2024).
9. Anas Bahi, Jihane Gharib, Youssef Gahi. Integrating Generative AI for Advancing Agile Software Development and Mitigating Project Management Challenges // *International Journal of Advanced Computer Science and Applications (IJACSA)*. 2024. Vol. 15, No. 3. <http://dx.doi.org/10.14569/IJACSA.2024.0150306>
10. Savio, R., Ali, J. Artificial Intelligence in Project Management & Its Future // *Saudi Journal of Engineering and Technology*. 2023. <https://doi.org/10.36348/sjet.2023.v08i10.002>
11. Taboada, I., Daneshpajouh, A., Toledo, N., de Vass, T. Artificial Intelligence Enabled Project Management: A Systematic Literature Review. *Appl. Sci*. 2023. Vol. 13. Article 5014. <https://doi.org/10.3390/app13085014>
12. Nenni, M.E., De Felice, F., De Luca, C. et al. How artificial intelligence will transform project management in the age of digitization: a systematic literature review. *Manag Rev Q*. 2024. <https://doi.org/10.1007/s11301-024-00418-z>

Received 19.10.2024.
Accepted 25.10.2024.

Forecasting the efficiency of agile projects using machine learning

The aim of the research is to address current challenges in the development of information systems and the integration of artificial intelligence technologies to optimize Agile methodologies. The study focuses on modeling and forecasting the efficiency of Agile processes, taking into account changing requirements over time and random deviations. Forecasting is understood as the process of estimating the future efficiency of Agile processes based on the analysis of time series and machine learning models. Efficiency forecasting is considered as time series forecasting, where parameters adapt depending on the system and data, allowing the model to generalize and take into account various factors affecting project outcomes.

The study examines the selection of the optimal model for forecasting Agile process efficiency by comparing machine learning methods such as Support Vector Machines (SVM),
ISSN 1562-9945 (Print) 183
ISSN 2707-7977 (Online)

Random Forest, and neural networks (CNN, LSTM). Comparative analysis showed that neural networks (LSTM, Bidirectional LSTM, CNN) are more effective for predicting project success, demonstrating high accuracy and lower errors. The practical value of the research lies in identifying LSTM and other neural network architectures as effective tools for predicting project success. This can serve as a guideline for implementing effective management systems in real-world conditions. For example, the application of LSTM in large IT companies for forecasting the success of Agile projects has significantly improved planning accuracy and reduced risks.

The research confirms that the integration of machine learning models, particularly LSTM and CNN, significantly improves the accuracy of forecasting and managing Agile projects. The application of these technologies can greatly enhance project management efficiency in information systems.

Keywords: Agile, forecasting, neural networks, artificial intelligence, information systems, machine learning, LSTM.

Палій Владислав Ігорович - аспірант кафедри цифрової економіки та системного аналізу Державного торговельно-економічного університету, м.Київ. v.paliy@knute.edu.ua.

Vladyslav Paliy - PhD Student, Department of digital economy and systems analysis, State University of Trade and Economics, Kyiv, Ukraine. v.paliy@knute.edu.ua.

**ДОСЛІДЖЕННЯ МОДЕЛЕЙ БАЗ ДАНИХ
ДЛЯ ЗБЕРІГАННЯ BIG DATA У СФЕРІ МОНІТОРИНГУ ДОВКІЛЛЯ**

Анотація. Питання, пов'язані з утворенням великих масивів даних, до кінця не вирішені. Кількість інформації у світі постійно збільшується, що призвело до виникнення проблеми її збереження. Створене для визначення цих даних поняття «big data» включає у собі наступні характеристики: кількість, швидкість обробки, різноманітність, достовірність, змінність та цінність. До цього типу інформації відносяться характеристики навколишнього середовища; дані визначають розподіл відповідних показників на Землі і дають можливість дати прогноз на майбутнє щодо зміни їх у часі і просторі, що є важливим для ведення господарства та сталого розвитку людства. Але інформація щодо ефективної організації збереження і обробки таких даних є недостатньою, що вимагає проведення додаткових досліджень. Таким чином, об'єктом дослідження є дані, що отримують на постах спостереження за станом навколишнього середовища. Предметом дослідження є зберігання даних, отриманих у результаті проведення моніторингу навколишнього середовища. Метою дослідження є розробка критеріїв для оцінки та порівняння різних типів сховищ даних з урахуванням специфічних вимог для їх зберігання; визначення типів інформації, яка буде зберігатися у базі даних параметрів навколишнього середовища; створення ER-діаграми визначеної бази даних. Отримані дані класифікуються в залежності від стану середовища, його розташування та забрудненості. Оскільки дані одержують з розгалуженої системи спостережень, вони поетапно потрапляють від місця їх реєстрації через міську, регіональну, державну і глобальну мережу до місця їх зберігання. Відповідно, повинні бути забезпечені наступні критерії до отриманої інформації: зберігання даних різних типів, швидкий доступ і обробка та можливість масштабування. Існує дві основні моделі сховищ даних: реляційні та нереляційні, кожна з яких має переваги і недоліки. Так, реляційні (SQL) мають жорсткі схеми, які забезпечують надійність збереження інформації, але є неефективними для обробки великої кількості запитів і не володіють значною масштабованістю. Нереляційні (NoSQL) зберігають дані у неструктурованому типі, легко масштабуються, і забезпечують велику швидкість обробки запитів. Висновки. Проведено дослідження щодо організації даних, отриманих з пунктів спостереження за навколишнім середовищем. Створено схему обробки отриманих даних. Окреслено групи параметрів, що будуть зберігатися у базі даних. Сформульовано основні критерії щодо зберігання даних, що дозволяють більш ефективно їх організувати. Реалізовано ER-діаграму для бази даних.

Ключові слова: інформація, база даних, big data, модель зберігання даних, вимірювання параметрів, навколишнє середовище, ER-діаграма.

Постановка проблеми. Кількість даних, що створюється та використовується людством, безперестанно зростає і обчислюється вже, за різними оцінками, у кількості від 10^6 до 10^9 байт, причому прогнозується їх збільшення до 2 екзабайт к 2025 року. Так, починаючи приблизно з 2009 року, вони подвоюються орієнтовно кожні 12 місяців [1]. Це створило проблему з їх зберіганням, обробкою та використанням. Для кожного окремого випадка необхідно визначати критерії та особливості зберігання інформації, створювати свій конкретний вид бази даних.

Аналіз останніх досліджень і публікацій. Будь-яка особа і організація, що працює у сферах державного управління, бізнесу, економіки і маркетингу, журналістики, медицини, біології, метеорології, інтернету, освіти використовує комп'ютери та інші цифрові пристрої, кожна хвилину продукує нову інформацію, яка потребує додатковий об'єм для зберігання та обробки [2-7]. Найбільшими розмірами володіють астрономічні та біологічні матеріали, зображення, звуки та відео [1, 8]. Для таких масивів даних було створено окреме поняття – «big data» [9], що включає, крім розмірів самої інформації, також джерела її отримання, апаратне і програмне забезпечення, інформаційні технології (засоби обробки і зберігання), методи аналізу та ін., тобто не існує чіткого визначення цього поняття. Спочатку вважалося, що достатніми основними ознаками «big data» є три «V»: volume – значні обсяги інформації; velocity – швидкість обробки (швидкість отримання даних та швидкість обробки запитів); variety – різноманітність (відмінності форматів, неструктурованість і неможливість описання їх звичними методами). З часом до визначень додалися ще три «V» – veracity (достовірність), variability (змінність, мінливість інформації), value (цінність).

У теперішній час до «великих даних» можна також віднести інформацію про стан та характеристики навколишнього середовища (температура, вологість, швидкість вітру, вміст хімічних речовин та ін.), що, починаючи з 60-70-х років ХХ століття, вимірюється на постійній основі на Земній кулі [10, 11]; в Україні також існує мережа служби моніторингу за станом атмосфери, гідросфери та літосфери. Отримання подібної інформації вкрай необхідно, оскільки дає можливість як встановити розподіл показників стану довкілля на земній кулі, так і зробити прогнози, що будуть впливати на функціонування багатьох галузей господарства. Так, від прогнозу погоди залежить робота авіаційного та морського транспорту, ведення сільського господарства. Крім цього, можна зробити висновки про перенесення забруднювачів в повітрі, воді та ґрунті; про вплив людини на навколишнє середовище; після цього визначити низку дій та кроків, спрямованих на розв'язання проблем погіршення стану навколишнього середовища. Це є важливим питанням сучасності, оскільки вплив людини на довкілля може стати причиною глобальної екологічної кризи на планеті і неможливості існування людства на Землі у майбутньому.

Вимірювання параметрів навколишнього середовища можна розділити на декілька категорій в залежності від мети дослідження: спостереження біля джерел забруднення, у містах та інших пунктах проживання людей, у сільській місцевості, вимірювання в умовно незабруднених місцях; контроль параметрів атмосфери, гідросфери, літосфери; вимірювання фізичних, хімічних параметрів тощо (рис. 1).



Рисунок 1 – Класифікація типів вимірювань параметрів навколишнього середовища

Концентрація речовин у викидах та біля підприємств та населених пунктах значно вище, ніж у сільській та незабрудненій місцевості. Викиди промисловості можуть містити цілу низку специфічних компонентів, які вступають у взаємодію з фоновими речовинами, тому необхідно проводити додаткові вимірювання цих компонентів. Отже щільність системи спостережень більш густа у заселених і промислових регіонах і майже відсутня над океанами і пустелями.

Отримання даних відбувається на стаціонарних та пересувних постах спостереження. Існують автоматизовані системи спостереження і контролю за якістю атмосферного повітря і пости ручного відбору проб та лабораторії, що обслуговуються операторами. Звичайно, використання автоматичних вимірювань дає більш швидкі та достовірні результати, оскільки не залежить від вимірювача або оператора. Але у багатьох випадках існують як неавтоматичні прилади (наприклад, аналогові), так і використовуються фізико-хімічні методи, що вимагають участі людини. Це призводить до утворення інформації різних типів та розбіжностей одиниць вимірювання. Крім того, дані можуть містити незв'язану інформацію: не на усіх пунктах вимірювання визначають весь можливий діапазон типів даних, що пов'язано з відсутністю даного компонента у повітрі, а також зі складністю або коштовністю обладнання, яке розміщують, наприклад, тільки на стаціонарних постах спостережень.

Метою дослідження є розробка критеріїв для оцінки та порівняння різних типів сховищ даних з урахуванням специфічних вимог для їх зберігання, визначення видів інформації для даного виду бази даних та створення ER-діаграми даної бази даних.

Виклад основного матеріалу дослідження. На міських станціях контролю за якістю атмосфери визначають наступне (рис. 2):

- фізичні параметри: температура, тиск, вологість, вітер (швидкість, напрямок, роза вітрів);

- хімічні параметри: вміст сульфуру(IV) оксиду, карбону(II) оксиду, нітрогену(IV) оксиду, амоніаку, сульфуру гідрогену, фенолу, гідрогену флуориду, формальдегіду, гідрогену хлориду;

- розташування місця вимірювання: дата, час, GPS координати.

На міських станціях контролю за якістю води визначають:

- гідрологічні параметри: витрата води, швидкість течії, рівень води;

- гідрохімічні параметри: температура, кольоровість, прозорість та зважені речовини, запахи, радіоактивність, розчинені гази (кисень, діоксид карбону), концентрації головних іонів (хлориди, сульфати, гідрокарбонати кальцію, магнію, натрію, калію), неорганічні сполуки (фосфор, азот, арсен, плюмбум, меркурій, купрум, кадмій, хлор, флуор тощо), рН, органічні сполуки та нафтопродукти.



Рисунок 2 – Досліджувані параметри атмосфери та гідросфери

При контролюванні якості ґрунтів визначають: вологість, рН, нітрати, пестициди, нафтопродукти, важкі метали, бактерії та інше.

Для промислових об'єктів додатково визначається параметри виробництва: назва, вид та технологія виробництва, місцезнаходження джерел викиду та їх кількість, параметри викиду (витрата, температура та ін.).

Всі системи об'єднуються у промислові, міські, регіональні, державні та глобальні мережі.

Після потрапляння інформації до мережі здійснюється її первинна обробка (переведення одиниць вимірювання, визначення числових характеристик); статистична обробка (визначення ймовірнісних характеристик); порівняння інформації про стан з параметрами в інших регіонах; складання алгоритму прогнозування стану довкілля (рис. 3).

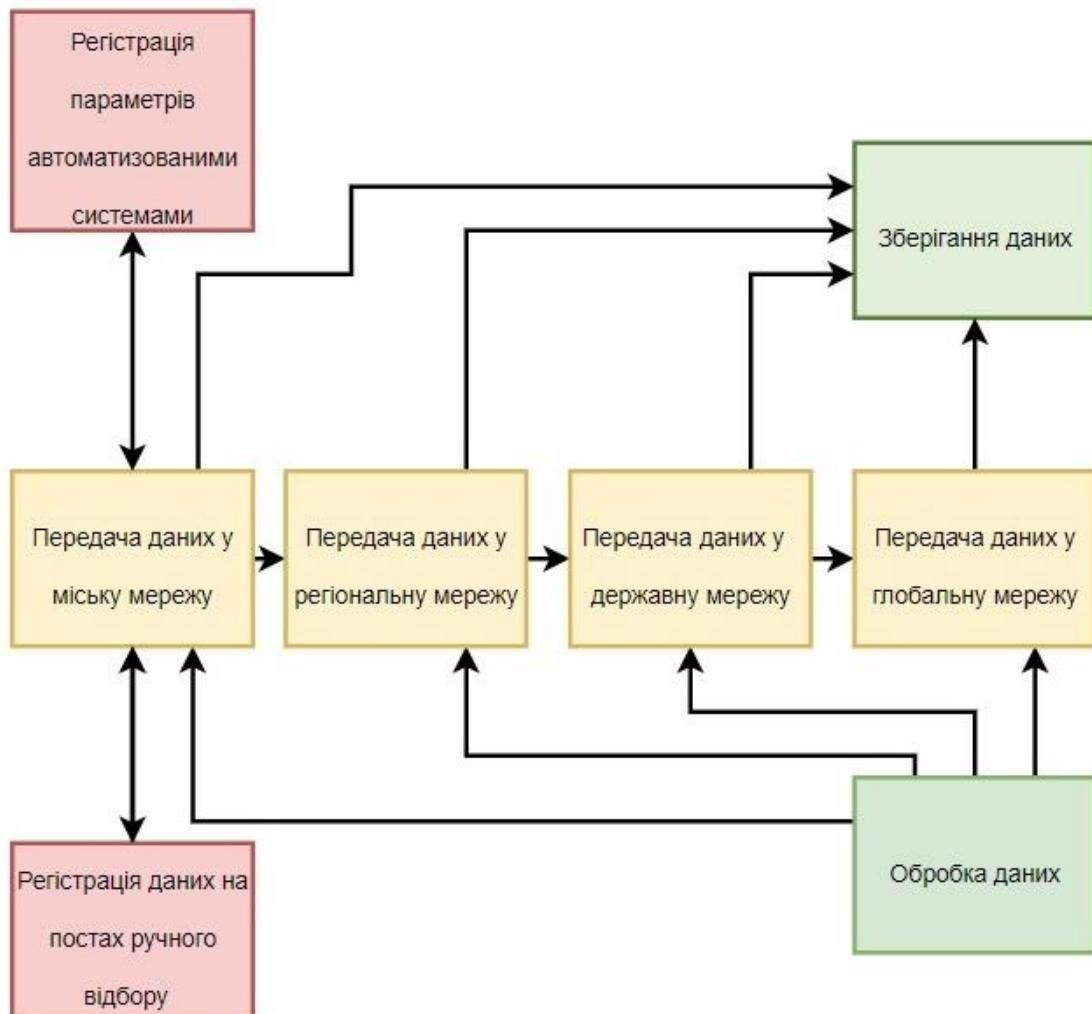


Рисунок 3 – Схема роботи з інформацією

Оскільки визначення параметрів середовища є комплексним і відбувається скрізь на земній кулі, необхідно, щоб дані постійно і швидко потрапляли у систему їх зберігання, були доступні для подальшої обробки, визначення характеристик та прогнозу.

Передавання даних на велику кількість серверів створює низку проблем:

- перевищення об'ємів інформації над продуктивністю пристрою, що їх опрацює;

- недостатня швидкість запису, переміщення, змін, обробки;
- різноманітність даних і складність їх впорядкування.

Таким чином, вибір моделі зберігання даних та моделі бази даних є вкрай важливим. Вибір моделі бази даних повинен враховувати наступні критерії:

- вид даних та їх структурованість;
- спосіб розташування та можливість масштабування;
- частота звернень до бази, оновлення та обробки інформації.

Як правило, бази даних, поділяють на дві категорії: реляційні та нереляційні.

У реляційних інформація зберігається у формі таблиць та зв'язків між ними. Кожна комірка повинна містити інформацію, вид якої визначається заздалегідь при проектуванні бази. Тобто дані є структурованими, і не передбачають швидких та частих змін. Перевагами таких баз даних є:

- простота представлення даних;
- використання правил проектування, заснованих на математичному апараті;
- незалежність даних.

Але вони мають і недоліки:

- не всі дані можуть бути представлені у табличному вигляді;
- необхідно, щоб усі комірки були заповнені інформацією;
- в результаті обробки даних з'являється множина таблиць, що ускладнює розуміння організації та структури бази даних і роботу з інформацією;
- база даних займає значне місце;
- швидкість обробки запитів при збільшенні розміру бази різко зменшується;
- для багатьох випадків обробка запитів даних, що розташовані на різних серверах, ускладнена.

У нереляційних базах інформація зберігається у неструктурованому вигляді, причому можна одразу поповнювати дані іншими типами, які до цього були відсутні [12-16]. Цей вид баз даних є масштабованим, і легко розподіляється на великій кількості серверів. Швидкість обробки даних і, відповідно, продуктивність, вище, ніж у реляційних баз даних. Таким чином, нереляційні:

- не зберігають інформацію у вигляді таблиць;
- використовують гнучкі схеми, що дають можливість пришвидшити розробку та забезпечити поетапну реалізацію;
- не використовують мову запитів SQL, що знайшло відображення у їх назві (NoSQL);
- підтримують можливість горизонтального масштабування і роботи з даними «big data»;
- дають можливість управляти даними у реальному часі;
- підтримують використання агрегатів і кластерів;
- застосовують динамічні гнучкі схеми.

Отже, для вибору моделі бази даних передусім потрібно розглянути характеристики даних, що необхідно розмістити у ній.

Таким чином, отримані в результаті вимірювання параметрів навколишнього середовища дані:

- майже не пов'язані одні з іншими, кожне вимірювання виконується окремо за допомогою відповідного обладнання та технічних засобів;
- заносяться до бази даних багатьма операторами або автоматично у місцях реєстрації інформації розгалуженої мережі спостережень;
- є сильно масштабованими;
- вимагають частого звернення до бази даних і обробки отриманої інформації.

Порівнюючи вимоги до даних та характеристики баз даних, можна зробити висновок, що більш ефективному збереженню та обробці даної інформації відповідають NoSQL бази даних.

Для визначених досліджуваних параметрів складено ER-діаграму бази даних (рис. 4).

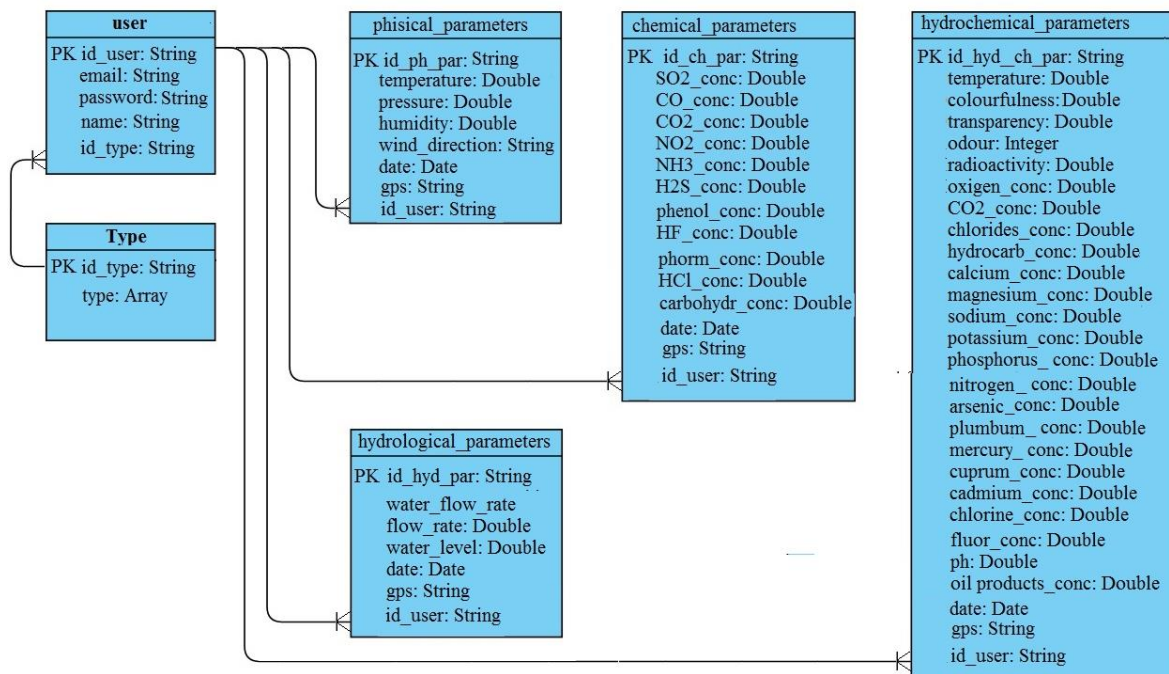


Рисунок 4 – Загальний вид ER-діаграми

Висновки. Проаналізовано моделі баз даних для зберігання інформації, отриманої на постах та пунктах спостережень за станом навколишнього середовища. Створено схему обробки отриманих даних.

Визначено критерії для даних та баз даних, які є наступними. Дані, що необхідно зберігати, є неструктурованими, оскільки їх кількість, вид та одиниці вимірювання залежить від багатьох факторів. Система збереження інформації повинна часто оновлюватися, зберігати інформацію, обробляти і робити висновки. База даних повинна мати значне горизонтальне масштабування.

Ці вимоги дають можливість зробити висновок, що під час подальших вимірювань параметрів навколишнього середовища і зберігання отриманих даних доцільно використовувати NoSQL бази даних.

ЛІТЕРАТУРА

1. Stephens Z. D. Big Data: Astronomical or Genomical? / Z. D. Stephens, S. Y. Lee, F. Faghri [et al] // PLOS Biology. – 2015. – №13(7). – e1002195. URL: <https://doi.org/10.1371/journal.pbio.1002195>.
2. Власенко Р. В. Концепція Big data в Україні : перспектива застосування в державних органах / Р. В. Власенко // Державне управління. – 2017. – № 4(60). – С. 97-101.
3. Стешенко Н. Л. Технології big data як система сучасних методів політичного впливу / Н. Л. Стешенко, Є. В. Болотіна, К. С. Чоста // Регіональні студії. – 2022. – № 31. – С. 44-48.
4. Dubey R. Education and training for successful career in big data and business analytics / R. Dubey, A. Gunasekaran // Industrial and Commercial Training. – 2015. – №47(4). – С. 174–181. URL: <https://doi.org/10.1108/ict-08-2014-0059>.
5. Шкирта І. М. Технологія Big Data: сутність, можливості для бізнесу / І. М. Шкирта, В. Ф. Лазар // Науковий вісник Мукачівського державного університету. Серія "Економіка" : збірник наукових праць / гол. ред. Т.В. Черничко. – Мукачево : МДУ, 2019. – Випуск 2(12). – С. 51-56.
6. Полюга В. Технологія великих даних (Big Data): основні характеристики та перспективи використання в журналістиці / В. Полюга // Діалог: Медіастудії. – 2019. – № 25. – С. 144-154. URL: <https://doi.org/10.18524/2308-3255.2019.25.195>.
7. Baig M. I. Big data in education: a state of the art, limitations, and future research directions / M. I. Baig, L. Shuib, E. Yadegaridehkordi // Int. J. Educ. Technol. High Educ. – 2020. – № 17(44). URL: <https://doi.org/10.1186/s41239-020-00223-0>.
8. Santana I. NoSQL Solution for Bioinformatics Data Provenance Storage / I. Santana, W. Silva, M. Holanda // New Knowledge in Information Systems and Technologies. Cham: Springer, 2019. – P. 528-537. URL: https://doi.org/10.1007/978-3-030-16181-1_50.
9. Кірей К. О. Розвиток і трансформація поняття big data / К. О. Кірей // Вісник Черкаського державного технологічного університету. – 2019. – № 1. – С. 33-40.
10. Erturk E. The Role of Big (Green) Data and Data Analytics in Environmental Sustainability / E. Erturk // Sustainable Futures: Symposium, 24 November 2020: abstracts. – Hamilton, New Zealand. URL: <https://doi.org/10.13140/RG.2.2.19228.10887/1>.
11. Методи вимірювання параметрів навколишнього середовища: підруч. / [Г. І. Гринь, В. І. Мохонько, О. В. Суворін та ін.]. – Сєверодонецьк : вид-во СНУ ім. В. Даля, 2019. – 420 с.
12. Резніченко О. В. Використання NoSQL технологій для обробки великих обсягів даних / О.В. Резніченко, О.А. Ляшенко // Молодь: Наука та Інновації : XI Міжн. наук.-техн. конф. студ., асп. та мол. вчених, 22-24 листопада 2023 р.: зб. матер. – Дніпро: НТУ «Дніпровська політехніка», 2023. – С. 27-28.
13. Hnatushenko V. V. Non-relational approach to developing knowledge bases of expert system prototype / V. V. Hnatushenko, Vik. V. Hnatushenko, N. L. Dorosh, N. O. Solodka, O. A. Liashenko // Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu. – 2022. – № 2 (188). – P. 112 - 117. URL: <https://doi.org/10.33271/nvngu/2022-2/112>.

14. Ляшенко О. А. Сравнительный анализ выполнения запросов к серверам баз данных MYSQL и MONGODB / О. А. Ляшенко, О. О. Конашков, Н. А. Солодка // Вісник Херсонського національного технічного університету. – 2019. – № 4 (71). – С. 114 - 124. URL: <https://doi.org/10.35546/kntu2078-4481.2019.4.13>.
15. Ляшенко О. А. Анализ производительности баз данных PostgreSQL/PostGIS и MONGODB для геопространственных запросов / О. А. Ляшенко, С. Н. Литвинов, Н. А. Солодка // Вісник Кременчуцького національного університету імені Михайла Остроградського. – 2019. – № 6 (119). – Р. 60 – 67. URL: <https://doi.org/10.30929/1995-0519.2019.6.60-67>.
16. Головка А. О. Проектування застосунку щодо підбору житла з використанням реляційної та нереляційної баз даних / А. О. Головка, О. А. Ляшенко // Комп'ютерне моделювання та оптимізація складних систем (КМОСС-2023): VIII Міжнар. наук.-техн. конф., 1-3 листопада 2023 р.: матер. конф.– Дніпро: ДВНЗ УДХТУ, 2023. – С. 157-158. URL: <https://udhtu.edu.ua/wp-content/uploads/2023/11/zbirnyk-tez-kmoss-2023.pdf>.

REFERENCES

1. Stephens Z. D. Big Data:Astronomical or Genomical? / Z. D. Stephens, S. Y. Lee, F. Faghri [et al] // PLOS Biology.– 2015.– №13(7). – e1002195. URL: <https://doi.org/10.1371/journal.pbio.1002195>.
2. Vlasenko R. V. Kontsepsiia Big data v Ukraini : perspektyva zastosuvannia v derzhavnykh orhanakh / R. V. Vlasenko // Derzhavne upravlinnia. – 2017. – № 4(60). – С. 97-101.
3. Steshenko N. L. Tekhnolohii big data yak systema suchasnykh metodiv politychnoho vplyvu / N. L. Steshenko, Ye. V. Bolotina, K. S. Chosta // Rehionalni studii. – 2022. – № 31. – С. 44-48.
4. Dubey R. Education and training for successful career in big data and business analytics / R. Dubey, A. Gunasekaran // Industrial and Commercial Training. – 2015. – №47(4). – С. 174–181. URL: <https://doi.org/10.1108/ict-08-2014-0059>.
5. Shkyrta I. M. Tekhnolohiia Big Data: sutnist, mozhlyvosti dlia biznesu / I. M. Shkyrta, V. F. Lazar // Naukovyi visnyk Mukachivskoho derzhavnoho universytetu. Seriia "Ekonomika" : zbirnyk naukovykh prats / hol. red. T.V. Chernychko. – Mukachevo : MDU, 2019. - Vypusk 2(12). – С.51-56.
6. Poliuha V. Tekhnolohiia velykykh danykh (Big Data): osnovni kharakterystyky ta perspektyvy vykorystannia v zhurnalistytsi / V. Poliuha // Dialoh: Mediastudii. – 2019. – № 25. – С. 144-154. URL: <https://doi.org/10.18524/2308-3255.2019.25.195>.
7. Baig M. I. Big data in education: a state of the art, limitations, and future research directions / M. I. Baig, L. Shuib, E. Yadegaridehkordi // Int. J. Educ. Technol. High Educ. – 2020. – № 17(44). URL <https://doi.org/10.1186/s41239-020-00223-0>.
8. Santana I. NoSQL Solution for Bioinformatics Data Provenance Storage / I. Santana, W Silva, M. Holanda // New Knowledge in Information Systems and Technologies. Cham: Springer, 2019. – P. 528-537. URL: https://doi.org/10.1007/978-3-030-16181-1_50.
9. Kirei K. O. Rozvytok i transformatsiia poniattia big data / K. O. Kirei // Visnyk Cherkaskoho derzhavnoho tekhnolohichnoho universytetu. – 2019. – № 1. – С. 33-40.

10. Erturk E. The Role of Big (Green) Data and Data Analytics in Environmental Sustainability / E. Erturk // Sustainable Futures: Symposium abstracts (Hamilton, New Zealand, 24 November 2020). URL: <https://doi.org/10.13140/RG.2.2.19228.10887/1>.
11. Metody vymiriuvannia parametriv navkolyshnoho seredovyscha: pidruch. / [H. I. Hryn, V. I. Mokhonko, O. V. Suvorin ta in.]. – Sievierodonetsk : vyd-vo SNU im. V. Dalia, 2019. – 420 s.
12. Reznichenko O. V. Vykorystannia NoSQL tekhnolohii dlia obrobky velykykh obsiahiv danykh / O. V. Reznichenko, O. A. Liashenko // Molod: Nauka ta Innovatsii : XI Mizhn. nauk.-tekhn. konf. stud., asp. ta mol. vchenykh, 22-24 lystopada 2023 r.: zb. mater. – Dnipro: NTU «Dniprovska politekhnika», 2023. – S. 27-28.
13. Hnatushenko V. V. Non-relational approach to developing knowledge bases of expert system prototype / V. V. Hnatushenko, Vik. V. Hnatushenko, N. L. Dorosh, N. O. Solodka, O. A. Liashenko // Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu. – 2022. – № 2 (188). – P. 112 - 117. URL: <https://doi.org/10.33271/nvngu/2022-2/112>.
14. Liashenko O. A. Sravnitelnyj analiz vypolneniya zaprosiv k serveram baz danykh MYSQL i MONGODB / O. A. Liashenko, O. O. Konashkov, N. A. Solodkaia // Visnyk Khersonskoho natsionalnoho tekhnichnoho universytetu. – 2019. – № 4 (71). – С. 114 - 124. URL: <https://doi.org/10.35546/kntu2078-4481.2019.4.13>.
15. Liashenko O. A. Analiz proizvoditelnosti baz danykh POSTGRES/POSTGIS i MONGODB dlya geoprostranstvennykh zaprosiv / O. A. Liashenko, C. N. Lytvynov, N .A. Solodkaia // Visnyk Kremenchutskoho natsionalnoho universytetu imeni Mykhaila Ostrohradskoho. – 2019. – № 6 (119). – S. 60 – 67. URL: <https://doi.org/10.30929/1995-0519.2019.6.60-67>.
16. Holovko A. O. Proiektuvannia zastosunku shchodo pidboru zhytla z vykorystanniam reliatsiinoi ta nereliatsiinoi baz danykh / A .O. Holovko, O. A. Liashenko // Kompiuterne modeliuвання ta optymizatsiia skladnykh system (KMOSS-2023): VIII Mizhnar. nauk.-tekhn. konf., 1-3 lystopada 2023 r.: mater. konf. – Dnipro: DVNZ UDKhTU, 2023. – S. 157-158. URL: <https://udhtu.edu.ua/wp-content/uploads/2023/11/zbirnyk-tez-kmoss-2023.pdf>.

Received 10.11.2024.
Accepted 15.11.2024.

Database models for storing big data

The issues related to the formation of large data collections are not fully resolved. The amount of information in the world is constantly increasing, which has resulted in the problem of its storage. The term "big data" created to define this data includes the following characteristics such as quantity, processing speed, variety, reliability, variability and value. This type of information includes environmental characteristics; the data determine the distribution of relevant indicators on the Earth and make it possible to make a forecast for the future regarding their changes in time and space, which is important for economic management and sustainable development of humanity. However, there is not enough information on the effective organisation of the storage and processing of such data, and further research is needed. Thus, the object of the study is the data obtained at environmental monitoring stations. The subject of the study is the storage of data obtained as a result of

environmental monitoring. The purpose of the study is to develop criteria for evaluating and comparing different types of data repositories, taking into account specific requirements for their storage; to determine the types of information to be stored in the database; to create an ER-diagram of a particular database. The received data are classified according to the state of the environment, its location and pollution. As the data is obtained from an extended system of observations, it passes in stages from the place of its registration through the city, regional, state and global network to the place of its storage. Accordingly, the following criteria for the information received must be provided such as the ability to store data of various types, quick access and processing, and scalability. There are two main models of data bases such as relational and non-relational, each of them has its advantages and disadvantages. For example, relational (SQL) data storage systems have rigid schemes that ensure the reliability of information storage, but are inefficient for processing a large number of queries and have no significant scalability. Non-relational (NoSQL) systems store data in an unstructured type, are easily scalable, and provide high speed of query processing. Conclusions. The research has shown that non-relational databases are more appropriate for storing data obtained from environmental monitoring stations. A scheme for processing the data was created. The groups of parameters that will be stored in the database are outlined. The main criteria for data storage were developed, allowing for more efficient data organisation. An ER diagram for the database was implemented.

Keywords: information, database, big data, data storage model, measurement of parameters, environment, ER-diagram.

Резніченко Олег Вікторович – аспірант кафедри інформаційних систем, ННІ УДХТУ, Український державний університет науки та технологій.

Ляшенко Оксана Анатоліївна – к.т.н., доцент кафедри інформаційних систем, ННІ УДХТУ, Український державний університет науки та технологій.

Архипова Вікторія Вікторівна – к.т.н., старший науковий співробітник Наукового центру інноваційних технологій, ННІ УДХТУ, Український державний університет науки та технологій; доцент кафедри цивільної інженерії, технологій будівництва та захисту довкілля, Дніпровський державний аграрно-економічний університет.

Reznichenko Oleh – Postgraduate student of Department of Information Systems, Ukrainian State University of Science and Technologies.

Liashenko Oksana – Candidate of Technical Sciences, Associate Professor of the Department of Information Systems, Ukrainian State University of Science and Technologies.

Arkhylova Viktoriia – Candidate of Technical Sciences, Senior Research Officer at the Scientific Centre for Innovative Materials and Technologies, Associate Professor of the Department of Civil Engineering, Construction Technologies and Environmental Protection, Dnipro State Agrarian and Economic University.

УМОВИ ЗМЕНШЕННЯ ДОВЖИНИ ПАРАЛЕЛЬНИХ УПОРЯДКУВАНЬ ВЕРШИН СПЕЦІАЛЬНИХ ОРГРАФІВ ПРИ НАЯВНОСТІ ПЕРЕРИВАНЬ

Анотація. Прикладні задачі планування виробництва часто можна змодельовати у вигляді оптимізаційних задач на орграфах, що відносяться до задач теорії розкладів. Однією з таких є задача паралельного упорядкування, в якій необхідно мінімізувати сумарний час виконання усіх робіт, на порядок виконання яких може бути накладено технологічні обмеження, при заданих ресурсах. Для випадків, коли орграф задачі має спеціальну структуру, досліджено способи побудови на основі комбінаторних конфігурацій та визначено, за яких умов дозвіл переривань при виконанні робіт може зменшити значення цільової функції.

Ключові слова: теорія розкладів, задачі комбінаторної оптимізації, паралельні упорядкування, орієнтовані графи, гусениці, граціозна розмітка, переривання, комбінаторна конфігурація.

Постановка проблеми. В теорії розкладів вивчаються як теоретичні, так і прикладні задачі, пов'язані з плануванням виробничих процесів, проектуванням обчислювальних систем тощо. Їх математична постановка часто може бути зведена до оптимізаційних задач на орієнтованих графах, де комплексу необхідних до виконання робіт відповідає задана множина вершин, а дугами задається частковий порядок, що моделює відповідні технологічні обмеження. Однією з них є задача паралельного упорядкування вершин таких графів.

У класичній постановці задачі упорядкування розглядається ациклічний орграф $G(V,U)$, вершини якого мають однакові вагові коефіцієнти, що приймаються за 1. Також заданий один з параметрів паралельного упорядкування: довжина або ширина. Довжина позначається l і відображає загальний час завершення всіх робіт. Ширина визначає максимальну кількість робіт, які в будь-який момент часу можуть виконуватися паралельно, і має позначення h . Як правило, той параметр, який не заданий в початкових умовах, необхідно мінімізувати [1].

Класична постановка задачі упорядкування має важливе теоретичне значення, однак на практиці тривалість виконання робіт часто є різною. У зв'язку з цим, розглядаються випадки, де ваги вершин задаються множиною, що складається з натуральних чисел, які відображають відносні відмінності у тривалості їх виконання. Через це виникає питання доцільності та ефективності застосування переривань при виконанні робіт. Було показано, що переривання можуть значно покращувати значення цільової функції,

але здебільшого це залежить від структури конкретного графа. Встановлено, що дозвіл переривань впливає на оптимальність у випадках, коли граф задачі є паралельно-послідовним [2], дводольним [3], або має структуру деяких підкласів дерев, для яких була доведена можливість граціозної розмітки їх вершин [4].

В даній роботі розглядаються задачі, де при заданому орієнтованому графі G та ширині упорядкування h необхідно мінімізувати довжину l . Розглянуто один з підкласів дерев, який допускає граціозну розмітку, що має назву гусениці (або одностантне дерево). Прикладні задачі, в яких комплекс робіт може бути представлений графами такого виду, включає зокрема планування будівництва. Вершини ланцюга наприклад можуть моделювати побудову конкретних поверхів. Тоді приєднані до цього ланцюга вершини залежно від спрямованості дуг відповідають або проміжним етапам будівництва, або навпаки — роботам, які можуть виконуватися лише після того, як конкретний поверх побудовано.

Аналіз останніх досліджень і публікацій. Розв'язання прикладних задач із планування виробничих процесів, складання різноманітних розкладів, оптимізації обслуговуючих систем часто супроводжується математичними моделями, які застосовують елементи теорії графів [5]. Окрім того, актуальним залишається питання впливу дозволу переривань на оптимальність розкладу [6]. Оскільки певні прикладні задачі моделюються графами спеціальної структури, то доцільно розробляти теоретичні основи розв'язання задач для відповідних підкласів графів. Деякі з них є предметом досліджень в питаннях розміток графів, як то граціозної [7], або таких, при дослідженні яких застосовуються комбінаторні конфігурації [8].

Мета дослідження. Вивчити прикладні задачі, що моделюються оргграфами малодосліджених структур. Запропонувати методи знаходження оптимальних розв'язків у задачах упорядкування, де графом є орієнтоване дерево спеціальної структури. Визначити необхідні умови, за яких дозвіл переривань може зменшити значення цільової функції в цих задачах.

Виклад основного матеріалу дослідження. Поміж підкласів дерев, які допускають граціозну розмітку, окремої уваги заслуговують ланцюги та похідні від них — гусениці.

Означення 1. Ланцюгом називається дерево, в якому дві вершини мають степені рівні 1, а решта вершин — степені рівні 2.

Орієнтований ланцюг являє собою послідовність k вершин, поєднаних послідовно $(k-1)$ дугами $(v_1, v_2), (v_2, v_3) \dots (v_{k-1}, v_k)$. Очевидно, що при такій структурі графа розглядати задачі паралельного упорядкування недоцільно, адже незалежно від кількості вершин та їх вагових коефіцієнтів жодні дві вершини ланцюга не можуть стати на одне місце упорядкування.

На відміну від ланцюгів, паралельне упорядкування вершин допускають графи, що є похідними від них, а саме гусеницями.

Означення 2. Гусеницею називається дерево, яке після вилучення з нього всіх вершин степені 1 перетворюється на ланцюг [7].

При цьому довжина ланцюга не обов'язково зберігається.

Розглянемо різні варіанти структури гусениць в залежності від спрямованості дуг. Нехай гусениця утворена з орієнтованого ланцюга довжини k шляхом додавання до нього висячих вершин степенів 1, які з'єднуються вхідними дугами з вершинами ланцюга, утворюючи з ними кореневі дерева (рис. 1).

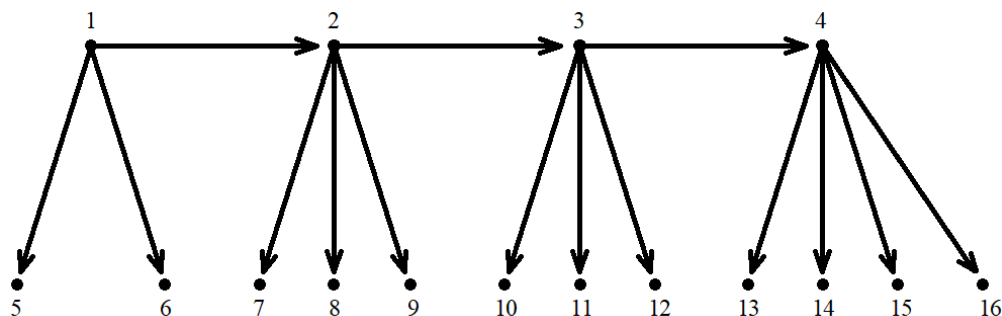


Рисунок 1 — Орієнтована гусениця із вихідними дугами

Вважатимемо, що вагові коефіцієнти усіх вершин є однаковими і рівні одиниці. В таких випадках на перші k місць упорядкування ставляться вершини заданого ланцюга. Решта вершин заносяться до упорядкування із урахуванням часткового порядку починаючи з вершин, які були приєднані до початку ланцюга. Для значень $h \geq 2$ якщо в побудованому оптимальному упорядкуванні без переривань на останньому місці стоїть менше ніж h вершин, то при дозволі переривань значення цільової функції може бути покращено за умови, що на двох останніх місцях стоять лише висячі вершини.

Для вершин графа з рис. 1 переривання зменшують довжину упорядкування у випадку $h = 2$.

$$h = 2: \left(\begin{array}{cccccccc} 1 & 2 & 6 & 7 & 8 & 10 & 12 & 14 & 16 \\ & 5 & 3 & 4 & 9 & 11 & 13 & 15 & \end{array} \right), l^* = 9;$$

$$h = 2: \left(\begin{array}{cccccccc} 1 & 2 & 6 & 7 & 8 & 10 & 12 & \{(14;0,5), (16;0,5)\} & (14;0,5) \\ & 5 & 3 & 4 & 9 & 11 & 13 & 15 & (16;0,5) \end{array} \right), l^* = 8,5.$$

При $h = 3$ переривання не сприяють зменшенню довжини.

$$h = 3: \left(\begin{array}{cccc} 1 & 2 & 3 & 4 & 11 & 14 \\ & 5 & 7 & 9 & 12 & 15 \\ & 6 & 8 & 10 & 13 & 16 \end{array} \right).$$

Інший варіант структури орієнтованої гусениці передбачає, що до орієнтованого ланцюга довжини k приєднані висячі вершини степенів 1, з яких прямують дуги до вершин цього ланцюга.

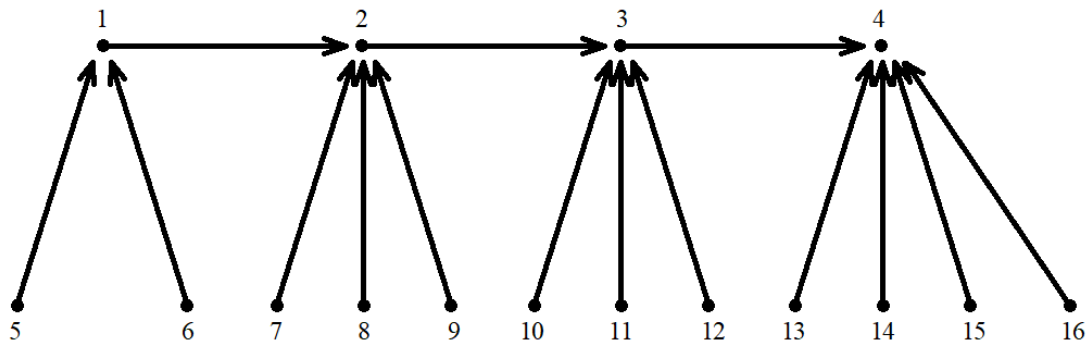


Рисунок 2 — Орієнтована гусениця із вхідними дугами

В такому випадку принцип побудови упорядкування буде наступним. Вершини ланцюга довжини k займають останні k місць в побудованому упорядкуванні. Далі починаючи з кінця упорядкування і рухаючись до його початку з урахуванням часткового порядку заносяться допустимі вершини, в тому числі на ті місця, які вже частково заповнені. Тоді якщо при деякому $h \geq 2$ на перших двох місцях упорядкування без переривань знаходяться лише вершини без вхідних дуг, а до того ж на першому місці їх кількість менша за h , то за рахунок дозволу переривань довжина упорядкування може бути зменшена. Побудоване таким чином упорядкування вершин графа з рис. 2 при $h = 2$ має наступний вигляд:

$$\begin{pmatrix} (5;0,5) & \{(5;0,5),(6;0,5)\} & 8 & 10 & 12 & 1 & 2 & 3 & 4 \\ (6;0,5) & 7 & 9 & 11 & 13 & 14 & 15 & 16 \end{pmatrix}.$$

Альтернативним способом побудови упорядкування для вершин даного графа є занесення на перші місця упорядкування вершин без вхідних дуг починаючи з тих, які приєднані до першої вершини ланцюга. Самі вершини ланцюга вносяться до упорядкування на перші для них допустимі місця після того, як усі відповідні приєднані до них уже розподілені.

$$h = 2: \begin{pmatrix} 5 & 1 & 8 & 2 & 11 & 3 & 14 & 16 & 4 \\ 6 & 7 & 9 & 10 & 12 & 13 & 15 \end{pmatrix}; h = 3: \begin{pmatrix} 5 & 1 & 2 & 12 & 3 & 4 \\ 6 & 8 & 10 & 13 & 15 \\ 7 & 9 & 11 & 14 & 16 \end{pmatrix}.$$

Тепер з'ясуємо, якими будуть упорядкування мінімальної довжини розглянутих структур графів у випадку, коли не всі значення вагових коефіцієнтів вершин є однаковими та чи вплине дозвіл переривань на оптимальність. Нехай вершини орієнтованого ланцюга довжини k мають довільні вагові коефіцієнти $p_1, p_2 \dots p_k$ такі, що $p_i \gg 1$ ($i = \overline{1, k}$), а решта вершин, приєднаних до цього ланцюга мають вагові коефіцієнти рівні 1. В разі, коли до вершини v_i орієнтованого ланцюга приєднується нова вершина v_r ($r > k$) введенням дуги (v_i, v_r) , маємо структуру графа з однією вершиною, що не має вхідних дуг. Приклад такого графа зображений на рис. 3.

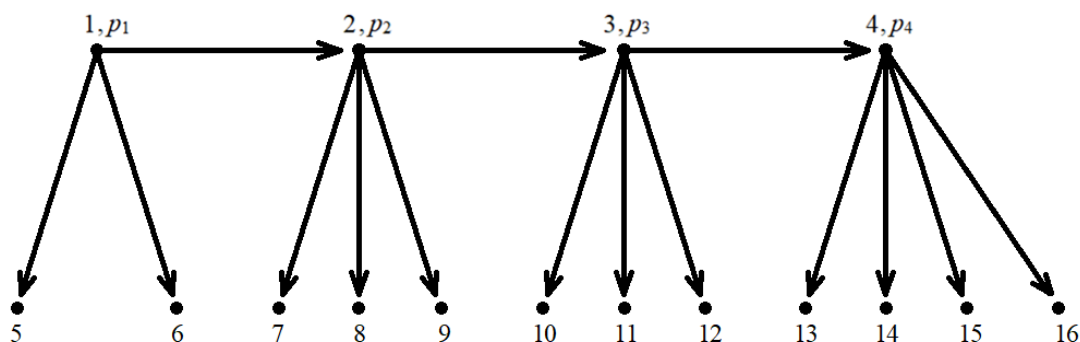
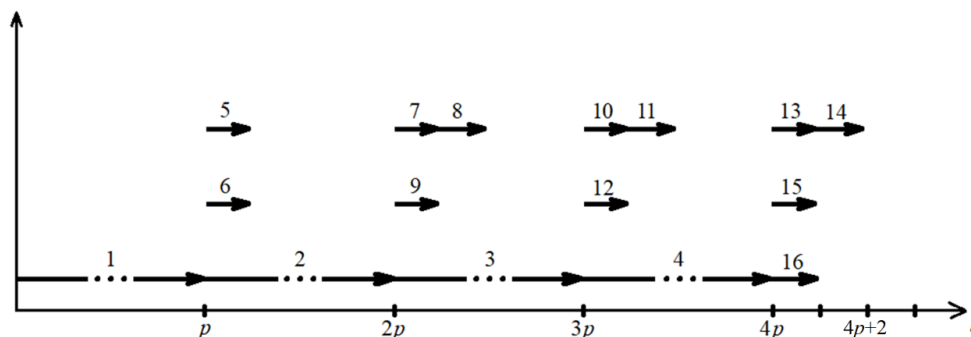


Рисунок 3 — Зважена гусениця із вихідними дугами

Оптимальне упорядкування вершин цього графа без переривань при заданій ширині $h = 3$ має довжину $l^* = \sum_{i=1}^4 p_i + 2$ і схематично виглядає наступним чином.

рину $h = 3$ має довжину $l^* = \sum_{i=1}^4 p_i + 2$ і схематично виглядає наступним чином.



За рахунок дозволу переривань довжина може бути скорочена до величини рівної

$l_{II}^* = \sum_{i=1}^4 p_i + 1\frac{1}{3}$, тобто менш ніж на одиницю, що вочевидь не є суттєвим покращенням,

враховуючи початкову умову $p_i \gg 1$. Очевидно, що у випадку, коли вершина v_r ($r > k$) приєднується до ланцюга введенням дуги (v_r, v_i) , іншими словами ми маємо граф з однією вершиною, що не має вихідних дуг, оптимальне упорядкування з перериваннями так само несуттєво змінить значення цільової функції.

Інший випадок, за якого не всі вершини мають однакові вагові коефіцієнти, передбачає, що навпаки, вершини ланцюга мають ваги рівні 1, а приєднані до них вершини відповідно — рівні p . На рис. 4 зображено структуру такого графа (де вагові коефіцієнти вершин 5-16 дорівнюють p).

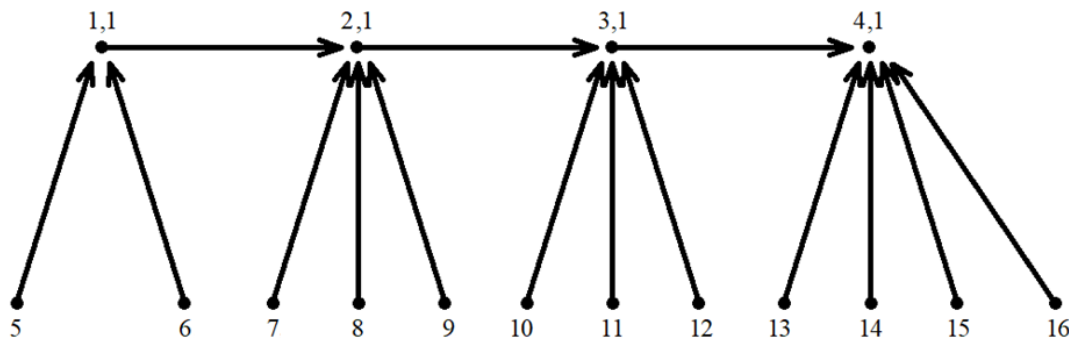
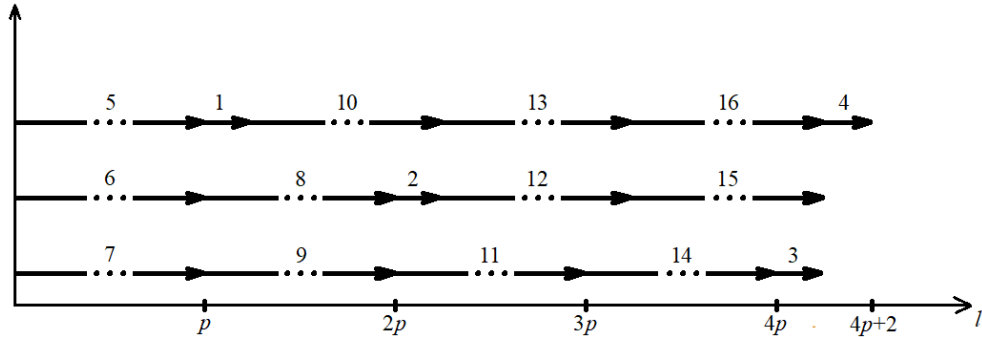


Рисунок 4 — Зважена гусениця із вхідними дугами

При $h = 3$ довжина оптимального упорядкування без переривань рівна $l^* = 4p + 2$, причому очевидно, що за рахунок переривань домогтися скорочення довжини неможливо, адже частково заповненим є лише останнє місце упорядкування, на якому розміщується остання вершина ланцюга.



Розглянемо як зміняться довжини упорядкувань з перериваннями та без них, якщо в графі з рис. 4 до однієї з вершин ланцюга приєднати ще одну (i для зручності перенумерувати вершини). Для визначеності, вважатимемо, що вона приєднана до вершини 2. Цей граф наведено на рис. 5.

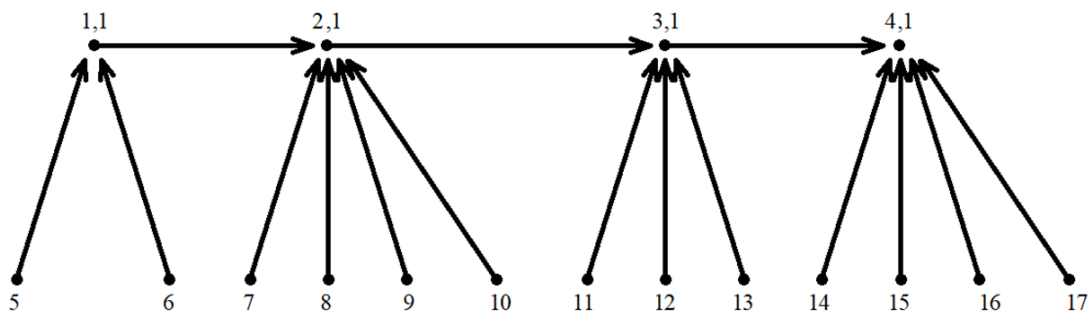
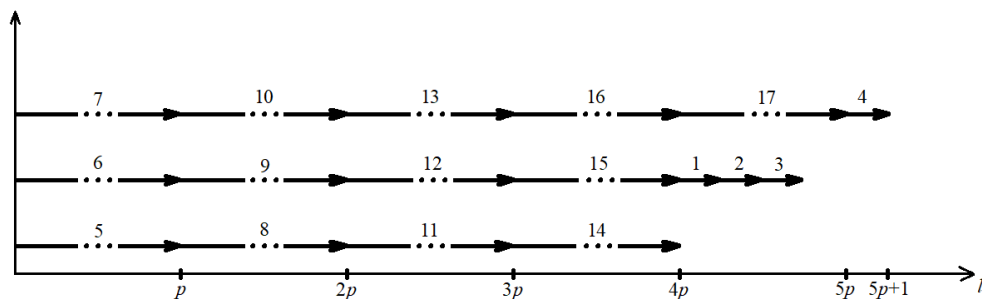
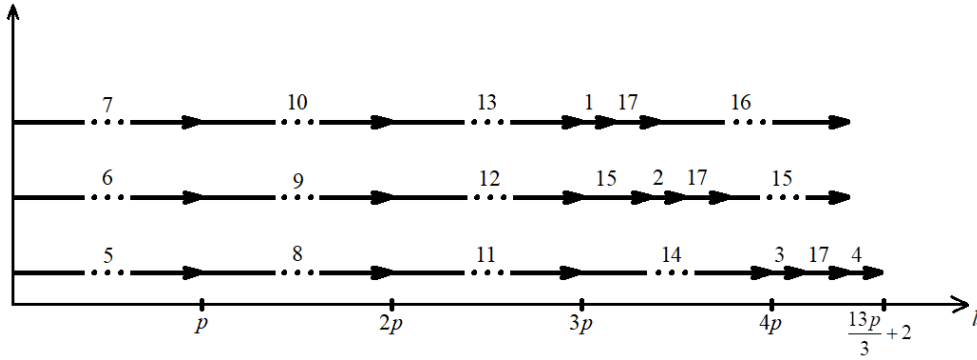


Рисунок 5 — Гусениця з додатково приєднаною вершиною

Упорядкування вершин цього графу без переривань при ширині $h = 3$ має наступну структуру.



При дозволених перериваннях отримуємо наступну структуру упорядкування.



При дозволених перериваннях довжина упорядкування скорочується з $l^* = 5p + 1$ до $l_{II}^* = 4\frac{1}{3}p + 2$. Різниця між цими величинами складає $\left(\frac{2}{3}p - 1\right)$, що може вважатися суттєвим покращенням значення цільової функції.

При розподілі тих вершин, які приєднані до орієнтованого ланцюга, пріоритетність при виборі кожної наступної вершини для розміщення визначається довжиною найдовшого шляху, який починається в цій вершині.

На наступному прикладі розглянемо, чи будуть доцільними переривання, якщо наприклад до вершини 1 приєднана лише одна, а до вершини 3 — більше половини вершин ваги p , тобто вони приєднані до ланцюга нерівномірно (рис. 6).

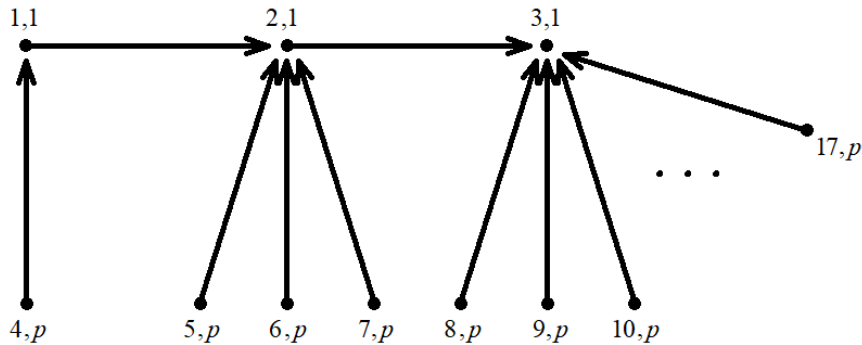
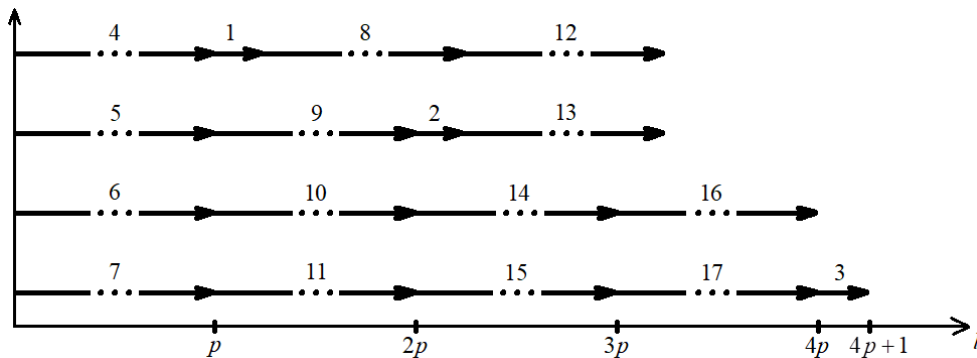


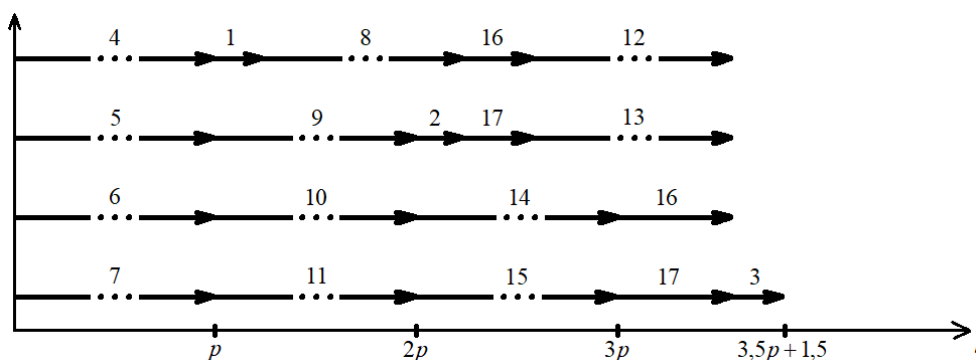
Рисунок 6 — Гусениця з перерозподіленими вхідними дугами

В даному графі вершини під номерами з 8 по 17 приєднані до вершини з номером 3. Схематично оптимальне упорядкування вершин даного графа без переривань при ширині рівній 4 має довжину $l^* = 4p + 1$ і виглядає наступним чином.



Дозволивши переривання, ця довжина може бути скорочена до величини рівної

$$l_{II}^* = 3\frac{1}{2}p + 1\frac{1}{2}, \text{ тобто на значення } \left(\frac{1}{2}p - \frac{1}{2}\right).$$



Таким чином на прикладах проілюстровано, що в загальному випадку множину вершин, що мають вагові коефіцієнти p можна розглядати в незалежності від того, до яких саме вершин ланцюга вони приєднані.

Якщо при приєднанні вершин ваги p суттєвим є врахування нумерації конкретних вершин, то кількість можливих варіантів розподілення приєднаних до ланцюга вершин може бути обчислена із застосуванням такої комбінаторної конфігурації, як сполучення. Представимо побудову графа у вигляді ітеративного процесу, при якому на кожному кроці до однієї з вершин ланцюга довжини k послідовно приєднується деяка кількість вершин ваги p . Нехай їх кількість для кожної i -ої вершини ланцюга складає m_i ($\forall i = \overline{1, k} \ m_i \geq 1$), відповідно загальна їх кількість $\sum_{i=1}^k m_i = m$. Для простоти запису почнемо приєднання починаючи з k -ої вершини у зворотному порядку. На першій ітерації

серед m вершин обирається m_k , кількість варіантів цього вибору складає $\frac{\left(\sum_{i=1}^k m_i\right)!}{m_k! \left(\sum_{i=1}^{k-1} m_i\right)!}$.

На наступній ітерації серед вершин, які залишилися обирається m_{k-1} , тобто кількість ва-

ріантів дорівнює $\frac{\left(\sum_{i=1}^{k-1} m_i\right)!}{m_{k-1}! \left(\sum_{i=1}^{k-2} m_i\right)!}$. Аналогічно для усіх наступних ітерацій, аж до приєд-

нання m_2 . На тому етапі, коли необхідно приєднати m_1 вершин, просто приєднується решта. Таким чином кількість способів приєднання вершин до орієнтованого ланцюга дорівнює

$$\sum_{j=0}^{k-2} \left(\frac{\left(\sum_{i=1}^{k-j} m_i\right)!}{m_{k-j}! \left(\sum_{i=1}^{k-j-1} m_i\right)!} \right).$$

Якщо при приєднанні вершин до ланцюга їх нумерація є несуттєвою, то потрібно враховувати лише можливу кількість вершин, які з'єднуються з відповідними вершинами ланцюга.

Сформулюємо умови, при яких переривання дозволяють зменшити значення цільової функції у вигляді наступної теореми.

Теорема. Дозвіл переривань при побудові оптимального упорядкування може зменшити значення цільової функції якщо:

1) множина вершин V ($|V| = n$) розбивається на дві неперетинні підмножини V_1 та V_2 ($|V_1| = n_1$, $|V_2| = n_2$, де $n_1 + n_2 = n$), $w_i \in \{c, p\}$ — вагові коефіцієнти вершин (де $i = \overline{1, n}$ та $p \gg c$), причому

$$\begin{cases} w_i = c, \text{ якщо } v_i \in V_1; \\ w_i = p, \text{ якщо } v_i \in V_2; \end{cases}$$

2) вершини підмножини V_1 утворюють орієнтований ланцюг;

3) $\forall v_j \in V_2 \exists!(v_j, v_k) \in U$, де $v_k \in V_1$;

4) $|V_2| > h$ та $\frac{|V_2|}{h} \notin N$.

При цьому $l^* - l_n^* > \frac{(h-q)}{h} p$, де q — остача від ділення $|V_2|$ на h .

Доведення. Оцінимо значення цільової функції в разі, якщо переривання заборонені. Неважко переконатися в тому, що $\left\lceil \frac{|V_2|}{h} \right\rceil p + c \leq l^* \leq \left\lceil \frac{|V_2|}{h} \right\rceil p + |V_1|c$. Дійсно, незалежно від того, до яких вершин орієнтованого ланцюга прямують дуги з вершин, що належать підмножині V_2 , завжди знайдеться спосіб їх розміщення в упорядкуванні, при якому повністю будуть заповнені $\left\lceil \frac{|V_2|}{h} \right\rceil p$ місць, а частково заповненими — відповідно ще p місць (за рахунок умови 4). З іншого боку, якщо наприклад всі вершини підмножини V_2 приєднані до першої вершини ланцюга, то вся підмножина вершин V_1 може бути розміщена в упорядкуванні лише послідовно після них, займаючи $|V_1|c$ місць. У разі, якщо до останньої вершини ланцюга приєднано не менше q вершин і $(|V_1| - 1)c \leq p$, то на p частково заповнених місць можна розмістити перші $(|V_1| - 1)$ вершини ланцюга, а відповідно останню — на місця безпосередньо після внесення решти вершин графа.

У випадку дозволених переривань оцінка довжини має наступний вигляд $\frac{|V_2|}{h} p + c \leq l^* \leq \frac{|V_2|}{h} p + |V_1|c$ з міркувань, аналогічних до випадку заборонених переривань. Єдиною відмінністю є те, що нижня оцінка досяжна за умови, коли до останньої

вершини ланцюга приєднано не менше $(h+q)$ вершин. Знайдемо різницю між відповідними значеннями довжини упорядкування. Для оцінки знизу:

$$l^* - l_{II}^* = \left\lceil \frac{|V_2|}{h} \right\rceil p + c - \left(\frac{|V_2|}{h} p + c \right) = \left\lceil \frac{|V_2|}{h} \right\rceil p - \frac{|V_2|}{h} p = \left(\left\lceil \frac{|V_2|}{h} \right\rceil - \frac{|V_2|}{h} \right) p.$$

Для оцінки зверху:

$$l^* - l_{II}^* = \left\lceil \frac{|V_2|}{h} \right\rceil p + |V_1|c - \left(\frac{|V_2|}{h} p + |V_1|c \right) = \left\lceil \frac{|V_2|}{h} \right\rceil p - \frac{|V_2|}{h} p = \left(\left\lceil \frac{|V_2|}{h} \right\rceil - \frac{|V_2|}{h} \right) p.$$

Тобто різниця буде однаковою в обох випадках. Враховуючи умову 4 представимо $|V_2|$ у вигляді $(mh + q)$, де $m \in N$. Тоді

$$\left(\left\lceil \frac{|V_2|}{h} \right\rceil - \frac{|V_2|}{h} \right) p = \left(\left\lceil \frac{mh+q}{h} \right\rceil - \frac{mh+q}{h} \right) p = \left(m + \left\lceil \frac{q}{h} \right\rceil - m - \frac{q}{h} \right) p = \left(\left\lceil \frac{q}{h} \right\rceil - \frac{q}{h} \right) p.$$

Оскільки q — остача від ділення на h , то очевидно, що $\frac{q}{h} < 1$ і відповідно

$\left\lceil \frac{q}{h} \right\rceil = 1$. З цього випливає, що

$$l^* - l_{II}^* = \left(1 - \frac{q}{h} \right) p = \left(\frac{h-q}{h} \right) p.$$

Висновки. Для одного з відомих підкласів неорієнтованих графів (гусениці) вперше розглянуто їх орієнтований аналог та запропоновано метод знаходження оптимального розв'язку однієї задачі упорядкування у випадку, коли граф, що задає частковий порядок, має відповідну структуру. Проаналізовано ефективність впливу переривань на значення цільової функції в залежності від орієнтації дуг, значень вагових коефіцієнтів вершин та структури конкретних графів.

Отримано теоретичні результати для оцінки впливу дозволу переривань при розв'язку задач упорядкування, коли граф є орієнтованою гусеницею, які сформульовано у вигляді теореми. Отримані результати мають не лише теоретичне значення в сфері розробки методів розв'язання задач комбінаторної оптимізації, але й з точки зору практичного застосування результатів до ряду прикладних задач.

ЛІТЕРАТУРА

1. Бурдюк В. Я. Алгоритмы параллельного упорядочения: учебное пособие. / В. Я. Бурдюк, В. А. Турчина. – Днепропетровск: ДГУ, 1985. 83 с.
2. Коваленко Є. О. Аналіз впливу структури графів на оптимальність розв'язку задач параллельного упорядкування з перериваннями [Електронний ресурс] / Є. О. Коваленко, В. А. Турчина // Питання прикладної математики і математичного моделювання. – 2021. – Вип. 21. – С. 130–137. – Режим доступу: <https://doi.org/10.15421/322113>.
3. Турчина В. А. Вплив початкових даних задачі параллельного упорядкування з перериваннями на оптимальність розв'язку [Електронний ресурс] / В. А. Турчина, Є. О. Ко-

валенко // Питання прикладної математики і математичного моделювання. – 2022. – Вип. 22. – С. 158-167. – Режим доступу: <https://doi.org/10.15421/322217>.

4. Турчина В. А. Дослідження задачі упорядкування з перериваннями для одного підкласу дерев [Електронний ресурс] / В. А. Турчина, Є. О. Коваленко // Питання прикладної математики і математичного моделювання. – 2023. – Вип. 23. – С. 118-125. – Режим доступу: <https://doi.org/10.15421/322313>.

5. Baker K. R. Principles of Sequencing and Scheduling [Electronic resource] / Kenneth R. Baker, Dan Trietsch. – Hoboken, NJ, USA : John Wiley & Sons, Inc., 2018. – Mode of access: <https://doi.org/10.1002/9781119262602>.

6. Lenstra J. K. On the complexity of scheduling unrelated parallel machines with limited preemptions [Electronic resource] / Jan Karel Lenstra, Nodari Vakhania // Operations Research Letters. – 2023. – Mode of access: <https://doi.org/10.1016/j.orl.2023.02.004>.

7. Петренюк Д. А. Граціозні дерева. Аналіз проблеми та перспективи [Електронний ресурс] / Д. А. Петренюк // Управляющие системы и машины. — 2016. — № 1. — С. 16–25. – Режим доступу: <https://doi.org/10.15407/usim.2016.01.016>.

8. Semeniuta M. F. Combinatorial Configurations in the Definition of Antimagic Labelings of Graphs [Electronic resource] / M. F. Semeniuta // Cybernetics and Systems Analysis. – 2021. – Vol. 57, no. 2. – P. 196–204. – Mode of access: <https://doi.org/10.1007/s10559-021-00344-y>.

REFERENCES

1. Burdyuk V. Ya. Parallel sequencing algorithms: textbook. / V. Ya. Burdyuk, V. A. Turchyna. – Dnipropetrovsk: DSU, 1985. 83 p.

2. Kovalenko Y. O. Analysis of the graph structure influence on the parallel sequencing problems with interruptions solution optimality [Electronic resource] / Y. O. Kovalenko, V. A. Turchyna // Problems of applied mathematics and mathematical modeling. – 2021. – Vol. 21. – pp. 130–137. - Mode of access: <https://doi.org/10.15421/322113>.

3. Turchyna V. A. The influence of the parallel sequencing problem with interruptions initial data on the solution optimality [Electronic resource] / V. A. Turchyna, Y. O. Kovalenko // Problems of applied mathematics and mathematical modeling. – 2022. – Vol. 22. – pp. 158-167. – Mode of access: <https://doi.org/10.15421/322217>.

4. Turchyna V. A. Investigation of the parallel sequencing problem with interruptions for one subclass of trees [Electronic resource] / V. A. Turchyna, Y. O. Kovalenko // Problems of applied mathematics and mathematical modeling. – 2023. – Vol. 23. – pp. 118-125. – Mode of access: <https://doi.org/10.15421/322313>.

5. Baker K. R. Principles of Sequencing and Scheduling [Electronic resource] / Kenneth R. Baker, Dan Trietsch. – Hoboken, NJ, USA : John Wiley & Sons, Inc., 2018. – Mode of access: <https://doi.org/10.1002/9781119262602>.

6. Lenstra J. K. On the complexity of scheduling unrelated parallel machines with limited preemptions [Electronic resource] / Jan Karel Lenstra, Nodari Vakhania // Operations Research Letters. – 2023. – Vol. 51, № 2. – P. 187-189. – Mode of access: <https://doi.org/10.1016/j.orl.2023.02.004>.

7. Petreniuk D. A. Graceful Trees: the State of Arts and the Prospects [Electronic resource] / D. A. Petreniuk // Control systems and computers. — 2016. — Vol. 1, № 2. — P. 16–25. — Mode of access: <https://doi.org/10.15407/usim.2016.01.016>.
8. Semeniuta M. F. Combinatorial Configurations in the Definition of Antimagic Labelings of Graphs [Electronic resource] / M. F. Semeniuta // Cybernetics and Systems Analysis. — 2021. — Vol. 57, № 2. — P. 196–204. — Mode of access: <https://doi.org/10.1007/s10559-021-00344-y>.

Received 12.11.2024.
Accepted 19.11.2024.

***Conditions for length reduction of the parallel sequence of special digraphs' vertices
in the presence of interruptions***

In the scheduling theory, both theoretical and applied problems related to the planning of production processes, the design of computer systems, etc. are considered. In particular, there is a subclass of problems called parallel sequencing problems. Their mathematical formulation can often be reduced to optimization problems on directed graphs, where a given set of vertices corresponds to a set of jobs required for execution, and arcs specify a partial order corresponding to various technological constraints. The classical variant implies that input data are the finite set of jobs with equal execution times (defined as 1) represented as digraph G and one of the sequence parameters (length or width). The undefined parameter needs to be minimized.

On the contrary to the classical ones, in applied problems, the duration of the jobs often differs. That is why the question of whether to allow or forbid interruptions during job execution is widely considered. The efficiency of using interruptions was proven for the cases when the corresponding digraph has a series-parallel, bipartite structure or belongs to the subclass of trees which allows graceful labeling. As graph structure plays a key role in the evaluation of profit from interruptions the graph labeling topics are of a big interest as well, such as those that use combinatorial configurations.

This paper is devoted to the research of possible profit from interruptions in cases when G is a directed caterpillar. The effectiveness of the impact of interruption allowance on the objective function value was analyzed depending on the arcs' orientation, the weight coefficient values of the vertices, and the structure of specific graphs. Theoretical results for evaluating the profit were obtained and formulated in the form of a theorem.

Турчина Валентина Андріївна – к.ф.-м.н., доцент, завідувачка кафедри обчислювальної математики та математичної кібернетики Дніпровського національного університету імені Олеся Гончара.

Коваленко Євген Олександрович – аспірант кафедри обчислювальної математики та математичної кібернетики Дніпровського національного університету імені Олеся Гончара.

Turchyna Valentyna – Candidate of Physical and Mathematical Sciences, Associate Professor, Head of the Department of Computational Mathematics and Mathematical Cybernetics of Oles Honchar Dnipro National University.

Kovalenko Yevhen – Postgraduate student of the Department of Computational Mathematics and Mathematical Cybernetics of Oles Honchar Dnipro National University.

Д.О. Остапець, О.О. Сухомлин

АНАЛІЗ ІСНУЮЧИХ ПІДХОДІВ ДО ФОРМУВАННЯ ФУНКЦІОНАЛЬНИХ ПРОФІЛІВ ЗАХИЩЕНОСТІ

Анотація. У роботі проведено аналіз існуючих підходів до формування функціональних профілів захищеності (ФПЗ) в процесі створення систем захисту інформації. Метою роботи є аналіз підходів до визначення функціональних профілів захищеності при проектуванні систем захисту інформації на основі їх порівняльної характеристики. *Вирішувані задачі:* формування вимог до характеристик підходів та методик визначення ФПЗ, аналіз існуючих підходів та методик визначення ФПЗ та їх порівняльна характеристика. *Сформовано перелік ключових характеристик відомих методик визначення ФПЗ. Проведено порівняльний аналіз методик, визначено їх переваги та обмеження. Надані рекомендації щодо покращення ефективності процесу формування ФПЗ.*

Ключові слова: функціональний профіль захищеності, інформаційна безпека, автоматизована система, інтелектуальні методи, експертна система.

Постановка проблеми. Забезпечення інформаційної безпеки (ІБ) в сучасних умовах цифровізації та активного розвитку автоматизованих систем (АС) є однією з ключових задач будь-якої організації. Особливо важливою складовою цього процесу є формування функціональних профілів захищеності (ФПЗ), які визначають рівень захисту інформації від несанкціонованого доступу (НСД). На сьогодні існує багато різних підходів до розробки ФПЗ, кожен із яких має свої переваги та обмеження. У сучасних дослідженнях відзначається важлива роль кваліфікації експерта при виборі та формуванні ФПЗ, що зумовлює актуальність розробки більш універсальних методик, які мінімізують вплив людського фактора. Ця стаття спрямована на огляд існуючих підходів до формування ФПЗ, а також їх аналіз та порівняння.

Аналіз останніх досліджень і публікацій. В нормативній базі України містяться документи в галузі захисту АС від НСД, що стосуються ФПЗ: НД ТЗІ 2.5-004-99 «Критерії оцінки захищеності інформації в комп'ютерних системах від несанкціонованого доступу» [1], НД ТЗІ 2.5-005-99 «Класифікація автоматизованих систем і стандартні ФПЗ оброблюваної інформації від несанкціонованого доступу» [2] та НД ТЗІ 2.7-010-09 «Методичні вказівки з оцінювання рівня гарантій коректності реалізації ФПЗ в засобах захисту інформації від несанкціонованого доступу» [3].

В результаті аналізу цих документів та наявних публікацій [4], стає зрозуміло наступне:

© Остапець Д.О., Сухомлин О.О., 2024

- існують суперечності у вимогах до функціональних профілів захисту ФПЗ та в політиках ІБ;

- надають нечітку методику щодо вибору відповідних ФПЗ, причому всі прийняті рішення ґрунтуються на експертних оцінках та висновках фахівців з ІБ та відповідних комісій.

Методики, що описані в роботах [5-11], мають велику залежність на професійність експерта, що виконує формування ФПЗ. Таким чином, ефективність кожного методу напряму залежить від кваліфікації та досвіду цього експерта. Отже, покращення експертної оцінки при формуванні ФПЗ є актуальною проблемою.

Мета дослідження. Метою роботи є аналіз підходів до визначення функціональних профілів захищеності при проектуванні систем захисту інформації на основі їх порівняльної характеристики. Результати роботи можуть бути використані при реалізації засобів автоматизації проектування відповідних систем. В роботі відповідно до мети поставлені такі задачі: формування вимог до характеристик підходів та методик визначення ФПЗ, аналіз існуючих підходів та методик визначення ФПЗ та їх порівняльна характеристика.

Основний матеріал дослідження. Для порівняння методик, щодо визначення ФПЗ, пропонується використовувати наступні характеристики: часові витрати, складність методики, вплив кваліфікація експерта, можливість вибору нестандартних ФПЗ:

- **Часові витрати** - показник, що відображає, скільки часу необхідно для побудови ФПЗ відповідно до обраної методики.

- **Складність методики** - відображає рівень складності опанування методики, тобто наскільки легко або важко її реалізувати на практиці.

- **Вплив кваліфікація експерта** - характеризує, якою мірою рішення експерта з ІБ впливають на результативність та правильність вибору ФПЗ.

- **Можливість вибору нестандартних ФПЗ** - визначає, чи дозволяє методика створювати та використовувати ФПЗ, які не є визначеними у нормативних документах як стандартні.

НД ТЗІ 2.5-005-99 [2] описує стандартні ФПЗ, які можуть застосовуватися в державному та банківському секторах, проте без детального визначення методів їх вибору. Таким чином, із змісту нормативних документів випливає, що після аналізу середовища експерт з ІБ повинен обирати ФПЗ зі стандартних, що перекладає відповідальність за остаточне рішення на досвід та компетентність самого експерта та експертної комісії.

Методику вибору зі стандартних ФПЗ експертом з ІБ, можна визначити наступні характеристики:

- **Часові витрати** - значні, експерт з ІБ має провести глибокий аналіз середовища циркуляції інформації, врахувати всі необхідні критерії безпеки та на основі цього вибрати найбільш підходящий стандартний ФПЗ або той, що найбільше відповідає вимогам.

- **Складність методики** - середня, через обмежену кількість стандартних ФПЗ процес вибору є спрощеним, що зменшує кількість можливих варіантів і полегшує процес прийняття рішення.

- **Вплив кваліфікація експерта** - значний, якість виконання методики повністю залежить від компетентності та досвіду експерта, оскільки саме він визначає релевантність і відповідність стандартного ФПЗ до конкретної системи.

- **Можливість вибору нестандартних ФПЗ** - ні, методика не передбачає можливості створення або вибору нестандартних ФПЗ, обмежуючи гнучкість і варіативність рішень стандартними профілями.

Авторами [5] аналізуються вимоги нормативних документів щодо формування ФПЗ інформаційних систем від НСД. Автори вказують на недоліки існуючого підходу до формування таких профілів і пропонують нові методи для їх розробки, а саме автори описують новий метод, який передбачає використання спеціальних таблиць для кожної функціональної послуги захищеності. Ці таблиці дозволяють експерту визначати, які послуги необхідні для конкретної системи, і як вони мають бути реалізовані.

Задля перевірки та верифікації збудованої ФПЗ, автори описують в роботі [6] метод, який дозволяє перевірити коректність та несуперечність збудованої ФПЗ. Верифікація виконується в 3 кроки: перевірка наявності усіх необхідних послуг, перевірка сумісності послуг, викриття помилок типу "неповнота".

Метод побудови ФПЗ опитувальними таблицями та верифікацією можна охарактеризувати наступним чином:

- **Часові витрати** - середні, незважаючи на використання спеціальних таблиць та алгоритмів, які зменшують час на побудову ФПЗ, усе одно експерту необхідно провести аналіз загроз та відповісти на питання щодо критеріїв безпеки та верифікувати результат. Але усе одно це буде швидше ніж перебор усіх можливих варіантів вручну.

- **Складність методики** - середня, через використання опитувальних таблиць та алгоритму верифікації, дозволяє спростити процес побудови ФПЗ.

- **Вплив кваліфікація експерта** - значний, метод найбільш ефективний для досвідчених фахівців. Висококваліфікований експерт, який знайомий з принципами побудови систем захисту інформації та вимогами нормативних документів, зможе використовувати метод ефективно і швидко. Водночас методика дозволяє дещо компенсувати недостатність кваліфікації завдяки інструментам самоперевірки та детальним таблицям питань і відповідей.

- **Можливість вибору нестандартних ФПЗ** - так, побудова ФПЗ "з нуля" здійснюється за допомогою цього методу завдяки гнучкості таблиць, які дозволяють додавати нові функції або змінювати рівні поточних послуг. Однак створення нестандартного профілю вимагатиме більше часу та глибоких знань експерта, оскільки потрібно враховувати всі взаємозалежності між послугами безпеки та верифікувати їх.

У статті [7], автори детально описано підхід до формалізації завдання вибору оптимального ФПЗ для систем захисту інформації (СЗІ). Підхід базується на кількох ключових етапах та використовує математичне моделювання для оцінки загроз і вибору відповідних заходів захисту. Кожна загроза має свою ймовірність появи, яка залежить від зовнішніх дестабілізуючих факторів, що впливають на об'єкт захисту. Експертним шляхом визначається ймовірність появи цих факторів та їх вплив на загрози. В основі

підходу лежить припущення, що загроза нейтралізується відповідними засобами захисту.

Для кожної загрози розраховується ймовірність її нейтралізації через функціональні послуги. Ці послуги групуються в ФПЗ, і завданням експертів є визначення необхідного набору послуг для досягнення оптимального рівня захищеності.

Метод оптимального вибору функціонального профілю захищеності, можна охарактеризувати наступним чином:

- **Часові витрати** - значні, модель передбачає детальну оцінку загроз і залежностей між ними, що може потребувати значного часу, особливо при створенні нестандартних профілів.

- **Складність методики** - значна, через використання математичних моделей і необхідність експертних оцінок методика є складною і вимагає від експертів глибоких знань в області інформаційної безпеки.

- **Кваліфікація експертів** - значна, успіх реалізації методу залежить від досвіду та компетентності експертів, які повинні визначати ймовірності загроз та рівні захисту.

- **Можливість вибору нестандартних ФПЗ** - так, метод дозволяє гнучко адаптувати профілі захисту для нестандартних об'єктів завдяки можливості додавання або зміни функцій та рівнів послуг.

Метод вибору ФПЗ, запропонований у роботі [8], можна назвати як ймовірісно-вартісний метод вибору ФПЗ. Він базується на формалізованій математичній моделі та охоплює кілька ключових етапів, які сприяють оптимальному вибору профілю захисту для інформаційних систем. Основна мета цього підходу – мінімізація збитків від загроз при одночасному контролі витрат на систему захисту.

Цей метод вибору ФПЗ допомагає знайти найкращий набір заходів для захисту інформаційної системи, при цьому враховуючи, скільки це буде коштувати і як добре система захисту може запобігти загрозам

Ймовірісно-вартісний метод вибору ФПЗ можна розбити на наступні ключові етапи:

3. залучення експертів;
4. збір інформації;
5. розрахунок ймовірностей - вираховуються, наскільки ймовірно, що загрози справді можуть трапитися, і наскільки ефективними будуть засоби захисту;
6. оцінка збитків - визначають, скільки збитків може нанести загроза, і скільки вдасться уникнути завдяки захисту;
7. пошук оптимального рішення - яке найкраще захищає від загроз та не перевищує допустиму вартість реалізації.

Цей підхід допомагає створити таку систему захисту, яка не тільки добре захищає інформацію, але й допомагає знайти оптимальне рішення щодо витрачений грошей. У результаті аналізу, можна оцінити наступні характеристики:

- **Часові витрати** - значні, метод є досить трудомістким і потребує значного часу, оскільки включає безліч необхідних етапів.

- **Складність методики** - значна, метод має високу складність, оскільки він базується на математичному моделюванні та розрахунках, що включають ймовірності, оцінку збитків і впровадження оптимального рішення.

- **Кваліфікація експертів** - значна, від рівня знань і досвіду експертів залежить точність оцінки загроз, правильний підбір засобів захисту та визначення їхньої ефективності. Низька кваліфікація може призвести до неточних оцінок і неефективного вибору ФПЗ.

- **Можливість вибору нестандартних ФПЗ** - так, це можливо завдяки його гнучкості методу. Експерти можуть створювати профілі з урахуванням специфічних вимог системи та додавати нові послуги або коригувати рівні існуючих.

У статті [9] автори запропонували свій підхід до визначення стандартних ФПЗ для АС від НСД. Цей підхід базується на нормативних документах технічного захисту інформації (ТЗІ), зокрема на НД ТЗІ 2.5-004-99 та НД ТЗІ 2.5-005-99. Цьому методу можна дати назву як - логіко-матричним методом вибору ФПЗ. З цього методу, можна виділити наступні етапи:

4. визначення залежностей між послугами;
5. розбиття критеріїв інформаційної безпеки на вимоги;
6. побудова логічної рівняння для визначення критеріїв захисту та їх рівнів;
7. використання матриці знань, що містить опис кожного критерію.

Такий підхід дозволяє створити автоматизовану систему для визначення та обґрунтування вибору стандартного ФПЗ. Цей метод можна оцінити наступним чином:

- **Часові витрати** - середні, метод є досить трудомістким з точки зору реалізації, але під час експлуатації дасть процес ас на вибору стандартного ФПЗ.

- **Складність методики** - середня, хоч під час реалізації експертної системи і потребує багато зусиль та уваги на побудову матриці знань та логічного опису критеріїв захисту. Проте, фахівцям у сфері ІБ, які мають досвід у побудові ФПЗ, цей підхід буде зрозумілим.

- **Кваліфікація експертів** - значна, експерта відіграє вирішальну роль, оскільки він відповідає за коректну оцінку вимог по захисту інформації, побудову функціональних профілів та правильну інтерпретацію результатів.

- **Можливість вибору нестандартних ФПЗ** - так, у цій роботі автори досліджували використання методу для вибору функціонального профілю захищеності зі списку стандартних. Однак, метод є досить гнучким, що дозволяє його адаптувати для створення нестандартних ФПЗ, що забезпечує більш ефективний захист інформації.

Метод вибору парето-оптимальних ФПЗ описаний автором у роботі [10] ставе задачу оптимізувати процес вибору ФПЗ. Особливу увагу приділено створенню моделі, яка враховує ресурсні обмеження. Також запропоновано метод пошуку парето-оптимальних проектних альтернатив при побудові ФПЗ.

Можна виділити наступні етапи цього методу:

3. Формування множини припустимих ФПЗ (для кожної АС формується множина можливих варіантів ФПЗ, де кожен варіант має свої показники безпеки та витрат на його реалізацію).

4. Визначення критеріїв оптимальності (вводяться критерії, за якими буде проводитися оцінка ФПЗ).

5. Пошук Парето-оптимальних рішень.

6. Звуження множини до єдиного рішення.

Таким чином, метод дозволяє знайти баланс між рівнем захищеності інформації та ресурсними витратами на впровадження ФПЗ. У результаті аналізу, метод вибору парето-оптимальних ФПЗ, можна оцінити наступним чином:

- **Часові витрати** - значні, метод є часоємним, оскільки включає декілька етапів – від формування множини варіантів ФПЗ до пошуку Парето-оптимальних рішень.

- **Складність методики** - значна, метод вимагає глибокого розуміння в математичній моделі та знаннях у сфері ІБ та побудові ФПЗ.

- **Кваліфікація експертів** - значна, від кваліфікації спеціаліста залежить точність і надійність кінцевого рішення.

- **Можливість вибору нестандартних ФПЗ** - так, хоча основний акцент робиться на стандартні ФПЗ, підхід дозволяє створювати нестандартні ФПЗ.

Адаптивний метод визначення ФПЗ АС на основі оптимізації загроз описано автором у роботі [11]. Метод ґрунтується на математичному моделюванні та теорії ймовірностей. Основна мета методу – підвищити ефективність побудови систем захисту АС, адаптуючи ФПЗ до поточного рівня загроз. В зазначеному методі можна виділити наступні етапи:

1. Аналіз загроз та вибір критеріїв (обираються такі критерії як ризик безпеки, гарантія безпеки, вартість реалізації захисту та інші).

2. Обчислення цільової функції (на основі обраних критеріїв і математичних моделей (включаючи оптимізацію по Беллману) проводяться обчислення).

3. Оцінка витрат та ризиків (оцінюється витрати на впровадження ФПЗ та ймовірність успішного протистояння загрозам).

Адаптивний метод визначення ФПЗ АС на основі оптимізації загроз дозволяє ефективно генерувати ФПЗ та оптимізувати його відповідно до рівня загроз та вимог безпеки. У результаті аналізу, можна оцінити наступні характеристики:

- **Часові витрати** - середні, метод зменшує час на вибір ФПЗ завдяки автоматизації процесу та використанню математичного моделювання, особливо в частині обчислення цільової функції та перевірки збіжності.

- **Складність методики** - значна, оскільки метод передбачає використання оптимізаційних моделей (зокрема, методу Беллмана) і застосування теорії ймовірностей, що вимагає не лише знань у галузі ІБ, але й розуміння математичних методів.

- **Кваліфікація експертів** - значна, метод вимагає значну кваліфікацію експерта, особливо на етапі вибору початкових даних, таких як критерії оцінки загроз та ризиків.

• **Можливість вибору нестандартних ФПЗ** - так, метод передбачає можливість вибору нестандартних ФПЗ, оскільки він підтримує гнучке налаштування критеріїв і адаптацію до поточних загроз.

На основі проведеного аналізу методик визначення ФПЗ, виконано їх порівняльну характеристику, результати якої представлено в табл. 1.

Порівнюючи методики визначення ФПЗ від НСД, можна побачити, що всі з них вимагають значних часових ресурсів та більшість з них потребують необхідних знань та навичок. Окрема увага приділяється кваліфікації експерта, від якої залежить ефективність побудованої ФПЗ. Більшість методів підтримують можливість розробки нестандартних ФПЗ, що свідчить про їх здатність відповідати сучасним вимогам у сфері ІБ. Таким чином, можна запропонувати покращити ефективність процесу визначення ФПЗ шляхом побудови експертних систем із використанням інтелектуальних методів, які можуть також зменшити вплив кваліфікації експерта на результат сформованого ФПЗ.

Таблиця 1

Порівняльна характеристика методик визначення ФПЗ

Методика побудови ФПЗ	Характеристики			
	Часові витрати	Складність методики	Вплив кваліфікації експерта	Побудова нестандартних ФПЗ
Вибір зі стандартних ФПЗ експертом	значні	середня	значний	ні
Метод побудови ФПЗ опитувальними таблицями та верифікацією	середні	середня	значний	так
Метод оптимального вибору ФПЗ	значні	значна	значний	так
Ймовірно-вартісний метод вибору ФПЗ	значні	значна	значний	так
Логіко-матричним методом вибору ФПЗ	середні	середня	значний	так
Метод вибору парето-оптимальних ФПЗ	значні	значна	значний	так
Адаптивний метод визначення ФПЗ АС на основі оптимізації загроз	середні	значна	значний	так

Висновки. У роботі проведено огляд та аналіз існуючих методик до формування ФПЗ автоматизованих систем. Визначено ключові характеристики, що описують методики побудови ФПЗ, зокрема кваліфікацію експерта, складність методик, часові витрати та можливість створення нестандартних профілів. Порівняння існуючих методик дозволило визначити їх переваги та обмеження, що дає можливість для подальшого вдосконалення процесу визначення ФПЗ. Зазначено, що покращити ефективність процесу формування ФПЗ можна шляхом побудови експертних систем із використанням інтелектуальних методів.

ЛІТЕРАТУРА

1. Критерії оцінки захищеності інформації в комп'ютерних системах від несанкціонованого доступу: НД ТЗІ 2.5-004-99. – Режим доступу: <http://www.dstszi.gov.ua/dstszi/doccatalog/document?id=41649>
2. Класифікація автоматизованих систем і стандартні функціональні профілі захищеності оброблюваної інформації від несанкціонованого доступу: НД ТЗІ 2.5-005-99. – Режим доступу: <http://www.dstszi.gov.ua/dstszi/doccatalog/>
3. Методичні вказівки з оцінювання рівня гарантій коректності реалізації функціональних послуг безпеки в засобах захисту інформації від несанкціонованого доступу: НД ТЗІ 2.7-010-09. – Режим доступу: <http://www.dsszzi.gov.ua/dsszzi/doccatalog/document?id=103247>
4. Паламарчук Н.А., Хлапонін Ю.І., Овсянніков В.В. Сучасний стан нормативно-правової бази в галузі технічного захисту інформації // Збірник наукових праць ВІТІ НТУУ «КПІ» – К.: ВІТІ НТУУ «КПІ», 2011. №3. С. 78 – 82. Режим доступу: http://viti.edu.ua/files/zbk/2011/11_3_2011.pdf
5. Леншин А.В., Буслов П.В. Метод формування функціональних профілів захищеності від несанкціонованого доступу // Радіоелектронні і комп'ютерні системи : науч. тр. – Х.: Нац. аерокосм. ун-т “ХАИ”, 2010. – Вып. 7(48). – С. 77–81. – Режим доступу: http://nbuv.gov.ua/UJRN/recs_2010_7_15
6. Потій О.В., Леншин А.В. Методи побудови та верифікації несуперечності і повноти функціональних профілів захищеності від несанкціонованого доступу // Научно-технический журнал “Прикладная радиоэлектроника. Тематический выпуск, посвященный проблемам обеспечения информационной безопасности”. – Х., 2010. – Том 9. – №3. – С.479 – 488. – Режим доступу: <http://openarchive.nure.ua/handle/document/410>
7. Пискун С. Ж., Хорошко В.О. Оптимізація вибору функціонального профілю захищеності // Сучасна спеціальна техніка. - 2011. - № 3. - С. 36-40. - Режим доступу: http://nbuv.gov.ua/UJRN/sstt_2011_3_8
8. Ткач Ю. М. Метод вибору функціонального профілю захищеності // Інформатика та математичні методи в моделюванні.- 2020. Т. 10, № 1-2. С. 68-74.
9. Юдін О.К., Бучик С.С., Мельник С.В. Теоретичні основи визначення стандартних функціональних профілів захищеності автоматизованої системи від несанкціонованого доступу. Наукоємні технології. 2016. Вип. 2. С. 195-205.

10. Берестов Д.С. Методика вибору проектних альтернатив системи захисту інформації в автоматизованій системі відомчого призначення на основі ієрархічних моделей: автореф. дис. канд. техн. наук: 05.13.21. Київ, 2015. - 21 с.

11. Потенко О.С. Методи визначення функціонального профілю захисту автоматизованої системи з урахуванням поточного рівня загроз: автореф. дис. канд. техн. наук : 05.13.21. Київ, 2024. - 24 с.

REFERENCES

1. Criteria for Assessing Information Security in Computer Systems from Unauthorized Access: ND TZI 2.5-004-99. – Available at: <http://www.dstszi.gov.ua/dstszi/doccatalog/document?id=41649>

2. Classification of Automated Systems and Standard Functional Security Profiles of Processed Information from Unauthorized Access: ND TZI 2.5-005-99. – Available at: <http://www.dstszi.gov.ua/dstszi/doccatalog/>

3. Methodical Guidelines for Assessing the Level of Correctness Guarantees in Implementing Functional Security Services in Information Protection Tools from Unauthorized Access: ND TZI 2.7-010-09. – Available at: <http://www.dsszzi.gov.ua/dsszzi/doccatalog/document?id=103247>

4. Palamarchuk N.A., Khlaponin Yu.I., Ovsyannikov V.V. The Current State of the Regulatory Framework in the Field of Technical Information Protection // Collection of Scientific Papers of VITI NTUU "KPI" – K.: VITI NTUU "KPI", 2011. No. 3. P. 78-82. Available at: http://viti.edu.ua/files/zbk/2011/11_3_2011.pdf

5. Lenshin A.V., Buslov P.V. Method of Forming Functional Security Profiles from Unauthorized Access // Radio-electronic and Computer Systems: Scientific Journal – Kharkiv: National Aerospace University “KhAI”, 2010. – Issue 7(48). – P. 77–81. – Available at: http://nbuv.gov.ua/UJRN/recs_2010_7_15

6. Potii O.V., Lenshin A.V. Methods of Building and Verifying the Consistency and Completeness of Functional Security Profiles from Unauthorized Access // Scientific and Technical Journal "Applied Radio Electronics. Thematic Issue Devoted to Information Security Problems". – Kharkiv, 2010. – Vol. 9. – No. 3. – P. 479–488. – Available at: openarchive.nure.ua/handle/document/410.

7. Piskun S.G., Khoroshko V.O. Optimization of Functional Security Profile Selection // Modern Special Equipment. - 2011. - No. 3. - P. 36-40. - Available at: http://nbuv.gov.ua/UJRN/ssst_2011_3_8

8. Tkach U. M. Method of Selecting a Functional Security Profile // Informatics and Mathematical Methods in Modeling. – 2020. – Vol. 10, No. 1-2. – P. 68-74.

9. Yudin O.K., Buchyk S.S., Melnyk S.V. Theoretical Foundations for Defining Standard Functional Security Profiles of Automated Systems from Unauthorized Access. High-Tech Technologies. 2016. Issue 2. P. 195-205.

10. Berestov D.S. Methodology for Selecting Design Alternatives for Information Protection Systems in Departmental Automated Systems Based on Hierarchical Models: PhD Thesis Abstract in Technical Sciences: 05.13.21. Kyiv, 2015. - 21 p.

11. Potenko O.S. Methods for Defining the Functional Security Profile of an Automated System Considering the Current Threat Level: PhD Thesis Abstract in Technical Sciences: 05.13.21. Kyiv, 2024. - 24 p.

Received 12.11.2024.
Accepted 19.11.2024.

Analysis of existing approaches to the formation of functional security profiles

The work examines existing methods for forming functional security profiles (FSP) for information protection systems. The purpose of the work is to analyze approaches to determining FSP when designing information protection systems based on their comparative characteristics. Solved tasks: formation of requirements for the characteristics of approaches and methods for determining the FSP, analysis of existing approaches and methods for determining the FSP and their comparative characteristics. The authors explore the key characteristics of various methodologies, such as time consumption, complexity, the influence of expert qualification, and the possibility of creating non-standard profiles. The work compares the advantages and limitations of existing methods. The research shows that all methodologies require significant time and are dependent on the qualifications of specialists. At the same time, most methods support the development of non-standard profiles, allowing them to be adapted to specific information security requirements. Based on the analysis, it is concluded that the use of automated systems based on intelligent methods is necessary to improve the efficiency and accuracy of FSP formation. Thus, the work makes a significant contribution to the field of information security research, providing practical recommendations for improving methodologies for forming FSP in the context of increasing threats in the information technology sphere.

Остапеч Денис Олександрович – к.т.н., доцент, доцент кафедри «Електронні обчислювальні машини» Українського державного університету науки і технологій, ORCID: 0000-0003-1778-7770

Сухомлин Олексій Олександрович – аспірант кафедри «Електронні обчислювальні машини» Українського державного університету науки і технологій, ORCID: 0009-0006-7928-4721

Ostapets Denys – candidate of technical sciences, associate professor, associate professor of «Electronic computers» department of Ukrainian State University of Science and Technologies, ORCID: 0000-0003-1778-7770

Sukhomlyn Oleksii – graduate student of «Electronic computers» department of Ukrainian State University of Science and Technologies, ORCID: 0009-0006-7928-4721

ДОСЛІДЖЕННЯ ОПЕРАЦІЇ ДІЛЕННЯ НА ДВА В СИСТЕМІ ЗАЛИШКОВИХ КЛАСІВ З УСІМА ПАРНИМИ МОДУЛЯМИ

Анотація. Розвиток сучасної техніки, інформаційних та управляючих систем потребує застосування нових принципів, зокрема, на підставі представлення чисел в системі залишкових класів. Традиційна система класів залишків — це система, у якій довільне число представляється у вигляді набору найменших невід'ємних залишків по модулях. При цьому, якщо модулі попарно взаємно прості, то такому представленню відповідає тільки одне число в інтервалі чисел. У той самий час реалізація нових тенденцій у системі залишкових класів вимагає поруч із традиційними системами взаємно простих модулів використання систем із взаємно непростими, зокрема, з усіма парними модулями. При цьому система всіх парних модулів, кожен з яких не є співмножником жодного з інших модулів цієї системи, побудована на основі системи взаємно простих модулів базису шляхом множення кожного модуля базису на парне число - коефіцієнт переходу. Однією зі складних операцій у такій системі є ділення числа на два. Запропонований підхід до розв'язання задачі є наступним. Виконується ділення на два залишків за модулями системи. Складається модульне рівняння, за результатами якого визначаються два значення залишку для кожного модуля, розташованих у різних інтервалах чисел і маючих протилежні парності. Оскільки в парній системі модулів всі залишки або парні, або непарні, формуємо набір з усіх парних залишків і набір з усіх непарних залишків. При діленні на два числа переводяться в нижню половину інтервалу чисел, тому виконуємо порівняння цих наборів на менший з них. Запропонований підхід забезпечує бажаний результат, і є доцільним застосувати його як перспективний напрямок дослідження складних операцій у системі залишкових класів з усіма парними модулями.

Ключові слова: залишкові класи; системи модулів; інтервал чисел, ділення числа на два.

Постановка проблеми

Розвиток сучасної техніки, інформаційних та управляючих систем потребує застосування нових принципів, зокрема, на підставі представлення чисел у системі залишкових класів. Відповідно до класичної роботи [1] системою залишкових класів називається система числення, у якій довільне число N представляється як набір найменших невід'ємних залишків по модулях m_1, m_2, \dots, m_n , тобто $N = (\alpha_1, \alpha_2, \dots, \alpha_n)$. Тут $\alpha_i = N \pmod{m_i}$. При цьому, якщо числа m_i попарно взаємно прості, то такому представ-

ленню відповідає тільки одне число N інтервалу $[0, M-1]$, де $M = m_1 m_2 \dots m_n$. До переваг такого представлення чисел належить незалежність розрядів числа один від одного і можливість їхньої паралельної обробки, мала розрядність залишків, висока точність, здатність системи до самокорекції. Недоліки обумовлені труднощами при реалізації немодульних операцій. У багатьох обчислювальних завданнях, що виконуються в системі залишкових класів, однією з трудомістких немодульних операцій є операція ділення числа на два. Разом з тим реалізація нових напрямів в системі залишкових класів вимагає поряд з використанням систем взаємно простих модулів застосування також і систем з взаємно непростими, зокрема, з усіма парними модулями. При цьому кожен із них не є співмножником жодного з інших модулів цієї системи. Деякі особливості таких систем розглянуті у роботах [2—6].

Аналіз останніх досліджень і публікацій

Системі залишкових класів із усіма парними модулями присвячені роботи [2—6]. Розглянуті проблеми визначення належності числа до даної половини інтервалу чисел та спосіб представлення чисел такої системи у поліадичному коді, досліджені позиційні характеристики при представленні чисел у прямому та зворотному кодах, наведений алгоритм перетворення псевдочисел системи залишкових класів з усіма парними модулями у числа системи. Роботи щодо реалізації операції ділення числа на два в такій системі модулів автором не виявлено.

Формулювання мети дослідження

Метою дослідження є аналітичний розгляд системи залишкових класів при всіх парних модулях для реалізації складної операції ділення числа на два.

Виклад основного матеріалу

Нехай у системі парних модулів p_1, p_2 задано число $V = (\beta_1, \beta_2)$, яке необхідно розділити на два. Для цього кожен із залишків ділиться на два, тобто $\gamma_1 = \frac{\beta_1}{2}, \gamma_2 = \frac{\beta_2}{2}$. Значення залишку δ_1 за модулем p_1 $\delta_1 = \gamma_1 + x_1$, значення залишку δ_2 за модулем p_2 $\delta_2 = \gamma_2 + x_2$. Запишемо модульне рівняння для залишків по модулю p_1

$$(2 * (\gamma_1 + x_1)) \pmod{p_1} = 2 * \gamma_1,$$

$$2 * x_1 = 0,$$

$$x_1^1 = 0, \quad x_1^2 = \frac{p_1}{2}.$$

Отже, $\delta_1^1 = \gamma_1$, $\delta_1^2 = \gamma_1 + \frac{p_1}{2}$, належать вони різним половинам інтервалу $[0, p_1 - 1]$ тому один з них є парним числом, другий – непарним.

Аналогічно, модульне рівняння для залишків по модулю p_2

$$(2 * (\gamma_2 + x_2)) \pmod{p_2} = 2 * \gamma_2,$$

$$2 * x_2 = 0,$$

$$x_2^1 = 0, \quad x_2^2 = \frac{p_2}{2}.$$

Отже, $\delta_2^1 = \gamma_2$, $\delta_2^2 = \gamma_2 + \frac{p_2}{2}$, належать вони різним половинам інтервалу $[0, p_2 - 1]$ тому один з них є парним числом, другий – непарним.

Таким чином, отримуємо поєднання залишків δ_1^i та δ_2^i , які представлені табл.1.

Таблиця 1

Рядок	Залишок	
1	δ_1^i парний	δ_2^i парний
2	δ_1^i парний	δ_2^i непарний
3	δ_1^i непарний	δ_2^i парний
4	δ_1^i непарний	δ_2^i непарний

Оскільки в системі парних модулів всі залишки мають бути або парними, або непарними, другий та третій рядки виключаємо з табл.1.

Ділення числа на два переводить його до нижньої половини інтервалу. Тому результат порівняння чисел з залишками першого та четвертого рядків на менше з них дає шуканий результат.

Висновки

Запропонований підхід до розв'язання задачі ділення числа на два в системі залишкових класів при всіх парних модулях. Згідно з цим підходом виконується ділення на два залишків за модулями системи. Складається модульне рівняння, за результатами якого для кожного модуля визначаються два значення залишку, які розташовані у різних інтервалах чисел і мають протилежні парності. Оскільки в парній системі модулів всі залишки або парні, або непарні, формуємо набір з усіх парних залишків і набір з усіх непарних залишків. При діленні на два числа переводяться в нижню половину інтервалу чисел, тому виконуємо порівняння цих наборів на менший з них. Запропонований підхід забезпечує бажаний результат, і є доцільним застосувати його як перспективний напрямок дослідження складних операцій у системі залишкових класів з усіма парними модулями.

ЛІТЕРАТУРА

1. Акушкин І.Я., Юдицький Д.І. Машинна арифметика у залишкових класах. М.: Радянське радіо. 1968. 440 с.
2. Поліський Ю.Д. Про систему залишкових класів із взаємно непростими модулями. *Проблеми математичного моделювання*: матеріали наук.-метод. конф., м. Кам'янське, 24-26 трав. 2017 р. Дніпропетровськ, 2017 С. 107–112.
3. Поліський Ю. Д. Перетворення псевдочисел системи залишкових класів з усіма парними модулями у числа системи. /Ю. Д. Поліський // *Електронне моделювання*. – 2018. – Т. 40. – №1. – С. 115-120.

4. Поліський Ю.Д. Реалізація деяких проблемних операцій у системах залишкових класів. *Математичне моделювання*. 2018. № 1 (38). С. 22–27.
5. Поліський Ю.Д. Про позиційну характеристику у системі залишкових класів. *Ways of Science Development in Modern Crisis Conditions: Proceedings of the 4th International Scientific and Practical Internet Conference*, June 8-9, 2023, Dnipro, Ukraine, P. 351–352.
6. Поліський Ю.Д. Дослідження операції визначення належності числа до даної половини інтервалу чисел в системі залишкових класів з усіма парними модулями. *Математичне моделювання*. - 2023.- № 2(49). – С.9-15.

REFERENCES

1. Akushsky I.Ya., Yudytsky D.I. (1968). Mashynna aryfmetryka u zalyshkovykh klasakh [Machine arithmetic in residual classes]. *Radyanske radio*
2. Polissky Yu.D. (2017). Pro systemu zalyshkovykh klasiv iz vzayemno neprostymy modulyamy [On the system of residual classes with mutually nonprime modules]. *Problemy matematychnoho modelyuvannya – Problems of mathematical modeling*, 107-112 [in Ukrainian].
3. Polissky Yu. D. (2018). Peretvorennya psevdochysel systemy zalyshkovykh klasiv z usima parnymy modulyamy u chysla systemy- Elektronne modelyuvannya]. *Електронне моделювання*. 115-120..
4. Polissky Yu.D. (2018). Realizatsiya deyakykh problemnykh operatsiy u systemakh zalyshkovykh klasiv [Implementation of some problematic operations in residual class systems]. *Mathematychno modelyuvannya – Mathematical modeling*, 22-27 [in Ukrainian].
5. Polissky Yu.D. (2023). Pro pozytsiynu kharakterystyku u systemi zalyshkovykh klasiv [About positional characteristics in the system of residual classes] *Ways of Science Development in Modern Crisis Conditions: Proceedings of the 4th International Scientific and Practical Internet Conference*, 351-352 [in Ukrainian].
6. Polissky Yu.D. (2023). . Doslidzhennya operatsiyi vyznachennya nalezhnosti chysla do danoyi polovyny intervalu chysel v systemi zalyshkovykh klasiv z usima parnymy modulyamy. [Study of the operation of determining whether a number belongs to a given half of the interval of numbers in the system of residual classes with all even modules]. *Mathematychno modelyuvannya - 2023.- № 2(49). – .9-15. [in Ukrainian].*

Received 27.11.2024.
Accepted 30.11.2024.

Study of the division operation by two in the remainder class system with all paired modules

The development of modern technology, information and control systems will require the establishment of new principles, focusing on the representation of numbers in the system of redundant classes. The traditional system of residue classes is a system in which an arbitrary number is represented as a set of smallest non-negative remainders modulo. Moreover, if the modules are pairwise coprime, then only one number in the range of numbers corresponds to this representation. At the same time, the implementation of new trends in the system of residual classes requires, along with the use of systems of co-prime modules, the use of systems with co-prime modules, in particular, with all even modules. Moreover, the system of all even modules, each of which is not a factor of any of the other

modules of this system, is built on the basis of a system of relatively simple modules - the basis - by multiplying each basis module by an even number - the transition coefficient. One of the complex operations in such a system is dividing a number by two. The proposed approach to solving the problem is as follows. The remainder is divided into two by modules of the system. A modular equation is compiled, the results of which determine two remainder values for each module, located in different number intervals and having opposite parities. Since in an even system of modules all remainders are either even or odd, we form a set of all even remainders and a set of all odd remainders. Since, when divided by two, numbers are transferred to the lower half of the range of numbers, we compare these sets to the smaller of them. The proposed approach provides the desired result, and it seems appropriate to apply it as a promising direction for studying complex operations in a system of residual classes with all even modules.

Поліський Юрій Давидович - кандидат технічних наук, старший науковий співробітник, Заслужений винахідник України, Науково-дослідний інститут автоматизації чорної металургії (НДІАчормет) м.Дніпро.

Polissky Yuriy Davydovych - Candidate of technical science, senior researcher, Honored Inventor of Ukraine, Research Institute of Automation of Ferrous Metallurgy (NDIAchormet), Dnipro.

В.С. Михайленко, Г.В. Коренкова, В.В. Лещенко, Л.Я.Мартинovich

МЕТОДИКА СИНТЕЗА НЕЙРО - НЕЧІТКИХ СИСТЕМ АДАПТИВНОГО УПРАВЛІННЯ З ФУНКЦІЄЮ ПАРАМЕТРИЧНОЇ ІДЕНТИФІКАЦІЇ

Анотація. У статті розглянуто методика розробки нейрон - нечітких систем автоматичного управління (САУ) для судових паротурбінних установок (СПТУ) з функцією ідентифікації параметрів в процесі їх експлуатації. Методика включає етапи моделювання динаміки СПТУ, розробку алгоритмів ідентифікації параметрів на основі нейронних мереж, а також їх інтеграцію з нечіткою логікою для прийняття рішень. Проведено аналіз можливостей запропонованого підходу щодо підвищення якості перехідних процесів та стабільності роботи судових енергетичних установок. Отримані результати демонструють, що така система здатна в режимі реального часу коригувати параметри моделі, що забезпечують точність управління і зменшує ризик виникнення аварійних ситуацій. Методика може бути впроваджена у реальних судових енергетичних системах, що потребують адаптивного управління складними процесами.

Ключові слова: нейро- нечітка система, параметрична ідентифікація, судова паротурбінна установка, нейронні мережі, методика, адаптивне управління, судові енергетичні установки.

Постановка проблеми. Багато складних судових об'єктів, зокрема парогенеруючі системи, характеризуються нелінійними властивостями, динамічними елементами, неконтрольованими шумами, перехресними зв'язками, що ускладнює створення оптимальних систем автоматичного управління (САУ). Паротурбінні установки (СПТУ), що є ключовими енергетичними агрегатами вантажних суден, мають складну динаміку, нелінійності та зазнають впливу збурень і змін параметрів у процесі експлуатації. Для забезпечення стабільної роботи судових енергетичних установок (СЕУ) необхідні сучасні системи автоматичного управління, здатні до адаптації у реальному часі [1-3].

Розробка математичних моделей об'єктів управління базується на двох підходах: експериментальному (моделі типу "чорна скринька", що будуються на основі експериментальних даних) та аналітичному (диференціальні рівняння, які описують фізичні процеси). Обидва підходи мають свої переваги, але страждають від обмежень, зокрема труднощів моделювання під час зміни режимів роботи або умов експлуатації.

Одним з підходів до розв'язання цієї проблеми є використання нейро - нечітких систем, які поєднують переваги нейронних мереж і нечіткої логіки [4-8]. Нейро - нечіткі системи здатні до ефективного управління складними об'єктами з невизначено-

стями та нелінійностями, а також можуть бути адаптовані до змін параметрів у процесі експлуатації. Це особливо важливо для суднових папротурбинних установок (СПТУ), де зміни параметрів роботи можуть значно впливати на ефективність та безпеку експлуатації.

Мета і завдання досліджень. Метою роботи є розробка методики побудови нейро - нечітких систем параметричної ідентифікації для адаптивного управління параметрами СПТУ та її перевірка за допомогою імітаційного моделювання.

Матеріали, методи та методи досліджень. Як об'єкт дослідження розглянемо структуру САУ тиском пари суднового парового котла (рис.1). Суднові парові агрегати працюють на різних режимах парового навантаження споживають багато пару, що впливає на процес стабілізації тиску. На схемі регулятор тиску (РС) отримує інформацію від датчика тиску (РЕ) і у випадку відхилення заданого значення від встановленого, виробляє команду на клапан, який регулює витрата палива на пальник парового котла.

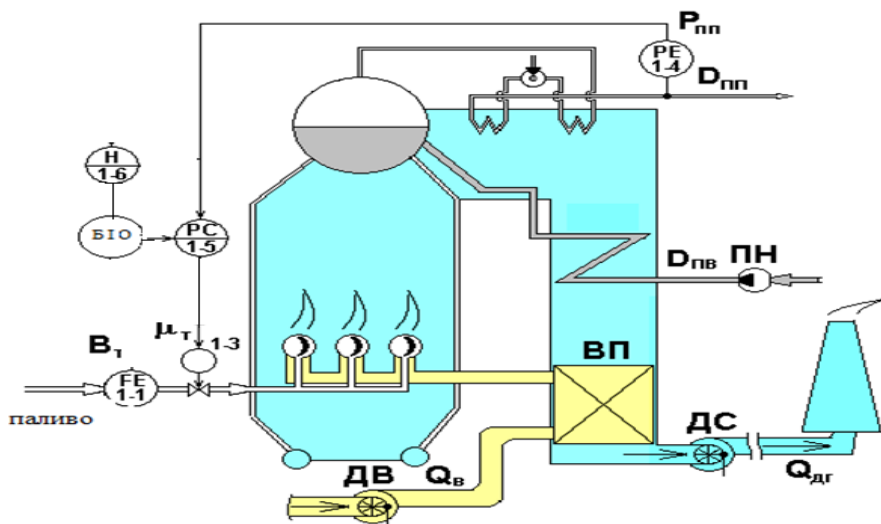


Рисунок 1 - САУ тиском пари: FE – датчик витрати; PC – контролер тиску пари; БІО – блок ідентифікації параметрів об'єкту; PE – датчик тиску; Н – пульт управління; ДВ – дугтевий вентилятор; ДС – димосос; ПН – живильний насос; ВП – повітряний підігрівач; $P_{пп}$ – тиск пари; B_m – витрата палива; $D_{пп}$ – витрата пари; $D_{пв}$ – витрата живильної води; Q_v – витрата повітря; $Q_{дг}$ – витрата димових газів; μ_T – управляючий вплив

На підставі аналізу ключових процесів, можна виділити два основних технологічних процеси: в судновому паровому котлі і паровій турбіні. Для процесу в котлі головними є канали передачі впливів, що зв'язують витрату палива, живильної води і повітря на вході з тиском і температурою перегріву пара на виході і в проміжній точці пароводяного тракту. Для процесу в турбіні головним служить вплив по каналу регулюючі клапани $h_{кл}$ – потужність N , для судновою паротурбіною установки (СПТУ) в цілому - по каналу теплове навантаження (паливо, вода і повітря) - потужність.

Динамічний режим судового парового котла (СПК) характеризується залежностями зміни тиску пара в часі від зміни $Q_{опс}$, $Q_{опв}$, $Q_{опл}$. Тиск пари залишається постійним

за умови дотримання теплового і матеріального балансів. Небаланс в підводі і відводі теплоти при перехідному режимі за час dt компенсується за рахунок зміни кількості теплоти $Q_{ак}$, акумульованої в пароводяної суміші, воді, парі, а також в металевих частинах СПУ. Таким чином, рівняння динаміки має наступний вигляд:

$$(Q_{nc} + Q_{ne} - Q_n)dt = dQ_{ак}. \quad (1)$$

З урахуванням важливості підтримки постійного тиску пара при різних теплових режимах СПК проводиться аналіз рівняння динаміки. Закон зміни тиску в парогенеруючій поверхні СПК при нестационарному (динамічному режимі) визначається за допомогою рішення рівняння виражають матеріальний і енергетично баланс, і рівняння стану. Рівняння матеріального балансу має вигляд [9,10]:

$$D_{ne} - D_n = \frac{d}{dt}(V_e \rho_e - V_n \rho_n), \quad (2)$$

де $\rho_{пв}$ – щільність води, $\rho_{п}$ – щільність пара ($\text{кг} / \text{м}^3$), V_v – водяний об'єм, $V_{п}$ – паровий обсяг (м^3).

Рівняння енергії встановлює, що різниця між припливом і стоком тепла йде на зміну теплової енергії, укладеної в обсягах пари, води і металу [11]:

$$D_{ne} i_{ne} - D_n i_n + Q = \frac{d}{dt}(V_e \rho_e i_{ne} + V_n \rho_n i_n + G_m C_m t_e), \quad (3)$$

де $i_{пв}$ – ентальпія живильної води; $i_{п}$ – ентальпія пара (кДж/кг); Q – кількість теплоти сприйманого поверхнею нагріву СПУ (кДж); G_m – маса металу поверхні, на якій випаровується пар (кг); C_m – питома теплоємність металу (кДж/кг К); t_v – температура киплячої води ($^{\circ}\text{C}$).

Щільність сухої насиченої пари і киплячої живильної води залежить від тиску пари, тому рівняння стану представлено як ряд співвідношень виду:

$$\begin{aligned} \rho_{ne} &= \rho_{ne}(P); \\ \rho_n &= \rho_n(P); \\ i_{ne} &= i_{ne}(P); \\ i_n &= i_n(P); \\ t_e &= t_e(P). \end{aligned} \quad (4)$$

Представлені вище рівняння є нелінійними. Після їх лінеаризації і припущень про рівність температури металу і киплячої води, отримано наступне рівняння динаміки СПК по тиску пари [11]:

$$T \frac{d\sigma}{dt} + K_c \sigma = \mu_m - \varphi, \quad (5)$$

де T – час розгону СПУ; K_c – коефіцієнт самовирівнювання СПК по тиску пари; $\sigma = \Delta P/P_n$; $\mu_m = \Delta Q_m/Q_{m,м}$; $\varphi = \Delta Q_{п}/Q_{п,м}$ – відносні зміни відповідно тиску пара, теплоти палива, що спалюється, теплоти відведеної з СПК з парою (тут $Q_{m,м}$, $Q_{п,м}$ – прийняті базові значення змінних при максимальному навантаженні СПК).

Відомо, що об'єкт по всіх розглянутих каналах має властивість самовирівнювання і інерцію, яка характеризується часом запізнювання і постійною часу T . В якості конкретних прикладів моделей динаміки наведено передавальні функції окремих процесів парових агрегатів СЕУ потужністю 40 МВт при постійних витратах палива та води, отримані в результаті тривалого розрахунку коефіцієнтів диференціальних рівнянь. Після перетворення Лапласа отримано значення передавальні функції ланки запізнювання та інерційного ланки першого порядку [10]:

Для котла (канал паливо B_T – температура пару за стельовим екраном у проміжній точці тракту)

$$W(s) = \frac{k_1 e^{-20s}}{120s + 1} \quad (6)$$

Для турбини (канал тиск – потужність)

$$W(s) = \frac{k_1 \cdot 150s e^{-2.5s}}{150s + 1} \quad (7)$$

де $K_1 = 25$ МВт см/кгс;

для енергоблоку в цілому (канал теплове навантаження - одночасно і узгоджене зміна витрат палива, води і повітря - потужність)

$$W(s) = \frac{k_2 \cdot e^{-17s}}{110s + 1} \quad (8)$$

де $K_2 = 45$ МВт год/т.

Передавальні функції для опису динаміки головного парогенератора СЕУ визначаються при фіксованому навантаженні (наприклад, 4 МВт). Для точного моделювання у широкому діапазоні навантажень потрібні функції для кожного усталеного режиму. Зазвичай ідентифікація здійснюється через ступеневі або синусоїдальні активні впливи, що погіршує якість управління. На діючих об'єктах СЕУ експериментальну інформацію, необхідну для параметричної ідентифікації, заносять в таблиці, що враховують мінливість параметрів моделей динаміки навантажень. На підставі цих значень виконується таблична адаптація типових регуляторів. Разом з тим, на думку авторів, даний спосіб адаптивного управління має цілий ряд недоліків, пов'язаних з неможливістю отримання оптимальних параметрів регулятора для всіх режимів роботи об'єкта, а також неможливістю врахування впливу випадкових перешкод, збурень і нелінійностей значень навантаження.

Виходячи з експериментальних характеристик (табл. 1), запропоновано нейронечітку мережу (ННМ), що враховує думку експертів по налаштуванню САУ, здатна до апроксимації для всіх можливих режимів навантаження парових агрегатів, а також до навчання за алгоритмом зворотнього поширення помилки [4, 8].

Виходячи з експериментальних даних (табл. 1) модель динаміки (передавальні функції по каналу витрата палива - тиск перегрітої пари що надходить на СПТУ) може бути доповнена таблицею значень K , T , τ у широкому діапазоні зміни навантажень СПТУ.

Таблиця параметрів математичних моделей які були отримані за допомогою аналітичних розрахунків

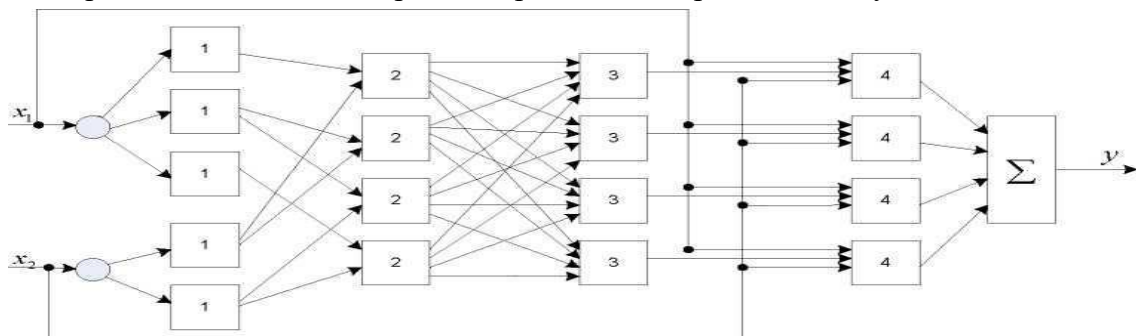
N , МВт	K , $^{\circ}\text{C}/\%$	τ , с	T , с
10	1,8	24	130
20	3,78	34,6	200
40	5.8	46,1	275

Для зменшення часу розрахунків диференціальних рівнянь та через неможливість проведення експериментів у складних умовах навігації запропоновано використовувати нейро-нечіткі мережі для інтерполяції й екстраполяції передавальних функцій на різних теплових режимах. Отримані дані слугуватимуть основою для табличної адаптації САУ.

Розглянемо принцип дії нечіткої нейронної продукційної мережі ANFIS, а саме варіант її реалізації із застосуванням алгоритму нечіткого виведення Сугено-Такагі, де використовується набір правил [5, 6].

На рисунку 2 зображено мережу ANFIS з двома вхідними лінгвістичними змінними x_1, x_2 , та чотирма нечіткими правилами.

Мережа ANFIS є п'ятишаровою, прямого поширення сигналу.



Рисинок 2 - Нечітка нейронна мережа ANFIS з застосуванням алгоритму Сугено-Такагі нульового порядку: 1-функції Гаусса; 2-антецеденти; 3-нормалізація сигналів; 4-укладення правил

Шар 1 - терми вхідних змінних. Входи мережі з'єднані тільки зі своїми термами. Кількість вузлів першого шару дорівнює сумі потужностей терм-множин вхідних змінних, де операція фазифікації виконана на Сингтон базі.

Шар 2 - посилення (антецеденти) нечітких правил. Кількість вузлів другого шару m . Вузол другого шару з'єднаний з тими вузлами першого шару, які формує посилення відповідного правила. Отже, кожен вузол другого шару може приймати від 1 до n сигналів. Виходом вузла є ступінь виконання правила, яка розраховується як добуток вхідних сигналів (по Ларсену). Позначимо виходи вузлів цього шару τ_r , $r = 1, \dots, \overline{m}$, где \overline{m} - кількість нечітких правил.

Шар 3 - нормалізація ступенів виконання правил. Кількість вузлів третього шару є m . Кожен вузол цього шару розраховує відносний ступінь виконання нечіткого правила (нормалізація) за формулою:

$$\tau_r^* = \frac{\tau_r}{\sum_{j=1}^m \tau_j}. \quad (9)$$

Шар 4 - складення правил. Кількість вузлів шару також m . Кожен вузол з'єднаний з одним з вузлів третього шару, а також з усіма входами мережі. Вузол четвертого шару розраховує внесок одного нечіткого правила у вихід мережі за формулою

$$y_r = \tau_r^* (b_{0,r} + b_{1,r}x_1 + \dots + b_{n,r}x_n). \quad (10)$$

Шар 5 - агрегування (композиція) результату, отриманого за різними правилами. Єдиний вузол цього шару підсумовує внески всіх правил.

$$y = \sum_{j=1}^m y_j. \quad (11)$$

Налаштування мережі ANFIS із двома входними змінними й чотирма нечіткими правилами виконується комбінацією градієнтного спуску (ЗПП) і методу найменших квадратів (МНК). Алгоритм ЗПП налаштовує функції приналежності, а МНК оцінює коефіцієнти укладення правил.

На першому етапі, запропонованої авторами методики, на входи подається навчальна вибірка і по нев'язці між бажаною і дійсною поведінкою мережі МНК знаходяться оптимальні параметри вузлів четвертого шару.

На другому етапі залишкова нев'язка коригує параметри вузлів першого шару, не змінюючи знайдені раніше коефіцієнти. Ітераційна процедура налаштування триває, поки нев'язка перевищує заздалегідь встановлене значення. Для налаштування функцій приналежності фазифікатора, крім методу ЗПП, можуть використовуватися і інші алгоритми оптимізації [9]. Гібридна мережа, яка реалізує алгоритм Сугено створюється виходячи з ряду етапів методики проектування показаних на рисунках 3 – 4.

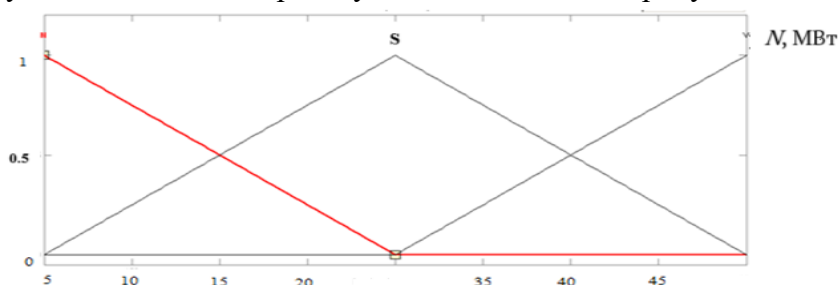


Рисунок 3 - Фазифікація вхідного параметра «навантаження СПТУ»

1.Розробка бази правил виду:

- ЯКЩО «навантаження велике» ТО «параметри моделі об'єкта $K_{об}$, $T_{об}$, τ - приймають великі значення» ІНАКШЕ,
- ЯКЩО «навантаження середнє» ТО «параметри моделі об'єкта $K_{об}$, $T_{об}$, τ - приймають середні значення» ІНАКШЕ,
- ЯКЩО «навантаження високе» ТО «параметри моделі об'єкта $K_{об}$, $T_{об}$, τ - приймають високе значення» і т.д.

2. Розробка структури гібридної мережі (ГМ) і її навчання на основі тестової вибірки (див. табл. 1). Структура ГМ виду ANFIS зображено на рисунку 4. Вибрано - кількість циклів навчання - 30, метод навчання - зворотного поширення помилки.

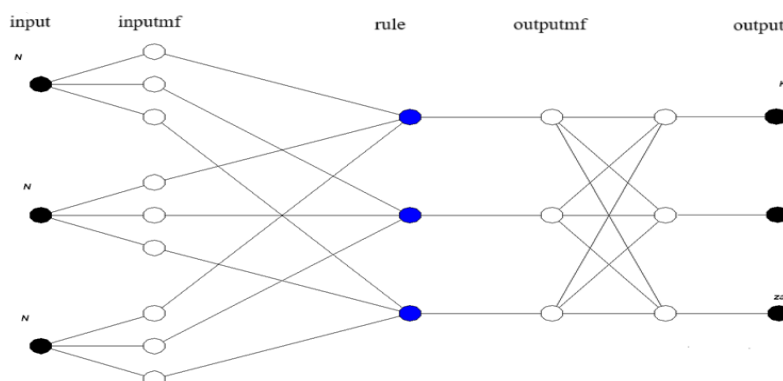


Рисунок 4 - Структура розробленої ННМ

Проміжні значення параметрів моделі об'єкта дослідження можливого навантаження СПТУ отримані в результаті апроксимації ГМ наведені в таблиці 2.

Таблиця 2

Таблиця параметрів математичної моделі для динамічного режиму роботи СПТУ

N, МВт	K, °C/%	τ , с	T, с
10	1,8	24	130
15	2,79	29	163
20	3,78	34,6	200
30	3,92	35,4	205
35	4,91	41	242
40	5.8	46,1	275

Для подальшого імітаційного моделювання інтелектуального блоку параметричної ідентифікації розроблено програму в додатку MatLab (Simulink) (рис.5). Розроблена ГМ була експортована до блоку нечіткого логічного контролера.

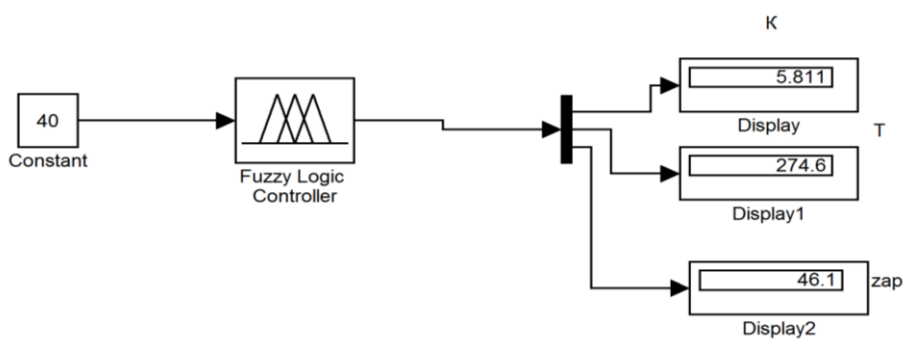


Рисунок 5 - Імітаційна модель перевірки роботи системи параметричної ідентифікації

Наступним етапом запропонованої методики є розробка програми для табличної адаптації (розрахунку оптимальних налаштувань ПІ-регулятора) САУ парових агрегатів.

Адаптація параметрів регулятора (табличне автоналаштування, рис. 6) передбачає використання заздалегідь визначених параметрів для різних умов роботи системи, що зберігаються у таблиці та застосовуються при необхідності.

Цей підхід доповнено блоком параметричної ідентифікації з нейро-нечіткою мережею. Поточне парове навантаження перетворюється навченою мережею у математичні моделі (передавальні функції). Ці функції надходять до блоку табличної адаптації, де формули розраховують оптимальні налаштування регуляторів (ПІ, ПІД).

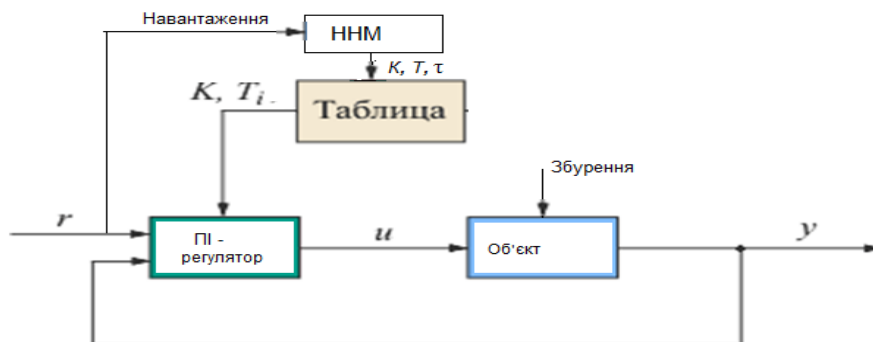


Рисунок 6 – Структурна схема запропонованої нейро – нечіткої адаптивної САУ

Як метод адаптивної оптимізації параметрів ПІ – регулятора автори обрали табличний метод А. Капеловича для САУ тепловими теплоенергетичними об'єктами.

Формула Капеловича використовується для налаштування ПІ-регулятора на основі параметрів об'єкта керування: його коефіцієнта передачі, сталої часу та часу запізнення. Для ПІ-регулятора ці параметри розраховуються наступним чином:

Нехай:

K — коефіцієнт передачі об'єкта.

T — стала часу об'єкта.

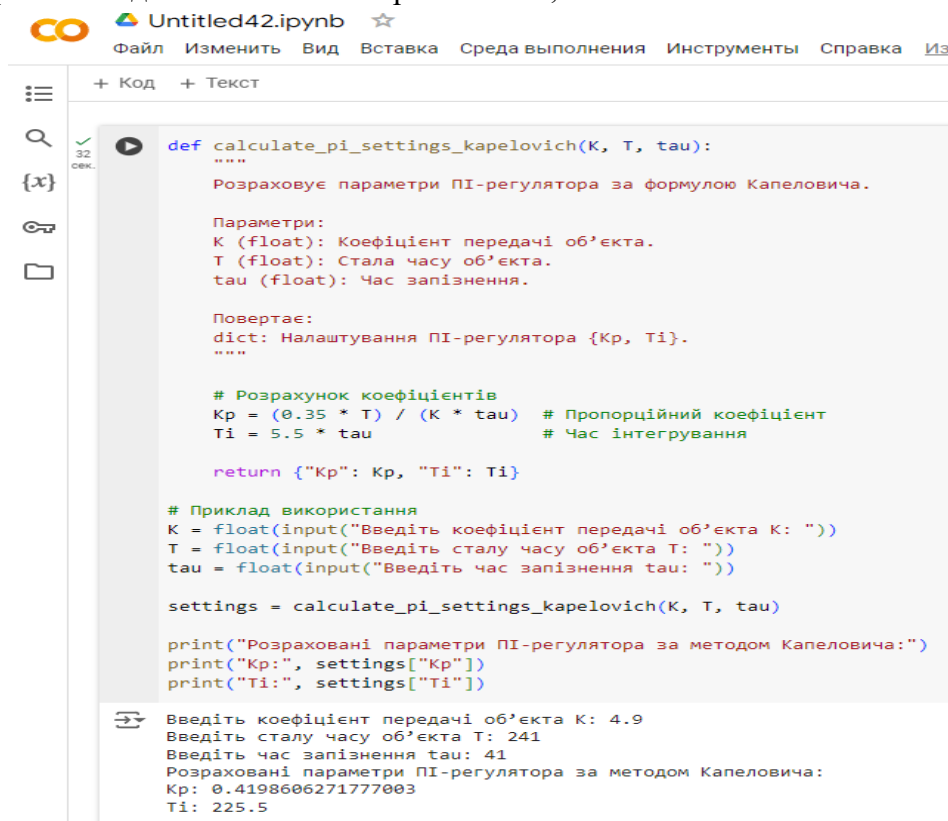
τ — час запізнення.

Тоді коефіцієнти ПІ-регулятора за формулою Капеловича визначаються як: $K_p=0.35 \cdot T / K$, $T_i=5.5 \cdot \tau$.

Код програми розрахунку адаптивних настроювальних параметрів типового регулятора в контурі САУ суднового парового агрегату та її робота наведено на рисунку 7.

Програма з вхідними параметрами K , T , τ розраховує коефіцієнти K_p і T_i за формулами Капеловича та виводить їх.

Ці значення допоможуть більш точно налаштувати ПІ-регулятор під об'єкт керування з урахуванням запізнення. При введенні коефіцієнту передачі об'єкта $K=4.9$, сталої часу об'єкта $T=241$ с та часу запізнення $\tau=41$ с розраховані параметри ПІ-регулятора за методом Капеловича: $K_p: 0.41$ Т/ К, $T_i: 225.5$ τ.



```
def calculate_pi_settings_kapelovich(K, T, tau):
    """
    Розраховує параметри ПІ-регулятора за формулою Капеловича.

    Параметри:
    K (float): Коефіцієнт передачі об'єкта.
    T (float): Стала часу об'єкта.
    tau (float): Час запізнення.

    Повертає:
    dict: Налаштування ПІ-регулятора {Kp, Ti}.
    """

    # Розрахунок коефіцієнтів
    Kp = (0.35 * T) / (K * tau) # Пропорційний коефіцієнт
    Ti = 5.5 * tau             # Час інтегрування

    return {"Kp": Kp, "Ti": Ti}

# Приклад використання
K = float(input("Введіть коефіцієнт передачі об'єкта K: "))
T = float(input("Введіть сталу часу об'єкта T: "))
tau = float(input("Введіть час запізнення tau: "))

settings = calculate_pi_settings_kapelovich(K, T, tau)

print("Розраховані параметри ПІ-регулятора за методом Капеловича:")
print("Kp:", settings["Kp"])
print("Ti:", settings["Ti"])

Введіть коефіцієнт передачі об'єкта K: 4.9
Введіть сталу часу об'єкта T: 241
Введіть час запізнення tau: 41
Розраховані параметри ПІ-регулятора за методом Капеловича:
Kp: 0.4198606271777003
Ti: 225.5
```

Рисунок 7 - Скріншот роботи програми виконаної мовою Пітон в інтерактивному середовищі Google Colab

Моделювання або апробація багаторежимного адаптивного регулятора підтримки заданого тиску перегрітої пари, що надходить на суднову парову турбіну, показана на рис. На рис. 8-9 канал зовнішнього збурення – зміна навантаження судових парових агрегатів.

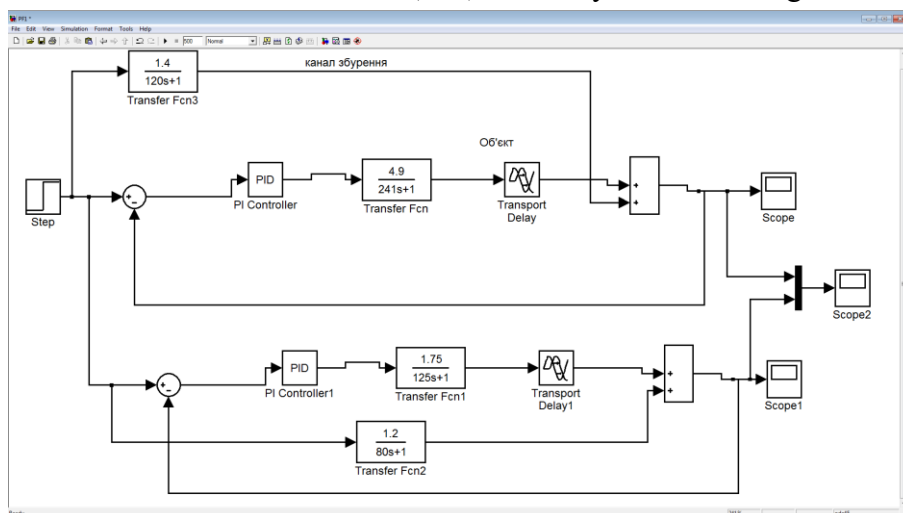
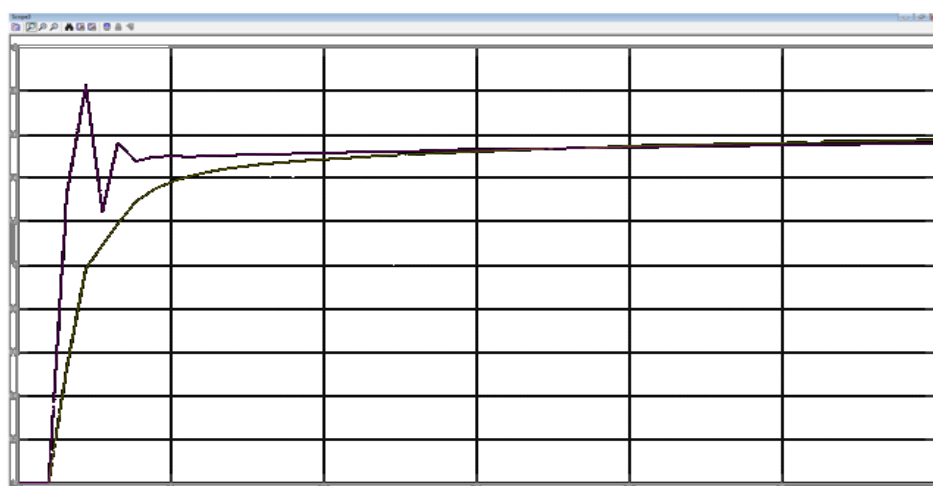


Рисунок 8 - Імітаційне моделювання адаптивної САУ

Перехідні процеси адаптивної САУ тиску пари парового котла показані на рисунку 10.

Р, МПа



t, c

Рисунок 9 - Перехідні процеси адаптивного САУ тиску пари під час роботи СПТУ на різних навантажувальних режимах

Аналіз показників якості перехідних процесів (час регулювання, помилка, занедбаність) відповідає нормативним значенням. Таким чином, запропонована адаптивна нейро-нечітка САУ успішно компенсує зовнішні збурення і є стійкою.

Висновки. Аналіз існуючих нейронечітких моделей показав, що вибір типу моделі залежить від багатьох факторів. Як орієнтир використовуються для підвищення ефективності управління наступні фактори: швидкість навчання, онлайн адаптивність, мінімізація помилок, простота розробки та обслуговування. При розробці ННМ біло вібрано - кількість циклів навчання - 30, метод навчання - зворотного поширення помилки, тип функцій належності вхідного та вихідних параметрів – трикутний. Таким чином,

запропонована методика дозволяє розробити ефективну систему параметричної ідентифікації та адаптації параметрів суднових систем управління в умовах частих змін парового навантаження. Впровадження такої системи на суднах з великою кількістю парогенеруючих установок та споживачів парової енергії може дозволити своєчасно та точно досягати заданих виробничих параметрів при роботі парових агрегатів на всьому діапазоні навантажень.

ЛІТЕРАТУРА

1. Åström KJ, Wittenmark B (1984) Computer controlled systems, theory and design. Prentice-Hall, Englewood Cliffs/ Доступ <https://bayanbox.ir/view/8821671619590593184/Adaptive-Control-Astrom-2nd-edition.pdf>
2. Kuo B. S. Automatic Control Systems / Benjamin C. Kuo. – 9th ed. – New York : John Wiley & Sons, 2010. – 1104 p.
3. Technical specifications of auxiliary and utilization boilers of the brands: KAV, KUP, Clayton, Zafa, Hitachi, Mitsubishi. URL: <http://shipservice.com.ua/>.
4. Mikhailenko V. S., Kharchenko R. Yu. Analysis of Traditional and Neuro Fuzzy Adaptive System of Controlling the Primary Steam Temperature in the Direct Flow Steam Generators in Thermal Power Stations / V. S. Mikhailenko, R. Yu. Kharchenko // Automatic Control and Computer Sciences. – 2014. – Vol. 48, № 6. – P. 334–344. DOI:10.3103/S0146411614060066
5. Mykhailenko V. S. Analysis of methods for adaptation of industrial control systems of thermal processes / V. S. Mykhailenko // Науковий вісник Національного гірничого університету. – 2014. – № 4. – С. 58–65. URL: http://nbuv.gov.ua/UJRN/Nvngu_2014_4_11.
6. Takagi T., Sugeno M. Fuzzy identification of system and its applications to modeling and control / T. Takagi, M. Sugeno // IEEE Transactions on Systems, Man, and Cybernetics. – 1985. – Vol. 15, № 1. – P. 166–132. <http://dx.doi.org/10.1109/TSMC.1985.6313399>
7. Mikhailenko V. S., Solodovnik M. S. Statement of the Synthesis Problem of the Intellectual System of Adaptive Management / V. S. Mikhailenko, M. S. Solodovnik // Lecture Notes in Economics and Mathematical Systems / Gil-Lafuente, Zopounidis Constantin (Eds.). – 2015. – Vol. 675. – 223 p. – P. 165–179. (Springer).
8. Zhang Y., Chen Z. Q., Yang P., Yuan Z. Z. Neural network-based PID predictive control for nonlinear time-delay systems / Y. Zhang, Z. Q. Chen, P. Yang, Z. Z. Yuan // Proceedings of International Conference on Machine Learning and Cybernetics. – 26–29 Aug. 2017. – Vol. 2. – P. 1014–1018. doi:10.1109/ICMLC.2004.1382336
9. Lee K. Y, L. Y. Ma, C. J. Boo, W.-H. Jung, and S.-Ho Kim. Inverse dynamic neuro-controller for superheater steam temperature control of a large-scale ultra supercritical (USC) boiler unit. // Proc. of the IFAC Symposium on Power Plants and Power Systems Control, in Tampere, Finland, July 5 – 8. – 2019.
10. Improving Direct Efficiency of Boilers – Role of Automation. URL: <https://www.forbesmarshall.com>.
11. Marine Boiler and Steam Turbine Generator. URL: https://www.mhimme.com/auxiliary_boilers.html

REFERENCES

1. Åström KJ, Wittenmark B (1984) Computer controlled systems, theory and design. Prentice-Hall, Englewood Cliffs/ Доступ
<https://bayanbox.ir/view/8821671619590593184/Adaptive-Control-Astrom-2nd-edition.pdf>
2. Kuo B. S. Automatic Control Systems / Benjamin C. Kuo. – 9th ed. – New York : John Wiley & Sons, 2010. – 1104 p.
3. Technical specifications of auxiliary and utilization boilers of the brands: KAV, KUP, Clayton, Zafa, Hitachi, Mitsubishi. URL: <http://shipservice.com.ua/>.
4. Mikhailenko V. S., Kharchenko R. Yu. Analysis of Traditional and Neuro Fuzzy Adaptive System of Controlling the Primary Steam Temperature in the Direct Flow Steam Generators in Thermal Power Stations / V. S. Mikhailenko, R. Yu. Kharchenko // Automatic Control and Computer Sciences. – 2014. – Vol. 48, № 6. – P. 334–344.
DOI:10.3103/S0146411614060066
5. Mykhailenko V. S. Analysis of methods for adaptation of industrial control systems of thermal processes / V. S. Mykhailenko // Науковий вісник Національного гірничого університету. – 2014. – № 4. – С. 58–65.
URL: http://nbuv.gov.ua/UJRN/Nvngu_2014_4_11.
6. Takagi T., Sugeno M. Fuzzy identification of system and its applications to modeling and control / T. Takagi, M. Sugeno // IEEE Transactions on Systems, Man, and Cybernetics. – 1985. – Vol. 15, № 1. – P. 166–132. <http://dx.doi.org/10.1109/TSMC.1985.6313399>
7. Mikhailenko V. S., Solodovnik M. S. Statement of the Synthesis Problem of the Intellectual System of Adaptive Management / V. S. Mikhailenko, M. S. Solodovnik // Lecture Notes in Economics and Mathematical Systems / Gil-Lafuente, Zopounidis Constantin (Eds.). – 2015. – Vol. 675. – 223 p. – P. 165–179. (Springer).
8. Zhang Y., Chen Z. Q., Yang P., Yuan Z. Z. Neural network-based PID predictive control for nonlinear time-delay systems / Y. Zhang, Z. Q. Chen, P. Yang, Z. Z. Yuan // Proceedings of International Conference on Machine Learning and Cybernetics. – 26–29 Aug. 2017. – Vol. 2. – P. 1014–1018. doi:10.1109/ICMLC.2004.1382336
9. Lee K. Y, L. Y. Ma, C. J. Boo, W.-H. Jung, and S.-Ho Kim. Inverse dynamic neuro-controller for superheater steam temperature control of a large-scale ultra supercritical (USC) boiler unit. // Proc. of the IFAC Symposium on Power Plants and Power Systems Control, in Tampere, Finland, July 5 – 8. – 2019.
10. Improving Direct Efficiency of Boilers – Role of Automation.
URL: <https://www.forbesmarshall.com>.
11. Marine Boiler and Steam Turbine Generator.
URL: https://www.mhimme.com/auxiliary_boilers.html

Received 07.12.2024.
Accepted 12.12.2024.

Methodology for the development of neuro-fuzzy automatic control systems with a function for identifying the parameters

The article discusses a methodology for developing neuro-fuzzy automatic control systems (ACS) for marine steam turbine installations (MSTI) with a parameter identification function during their operation. The proposed methodology includes stages of MSTI dynamic modeling, the development of parameter identification algorithms based on neural networks, and their integration with fuzzy logic for decision-making. An analysis of the proposed approach's capabilities regarding the enhancement of reliability and stability of marine power plants has been conducted. The results obtained demonstrate that such a system can adjust model parameters in real-time, ensuring control accuracy and reducing the risk of emergency situations. The methodology can be implemented in real marine power systems that require automated control of complex processes.

Keywords: neuro-fuzzy system, marine steam turbine installation, neural networks, dynamic systems, adaptive control, automatic control systems, thermal power plants.

Михайленко Владислав Сергійович – д.т.н., професор, професор кафедри електрообладнання і автоматики суден Одеського національного університету «Одеська морська академія».

Коренкова Ганна Валентинівна – к.ф.-м.н., доцент, доцент кафедри комп'ютерних систем та технологій Одеського національного університету імені І.І.Мечникова.

Лешенко Валерій Владиславович – ст. викладач кафедри електрообладнання і автоматики суден Одеського національного університету «Одеська морська академія».

Мартинович Лариса Ярославівна – ст. викладач кафедри комп'ютерних систем та технологій Одеського національного університету імені І.І.Мечникова.

Mikhailenko Vladyslav Serhiiovych – Doctor of Technical Sciences, Professor, Professor of the Department of Ship Electrical Equipment and Automation, Odesa National University "Odessa Maritime Academy".

Korenkova Hanna Valentynivna – Candidate of Physical and Mathematical Sciences, Associate Professor, Associate Professor of the Department of Computer Systems and Technologies, Odesa I.I. Mechnikov National University.

Leshchenko Valerii Vladyslavovych – Senior Lecturer of the Department of Ship Electrical Equipment and Automation, Odesa National University "Odessa Maritime Academy".

Martynovych Larysa Yaroslavivna – Senior Lecturer of the Department of Computer Systems and Technologies, Odesa I.I. Mechnikov National University.

Н.А. Гук, М.Ю. Мітіков

МАТЕМАТИЧНА МОДЕЛЬ ОПТИМІЗАЦІЇ ПОШУКУ ДУБЛІКАТІВ ОБ'ЄКТІВ ТИПА STRING У ЗНІМКАХ ПАМ'ЯТІ

Анотація. Мета цієї роботи полягає у виявленні збільшеного використання пам'яті програмними додатками. Сучасний цикл розробки програмного забезпечення зосереджений на функціональності і часто ігнорує аспекти оптимального використання ресурсів. Обмежене фізичне масштабування задає верхній ліміт на пропускну здатність системи оброблювати запити. Наявність незмінних об'єктів з однаковою інформацією є ознакою збільшеної витрати пам'яті. Уникнення дублікатів об'єктів в пам'яті дозволяє більш раціонально використовувати існуючий ресурс і збільшити обсяги оброблюваної інформації. Існуючі наукові публікації фокусуються на дослідженні проблем витоків пам'яті, та обмежують увагою саме надмірне використання пам'яті через відсутність уніфікованої моделі пошуку надмірного використання пам'яті. Варто зазначити, що існуючі шаблони програмування містять шаблон «нул об'єктів», але залишають висновок про доцільність його впровадження інженерам, не надаючи математичного підґрунтя. Представлено розробку математичної моделі для процесу виявлення дублікатів об'єктів з властивістю незмінності типу String в знімку пам'яті. Проаналізовано промислові системи, які вимагають сотні ГБ оперативної пам'яті для роботи та містять мільйони об'єктів в оперативній пам'яті. За таких масштабів даних, існує необхідність оптимізувати саме процес пошуку дублікатів. Методом дослідження є аналіз знімків пам'яті високонавантажених систем за допомогою програмного коду, розробленого на технології .NET та бібліотеці ClrMD. Знімок пам'яті відображає стан досліджуваного процесу у момент часу, містить усі об'єкти, потоки та виконувані операції. Бібліотека ClrMD дозволяє програмно досліджувати об'єкти, їх типи, отримувати значення полів, будувати графи зв'язків між об'єктами. За результатами дослідження було запропоновано оптимізацію яка дозволяє пришвидшити процес пошуку дублікатів у декілька разів. Науковий внесок дослідження полягає в створенні математично обґрунтованого підходу, який сприяє значному зменшенню використання ресурсів пам'яті та оптимізації обчислювальних процесів. Практична користь моделі підтверджується результатами оптимізації досягнутих завдяки отриманим рекомендаціям, зниженням витрат на хостинг (що забезпечує більшу економічну ефективність у розгортанні та використанні програмних систем у промислових умовах), а також збільшення обсягів оброблених даних.

Ключові слова: математична модель; оптимізація; алгоритм; продуктивність; знімок пам'яті; дублювання; строка

1. Вступ. Розвиток інформаційних технологій поширив всебічне застосування програмних додатків в багатьох сферах сьогодення. Необхідність оброблювати все більші обсяги інформації може компенсуватись збільшенням розрахункової потужності [1] або оптимізацією існуючих програмних реалізацій [2].

Враховуючи лінійне зростання вартості одиниці оперативної пам'яті у хмарному хостингу [3], складається хибне враження можливості лінійного масштабування сукупних експлуатаційних витрат. Існуючі пропозиції віртуальних машин пропонуються з розмірами оперативної пам'яті кратними ступеню двійки (4ГБ – 8ГБ – 16ГБ – 32ГБ -..). При потребі одного додаткового гігабайта понад наявної пам'яті виникає необхідність масштабування на наступну пропозицію яка мінімум у 2 рази дорожче за попередню (з урахуванням кількості процесорної потужності).

Оптимізації використання пам'яті програмним додатком потребують глибокого розуміння функціонування системи та доступу до процесу розробки, що також робить цей метод економічно необґрунтованим при наявності доцільнішої можливості масштабування. Масштабування розрахункових потужностей є вигідним при малих розмірах, але стає дедалі менш економічно-обґрунтованим при кожному збільшенні розміру віртуальної машини.

Варто зазначити, що сучасні методології розробки базуються на наявності декількох середовищ для тестування перед випуском версії у публічний доступ [4] і наявності вторинної репліки/регіону (5) для підвищення доступності. Таким чином, перехід на вищий рівень суттєво збільшує загальну вартість експлуатаційних витрат.

Існуючі публікації в більшості [6], [7] фокусуються на дослідженні проблеми витоків пам'яті, приділяючи менше уваги надмірному використанню пам'яті. Відомо, що витік пам'яті – це стан, коли програма помилково не може вивільнити більше непотрібну пам'ять для повторного використання, тож продовжує виділяти нові блоки.

На рисунку 1 представлено стрімке збільшення використання пам'яті програмним додатком, що призводить до зниження продуктивності і необхідності перезапуску системою моніторинга кожні декілька годин. Високонавантажені системи, які працюють з великими обсягами даних особливо вразливі до цієї проблеми.

Для знаходження першопричини витоків використовуються знімки пам'яті [8] – запис вмісту оперативної пам'яті у певний момент часу.

Знімки пам'яті також використовуються для пошуку шкідливого програмного коду [9], особливо при ін'єкції блоку на виконання в довірений процес. При наявному фокусі на ідентифікацію витоків, існуючі дослідження обділяють увагою проблему надмірного використання пам'яті – коли програмний додаток використовує більше пам'яті ніж того потребує розв'язувана задача. Зазначена проблема може виникнути з різних причин, але одна з найпоширеніших це дублювання незмінних об'єктів. Якщо програма створює декілька об'єктів з однаковими значеннями замість одного з властивістю незмінності, то це призводить до необхідності зберігати в пам'яті дублюючу інформацію і збільшує вимоги щодо обсягу пам'яті.

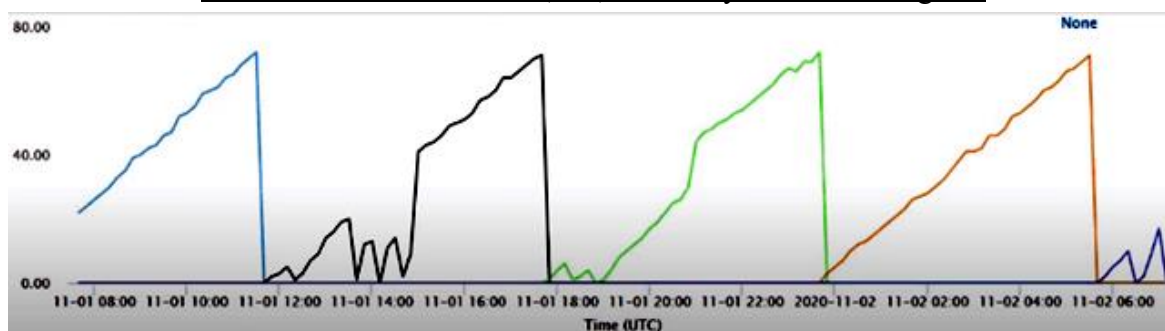


Рисунок 1 - Динаміка використання пам'яті

Однією з поширених проблем функціонування сучасних систем є дублювання об'єктів типу String (строка), що є серйозним викликом, який призводить до збільшення загальної вартості експлуатаційних витрат і постає непомітним для існуючих діагностичних систем бо саме критерій витоку пам'яті не виконується.

Для знаходження дублікатів необхідно виконати порівняння усіх строк між собою, що призведе до надвеликої кількості операцій і впливатиме на швидкодію додатку у негативному ключі.

Метою роботи є розробка нової математичної моделі виявлення дублікатів об'єктів типу «строка» на основі аналізу знімків пам'яті.

2. Основні результати. Візуалізація вартості користування хмарними обчислювальними машинами [3] в залежності від розміру оперативної пам'яті відображає збільшення вартості мінімум вдвічі при вичерпанні обсягу пам'яті віртуальної машини і необхідності використовувати наступний рівень.

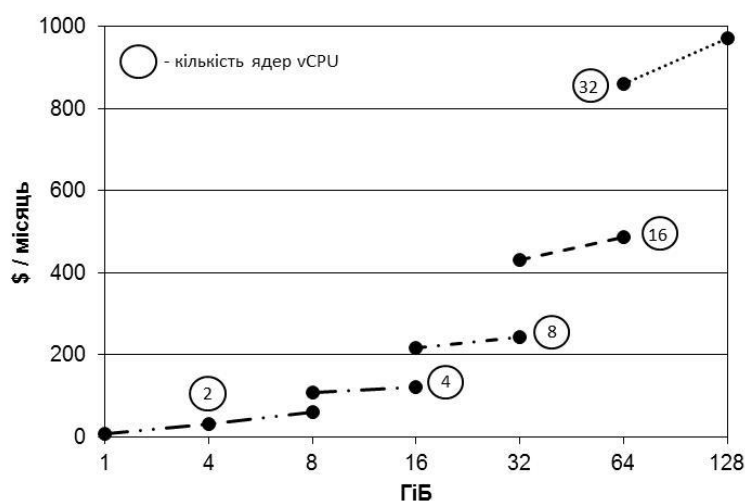


Рисунок 2 - Залежність вартості хостингу від об'єму оперативної пам'яті

Дослідники Adriaan Labuschagne та Laura Inozemtseva [4] у роботі по дослідженню регресійного тестування наочно продемонстрували методи для підвищення якості продукту. Тестування продукту у схожих до реального середовищах дозволяє суттєво зменшити ризики появи помилок у функціоналі [10], але вимагає наявності середовищ і збільшує кількість розрахункових вузлів (віртуальних машин).

Велика увага приділяється механізмам знаходження витоків пам'яті в роботах Gene Novark, Emery D. Berger, Benjamin G. [6] та Markus Weninger [7].

Jon Louis Bentley [2] описує підходи до зменшення використання пам'яті, але механізми по знаходженню кандидатів для оптимізації знаходяться поза рамками його робіт. Ioannis T. Christou, Sofoklis Efremidis [11] описують використання патерну проектування «пул об'єктів». За результатами їх досліджень, використання шаблону проектування може значно зменшити час відгуку високопродуктивних багатопотокових застосунків, особливо у середовищах з обмеженою пам'яттю. Механізм знаходження належних використань не описується.

Автори цих досліджень сходяться у думці, що знаходження застосувань цього шаблону залежить від конкретного програмного продукту і даними якими він оперує. Надмірне використання пам'яті може бути серйозною проблемою для високонавантажених систем. Однак, з розумінням причин неоптимального розподілу пам'яті та використанням відповідних стратегій для їх уникнення, можна підвищити стабільність та ефективність роботи системи.

Такий сценарій може бути важко виявити, оскільки він не проявляється явно – критерій витоку пам'яті не виконується.

Для виявлення дублювання даних необхідно проаналізувати пам'ять програми і виділити об'єкти які мають властивість незмінності. Властивість незмінності може задаватись як на рівні мови програмування [12], так і проявлятися для існуючого коду через відсутність викликів методів які можуть змінювати стан об'єкту.

Знаходження властивості незмінності є ключем до зменшення витрат пам'яті через повторне використання однакових об'єктів [13], і потребує наявності повної кодової бази виконуваної програми.

Знімок пам'яті програми може бути зображенням у вигляді структурованого масиву байтів в такий спосіб:

$$M_{time} = [b_1, b_2, \dots, b_C],$$

де *time* – момент часу, у який відбувається зняття знімку пам'яті; b_1, b_2, \dots, b_C – байти, з яких складається знімок; *C* – кількість байтів.

В знімку пам'яті відображається інформація про операції, що виконуються, стан потоків виконання, об'єкти, якими ці потоки керують та програмний код. Відомості про об'єкти вміщують їх типи, стани та взаємозв'язки.

Знімки пам'яті широко використовуються для аналізу та виявлення шкідливого коду який може бути введений в пам'ять довіреного процесу. Традиційні статичні та динамічні методи неефективні у виявленні шкідливих програм, що проживають у пам'яті [14]. Крім того, існуючі рішення з використанням форензичного аналізу пам'яті показують незадовільну ефективність у плані швидкості виявлення та залежать від масивних знань експертів у аналізі пам'яті. Рішенням цієї проблеми є зняття знімків пам'яті здорових процесів для використання у якості навчальної бази для алгоритмів машинного навчання. При виявленні невідповідності, система безпеки має змогу швидко відреагувати попри відсутність фізичного файлу-загрози.

Враховуючи наявність великої кількості об'єктів в пам'яті промислових програмних додатків, ручний пошук дублікатів поміж тисяч різних типів не є доцільним. Необхідно обмежити пошук до типів які займають найбільшу кількість місця в пам'яті, у більшості випадків це об'єкти типу «строка».

Крім того, створення, модифікація та видалення об'єктів типу «строка» вимагає певних ресурсів. На рисунку 3 наведений приклад дублювання об'єктів з досліджуваної промислової системи, яка споживає більше 20ГБ пам'яті. Цей програмний додаток не відповідає ознакам витоку пам'яті, а саме монотонному зростанню використаної пам'яті, тож існуючі системи моніторингу не знаходять проблем з використанням пам'яті. Завдяки знімку пам'яті, були знайдені дублікати об'єктів з властивістю незмінності, тож ідентифіковане збільшене використання оперативної пам'яті.

```
00000181696ce700 (Model.CountryToCountrySetting)
<ObjectState>k__BackingField:0x0 (Undefined) (Data.CacheObjectStates)
<SourceCountryId>k__BackingField:0xe2 (System.Int64)
<TargetCountryId>k__BackingField:0x20 (System.Int64)
<SettingName>k__BackingField:00000181696ce740 (System.String) Length=34, String="AddShippingCostToCommercialInvoice"
<SettingValue>k__BackingField:00000181696ce7a0 (System.String) Length=6, String="5;USD"
<SettingDescription>k__BackingField:00000181696ce7c8 (System.String) Length=111, String="Value for Commercial and Parcel Invoice,
00000181696ce8c0 (Model.CountryToCountrySetting)
<ObjectState>k__BackingField:0x0 (Undefined) (Data.CacheObjectStates)
<SourceCountryId>k__BackingField:0xe2 (System.Int64)
<TargetCountryId>k__BackingField:0x21 (System.Int64)
<SettingName>k__BackingField:00000181696ce900 (System.String) Length=34, String="AddShippingCostToCommercialInvoice"
<SettingValue>k__BackingField:00000181696ce960 (System.String) Length=6, String="5;USD"
<SettingDescription>k__BackingField:00000181696ce988 (System.String) Length=111, String="Value for Commercial and Parcel Invoice,
00000181696cea80 (Model.CountryToCountrySetting)
<ObjectState>k__BackingField:0x0 (Undefined) (Data.CacheObjectStates)
<SourceCountryId>k__BackingField:0xe2 (System.Int64)
<TargetCountryId>k__BackingField:0x22 (System.Int64)
<SettingName>k__BackingField:00000181696ceac0 (System.String) Length=34, String="AddShippingCostToCommercialInvoice"
<SettingValue>k__BackingField:00000181696ceb20 (System.String) Length=6, String="5;USD"
<SettingDescription>k__BackingField:00000181696ceb48 (System.String) Length=111, String="Value for Commercial and Parcel Invoice,
00000181696cec40 (Model.CountryToCountrySetting)
<ObjectState>k__BackingField:0x0 (Undefined) (Data.CacheObjectStates)
<SourceCountryId>k__BackingField:0xe2 (System.Int64)
<TargetCountryId>k__BackingField:0x23 (System.Int64)
<SettingName>k__BackingField:00000181696cec80 (System.String) Length=34, String="AddShippingCostToCommercialInvoice"
<SettingValue>k__BackingField:00000181696cece0 (System.String) Length=6, String="5;USD"
<SettingDescription>k__BackingField:00000181696ced08 (System.String) Length=111, String="Value for Commercial and Parcel Invoice,
```

Рисунок 3 - Приклад дублювання строк в об'єктах

Сканування знімків пам'яті – це важливий процес для розуміння механізму роботи програмного додатку та діагностування проблем [2]. Автоматизація процесу сканування знімків пам'яті - це ефективний метод, який допомагає швидко проводити аналіз роботи високонавантажених систем. Тому автоматизація сканування може принести значні переваги, зокрема зменшення часу на аналіз, підвищення точності виявлення проблем та покращення загальної продуктивності системи [3].

Ідентифікуємо об'єкти у знімку пам'яті за їх властивостями, і визначимо це як метрику. Метрика об'єктів вводиться для визначення «відстані» між об'єктами та їх схожості. Це можуть бути властивості об'єктів як то час, «вага», «довжина» і т.ін. Якщо таких властивостей нема, тоді створюємо метрику штучно.

Автоматизована система сканування утворить множину об'єктів O [1]. Застосуємо функцію групування за типами і знайдемо підмножину об'єктів, які мають незмінні типи O_i . У цій підмножині виділимо підмножину $STRN$ всіх об'єктів типу «строка» і підмножину STR об'єктів, які мають властивість незмінності. Таким чином, виділення підмножини «строк»:

$$STR \subset STRN \subset O_{t_i} \subset O. \quad (1)$$

Об'єкт типу «строка» має нульовий байт $b(0)$, довжину об'єкта у байтах і інформацію у цих байтах. Таким чином, об'єкт типу «строка» можна записати:

$$String(b(0), b).$$

Введемо додаткову властивість s об'єкту, як арифметичну суму байтів, що виділяються для збереження змісту об'єкта. Що дозволяє вираз для об'єкту типу «строка» записати у такий спосіб

$$String(b(0), b, s).$$

Кожен екземпляр об'єкту $String$ належить множині:

$$STRN = \{String_1(b_1(0), b_1, s_1), String_2(b_2(0), b_2, s_2), \dots, String_N(b_N(0), b_N, s_N)\}. \quad (2)$$

де N – кількість екземплярів об'єктів типу «строка». У цю множину ввійдуть всі об'єкти типу «строка», в тому числі дублікати.

Кожен екземпляр об'єкта «строка» ідентифікується по параметром s , який утворює множину

$$S = \{s_1, s_2, \dots, s_U\}.$$

де U – кількість унікальних розмірів об'єктів типу «строка», $U \leq N$.

У множині S знайдемо мінімальне a і максимальне b значення. Відрізок $[a, b]$ розіб'ємо на m напівінтервалів довжиною

$$\Delta = \frac{b-a}{m}, \quad m \leq U. \quad (3)$$

У кожен напівінтервал може ввійти кілька значень параметра s об'єктів типу «строка», а може і жодного. Розподіл по інтервалам параметра s проводиться з підпорядкуванням підмножин:

$$k = \prod_{i=1}^N \left(\frac{s_i}{\Delta} \right), \quad k \leq U, \quad (4)$$

де k – номер інтервалу.

Таким чином, відбувається часткове упорядкування множини S по напівінтервалам. Зауважимо, що отримані напівінтервали можуть містити строки з однаковим значенням s_i , але бути різними за вмістом. Наочним прикладом є строки, в яких змінений порядок слів, але алгебраїчна сума байтів що відводиться під їх збереження, буде однаковою. Функція підсумовування було обрано як швидку хеш-функцію. Серед множини напівінтервалів знайдуться такі, що будуть містити кілька неупорядкованих значень. В таких напівінтервалах слід провести додаткове упорядкування. Тому ставиться задача здійснити мінімальну кількість упорядкувань. Позначимо функцією $G(m, L)$ загальну кількість упорядкувань. Таким чином, виникає задача знайти мінімум $G(m, L)$. Введемо припущення, у кожен інтервал попадає однакова кількість чисел L , в такому разі загальна кількість упорядкувань дорівнює

$$G(m, L) = mL \frac{L-1}{2}. \quad (5)$$

Треба встановити значення m і L при яких функція $G(m, L)$ набуде найменше

значення. Оскільки $m = \frac{U}{L}$, то після підстановки загальна кількість упорядкувань

$$G(U, L) = U \frac{L-1}{2}. \quad (6)$$

Мінімум функції $G(m, L)$ залежить від значення параметра L . Тривіальний варіант $L=1$ не розглядаємо. Під час виконання операції сортування потрібно щонайменше два значення, тому мінімальне значення функція $G(m, L)$ набуде при $L=2$. В такому разі:

$$\min G(U, L) = \frac{U}{2}. \quad (7)$$

За такого значення L кількість напівінтервалів дорівнює $m = \frac{U}{2}$.

Оскільки заздалегідь невідомо, на яку кількість нерівних напівінтервалів слід робити множину S так, щоб ці напівінтервали містили однакову кількість чисел, вводимо оцінку:

$$\inf G(U, L) = \frac{U}{2}. \quad (8)$$

Верхню границю значень для функції G знайдемо з рівняння (5) при $L=U$ і $m=1$:

$$\sup G(U, L) = U \frac{U-1}{2}. \quad (9)$$

Побудуємо розподіл на основі даних зі знімку пам'яті однієї з досліджуваних промислових систем.

3. Приклади. На рисунку 4 показано розподіл об'єктів типу «строка» для типу «Model.CountryToCountrySetting» по параметру s . На графіку присутні три піки які демонструють значне збільшення елементів з однаковим значенням контрольної суми s_i , і відповідають відповідно трьом об'єктам «5;;USD», «AddShippingCostToCommercialInvoice», «Value for Commercial and Parcel Invoice». Відповідно схемі побудови напівінтервалів і сортуванню кожна вершина графіку відповідає одному значенню s_i та повністю ідентифікує об'єкт. Що дозволяє швидко проводити аналіз і ідентифікацію.

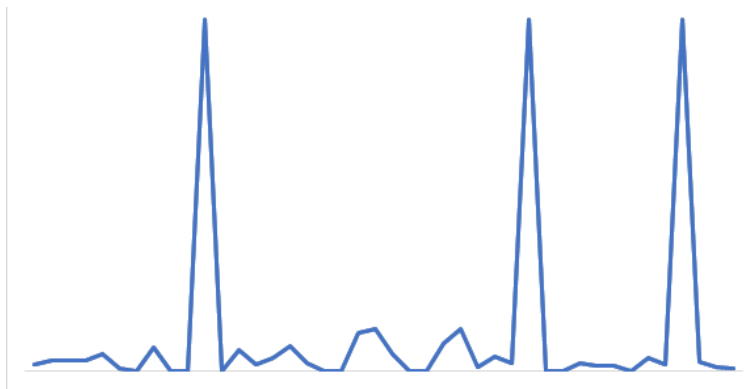


Рисунок 4 - Розподіл по множині S

Проведене групування за параметром s_i підтвердило наявність 220 тисяч унікальних значень і 14.5 мільйона дублікатів. Знайдено найбільш дубльовані строки, таким чином рисунок 5 підтверджує початкове припущення про наявність аномального дублювання у знімку пам'яті. Утворимо множину s^1 параметрів s_i цих об'єктів, $s^1 = \{s_{k1}, s_{k2}, s_{k3}\}$. Множина s^1 входить у множину S , тобто $s^1 \subset S$. Видаливши множину s^1 з множини S , отримаємо нову множину $S^1 = S \setminus s^1$. Знову будемо розподіл по множині S^1 .



Рисунок 5 - Розподіл по множині S^1

Отримуємо новий розподіл. По якому розподілу були знайдені об'єкти (таблиця 1).

Таблиця 1

Назва таблиці

«Строка»	Кількість дублікатів	Об'єм пам'яті
True	635181	2.4 MB
HideStandardShippingMethod	390073	9.7 MB
false	318273	1.5 MB
Restricted due to ticket	257600	8.4 MB
Visa	241913	945 KB

Визначимо ефективність цього метода по кількості операцій, порівняно з простим перебором [4]. При простому переборі буде виконано $U \frac{U-1}{2}$ операцій порівнянь.

4. Доведення. Визначимо коефіцієнт ефективності K_{ef} для $inf G(U, L)$ та $sup G(U, L)$ з додаванням U операцій розподілу по напівінтервалам та U операцій для визначення значень a і b

$$K_{ef}^{inf} = \frac{U \frac{U-1}{2}}{\frac{U}{2} + 2U} = \frac{U-1}{5},$$
$$K_{ef}^{sup} = \frac{U \frac{U-1}{2}}{U \frac{U-1}{2} + 2U} = \frac{U-1}{U+3},$$

Таким чином коефіцієнт ефективності буде знаходитись у межах

$$\frac{U-1}{U+3} \leq K_{ef} \leq \frac{U-1}{5}.$$

Наступним кроком оптимізації процесу пошуку дублікатів є зменшення початкового відрізка Δ (формула 3) з урахуванням розподілу строк, а саме великої кількості дублікатів до 150 символів. Таким чином було досягнуто більш рівного розподілу значень по напівінтервалам.

5. Висновки. Більшість публікацій за тематикою дослідження фокусуються на проблемі витоків пам'яті, обходячи надмірне споживання через неможливість його діагностування. Дослідження по застосуванню шаблонів проектування для зменшення використання пам'яті наводять механізми впровадження та ефект, але залишають стороною критерії та механіки пошуку кандидатів через велику варіативність сценаріїв та програмних продуктів.

Дана публікація фокусується саме на алгоритмі пошуку кандидатів за допомогою знімків пам'яті. Хоча знімки пам'яті використовуються для пошуку витоків пам'яті, або введеного шкідливого коду, їх використання для пошуку надмірного споживання пам'яті є новітнім впровадженням.

Для об'єктів типу «строка» була створена метрика. Побудовано покращений алгоритм пошуку дублікатів об'єктів типу «строка» у знімках пам'яті. Проведено тестування алгоритму на реальній промисловій системі який пришвидшує аналіз з 7464 до 1894 мс – майже в чотири рази.

Результати роботи оптимізованого алгоритму

Кількість підмножин	Максимальна кількість елементів в підмножині	Розмір Δ	Час виконання, мс.
2	2160701	46795	3321.95
4	2142365	23398	2653.41
8	1753658	11699	3152.39
16	1715889	5850	3100.36
32	1558702	2925	2323.03
64	1260964	1463	2274.14
128	862321	732	2089.94
256	535165	366	1968.16
512	325051	183	1894.38
1024	262314	92	2253.57
2048	174401	46	2074.31
10000	91012	10	1708.93
20000	50420	5	1789.49
30000	40239	4	1831.46
40000	30639	3	1881.13
50000	20531	2	1994.70

При збільшенні кількості підмножин, час виконання розбиття зменшується непропорційно зростанню використаної пам'яті процесом групування.

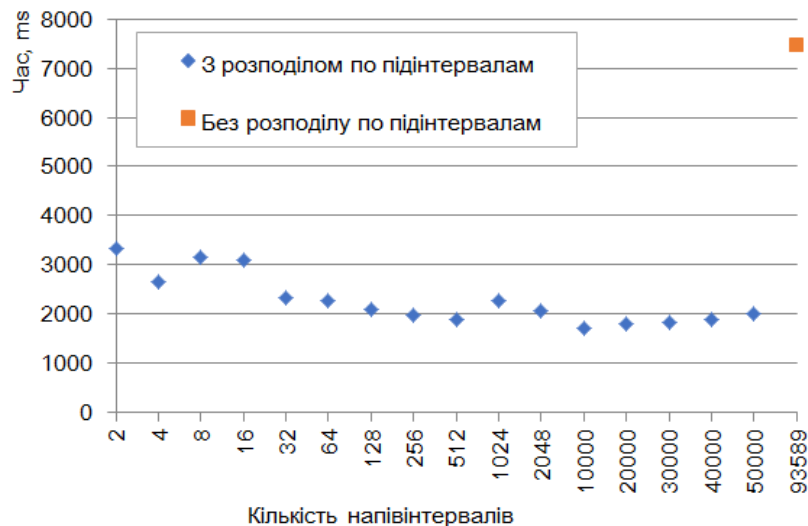


Рисунок 6 - Залежність часу виконання від кількості підмножин

Зменшення максимальної кількості елементів в підмножинах також спадає непропорційно збільшенню кількості підмножин.

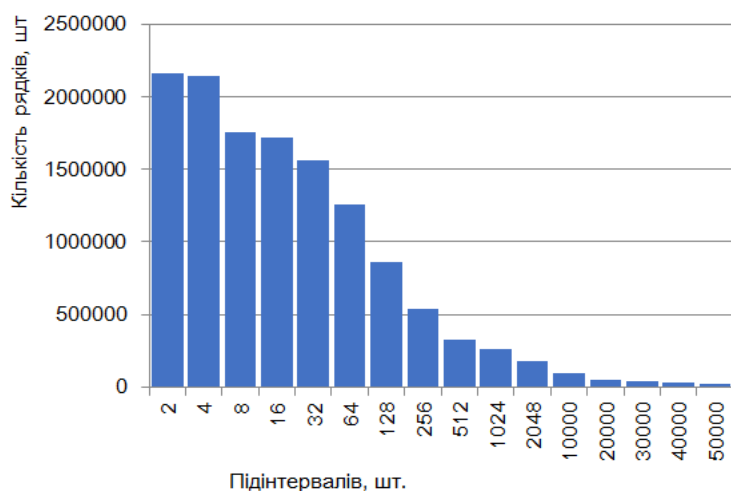


Рисунок 7 - Залежність максимальної кількості елементів в підмножині від її обсягу

Згідно зробленого припущення, найкраща швидкодія розбиття досягається при схожій кількості елементів в підмножинах для зменшення частки строк з однаковою контрольною сумою, але різних за змістом. Це припущення було підтверджене серією експериментів зі знімками пам'яті досліджуваних промислових систем, кількість підмножин для пришвидшення розбиття знаходиться у діапазоні (460 – 520). Подальше збільшення кількості підмножин не дає вагомого пришвидшення операції, але призводить до необхідності тримати в пам'яті значно більше число груп, тож збільшує вимоги до кількості пам'яті необхідної для проведення аналізу.

Внесок авторів.

Наталія Гук – концептуалізація, методика.

Миколай Мітіков – концептуалізація, математична модель, програмне забезпечення, проведення обчислювального експерименту, збір і перевірка емпіричних даних, аналіз джерел, підготовка огляду літератури або теоретичних основ дослідження.

ЛІТЕРАТУРА

1. Gregg, Brendan. 2.7.3 Scaling solutions. Systems Performance, Second Edition. Boston : Addison-Wesley, 2021, с. 929.
2. Bentley, Jon Louis. Writing efficient programs. Englewood Cliffs, N.J. : Prentice-Hall, 1982. ISBN-13/ 978-0139702440.
3. Microsoft. Windows Virtual Machines Pricing. Cloud Computing Services | Microsoft Azure. [Онлайнвий] Microsoft . [Цитовано: 8 February 2024 p.] <https://azure.microsoft.com/en-us/pricing/details/virtual-machines/windows/>.
4. Measuring the Cost of Regression Testing in Practice. Labuschagne, Adriaan, Laura Inozemtseva, Reid Holmes. New York : Association for Computing Machinery, 2017. 978-1-4503-5105-8.
5. Kleppmann, Martin. CHAPTER 5: Replication. Designing Data-Intensive Applications. місце видання невідоме : O'Reilly Media, 2017.

6. Efficiently and precisely locating memory leaks and bloat. Gene Novark, Emery D. Berger, Benjamin G. Zorn. Dublin : Proceedings of the 30th ACM SIGPLAN Conference on Programming Language Design and Implementation, 2009. 978-1-60558-392-1.
7. Analyzing Data Structure Growth Over Time to Facilitate Memory Leak Detection. Weninger, Markus. Mumbai : Association for Computing Machinery, 2019. 978-1-4503-6239-9.
8. Hewardt, Mario. Advanced .NET Debugging. Addison-Wesley Professional, 2009. 978-0-321-57889-1.
9. Hamad Naeem, Shi Dong, Olorunjube James Falana, Farhan Ullah. Development of a deep stacked ensemble with process based volatile memory forensics for platform independent malware detection and classification. Expert Systems with Applications. 2023 p., Т. 223.
10. Roy Oshero, Vladimir Khorikov. The Art of Unit Testing. місце видання невідоме : Manning, 2024. ISBN 9781617297489.
11. Ioannis T. Christou, Sofoklis Efremidis. To Pool or Not To Pool? Revisiting an Old Pattern. Marousi : Athens Information Technology, 2018. arXiv:1801.03763.
12. Immutable Objects for a Java-Like Language. C. Haack, E. Poll, J. Schäfer, A. Schubert. Berlin : Springer, 2007. 978-3-540-71316-6.
13. М.Ю. Мітіков, Н.А. Гук Огляд методів виявлення та аналізу проблем продуктивності в програмному забезпеченні: підходи, виклики та перспективи // Питання прикладної математики і математичного моделювання [Текст]: зб. наук. пр. / редкол.: О.М. Кісельова (відп. ред.) [та ін.]. – Дніпро, 2023. – Вип. 23. – с. 171 – 178. DOI: 10.15421/322318
14. MRm-DLDet: a memory-resident malware detection framework based on memory forensics and deep neural network. Liu, J., Feng, Y., Liu, X. 21, : Cybersecurity, 2023 p., Т. 6. 10.1186/s42400-023-00157-w.

REFERENCES

1. Gregg, Brendan. 2.7.3 Scaling solutions. Systems Performance, Second Edition. Boston : Addison-Wesley, 2021, с. 929.
2. Bentley, Jon Louis. Writing efficient programs. Englewood Cliffs, N.J. : Prentice-Hall, 1982. ISBN-13/ 978-0139702440.
3. 3. Microsoft. Windows Virtual Machines Pricing. Cloud Computing Services | Microsoft Azure. [Online] Microsoft . [Cited: 8 February 2024 p.] <https://azure.microsoft.com/en-us/pricing/details/virtual-machines/windows/>.
4. Measuring the Cost of Regression Testing in Practice. Labuschagne, Adriaan, Laura Inozemtseva, Reid Holmes. New York : Association for Computing Machinery, 2017. 978-1-4503-5105-8.
5. Kleppmann, Martin. CHAPTER 5: Replication. Designing Data-Intensive Applications. місце видання невідоме : O'Reilly Media, 2017.
6. Efficiently and precisely locating memory leaks and bloat. Gene Novark, Emery D. Berger, Benjamin G. Zorn. Dublin : Proceedings of the 30th ACM SIGPLAN Conference on Programming Language Design and Implementation, 2009. 978-1-60558-392-1.
7. Analyzing Data Structure Growth Over Time to Facilitate Memory Leak Detection. Weninger, Markus. Mumbai : Association for Computing Machinery, 2019. 978-1-4503-6239-9.

8. Hewardt, Mario. Advanced .NET Debugging. Addison-Wesley Professional, 2009. 978-0-321-57889-1.
9. Hamad Naeem, Shi Dong, Olorunjube James Falana, Farhan Ullah. Development of a deep stacked ensemble with process based volatile memory forensics for platform independent malware detection and classification. Expert Systems with Applications. 2023 p., T. 223.
10. Roy Oshero, Vladimir Khorikov. The Art of Unit Testing. Manning, 2024. ISBN 9781617297489.
11. Ioannis T. Christou, Sofoklis Efremidis. To Pool or Not To Pool? Revisiting an Old Pattern. Marousi : Athens Information Technology, 2018. arXiv:1801.03763.
12. Immutable Objects for a Java-Like Language. C. Haack, E. Poll, J. Schäfer, A. Schubert. Berlin : Springer, 2007. 978-3-540-71316-6.
13. M. Yu. Mitikov, N.A. Huk Overview of methods for identifying and analyzing performance problems in software: approaches, challenges and prospects // Issues of applied mathematics and mathematical modeling [Text]: coll. of science pr. / editor: O.M. Kiselyova (corresponding editor) [etc.]. – Dnipro, 2023. – Issue 23. - p. 171 – 178. DOI: 10.15421/322318
14. MRm-DLDet: a memory-resident malware detection framework based on memory forensics and deep neural network. Liu, J., Feng, Y., Liu, X. 21, Cybersecurity, 2023 p., T. 6. 10.1186/s42400-023-00157-w.

Received 02.12.2024.
Accepted 12.12.2024.

***Mathematical optimisation model for searching duplicate string objects
in the memory snapshot***

The purpose of this paper is to identify the increased memory usage of software applications. The modern software development cycle focuses on functionality and often ignores aspects of optimal resource usage. Limited physical scaling sets an upper limit on the system's capacity to process requests. The presence of unchanged objects with the same information is a sign of increased memory consumption. Avoiding duplicate objects in memory allows for a more rational use of the existing resource and an increase in the amount of information processed. Existing scientific publications focus on the study of memory leakage problems, and limit their attention to excessive memory usage due to the lack of a unified model for finding excessive memory usage. It is worth noting that existing programming templates contain the 'object pool' template, but leave the conclusion about the feasibility of its implementation to engineers without providing a mathematical basis. The paper presents the development of a mathematical model for the process of detecting duplicate objects with the immutability property of the String type in a memory snapshot. Industrial systems that require hundreds of GB of RAM to operate and contain millions of objects in RAM are analysed. Given this scale of data, there is a need to optimise the duplicate detection process. The research method is to analyse memory snapshots of highly loaded systems using the software code developed on .NET technology and the ClrMD library. The memory snapshot reflects the state of the process under study at a given time, contains all objects, threads and operations performed. The ClrMD library allows you to programmatically examine objects, their types, get field values, and build graphs of relationships between objects. Based on the results of the study, an opti-

misation was proposed that allows to speed up the process of finding duplicates several times. The scientific contribution of the study is the creation of a mathematically sound approach that significantly reduces the use of memory resources and optimises computing processes. The practical usefulness of the model is confirmed by the optimisation results achieved through the recommendations, reduced hosting costs (which provides greater cost-effectiveness in the deployment and use of software systems in industrial environments), and increased data processing.

Keywords: mathematical model; optimisation; algorithm; performance; memory snapshot; duplication; string

Гук Наталія Анатоліївна – завідувачка кафедри комп’ютерних технологій, доктор фізико-математичних наук, професорка, Дніпровський національний університет імені Олеся Гончара, huk_n@fpm.dnu.edu.ua, ORCID ID: 0000-0001-7937-1039

Мітіков Микола Юрійович – аспірант кафедри прикладної математики, Дніпровський національний університет імені Олеся Гончара, mitikov.m22@fpm.dnu.edu.ua, ORCID ID: 0009-0002-1297-5676

Huk Nataliia - Head of the Department of Computer Technologies, Doctor of Physical and Mathematical Sciences, Professor, Oles Honchar Dnipro National University, huk_n@fpm.dnu.edu.ua, ORCID ID: 0000-0001-7937-1039

Mitikov Nikolay - PhD student, Department of Applied Mathematics, Oles Honchar Dnipro National University, mitikov.m22@fpm.dnu.edu.ua, ORCID ID: 0009-0002-1297-5676

Д.А. Миргородський, Т.В. Селівьорстова

ОСОБЛИВОСТІ МОДЕЛЮВАННЯ СТАЛЕПЛАВИЛЬНИХ ПРОЦЕСІВ У РЕАЛЬНОМУ ЧАСІ ЗА ДОПОМОГОЮ UNREAL ENGINE 5

Анотація. У статті розглядаються особливості моделювання сталеплавильних процесів у реальному часі за допомогою Unreal Engine 5. Аналізується доцільність використання віртуальних симуляцій та інтерактивних технологій для візуалізації та оптимізації технологічних процесів машин безперервного лиття заготовок (МБЛЗ). Описано підхід до створення 3D-моделей обладнання, впровадження фізичних симуляцій процесів охолодження та твердіння сталі, а також інтеграції гейміфікованих навчальних сценаріїв у VR-середовищі. Визначено переваги використання реалістичних графічних рушіїв для навчання та підготовки спеціалістів у металургійній галузі.

Ключові слова: Unreal Engine 5, моделювання, сталеплавильні процеси, віртуальна реальність, МБЛЗ, симуляція в реальному часі, гейміфікація.

Вступ. Сучасна металургійна промисловість вимагає високого рівня підготовки фахівців, здатних працювати з автоматизованими системами та складними технологічними процесами, зокрема з машинами безперервного лиття заготовок (МБЛЗ). Традиційні методи навчання, які базуються на теоретичних знаннях і лабораторних роботах, часто є недостатньо ефективними для формування практичних навичок роботи в реальному виробничому середовищі.

Розвиток технологій віртуальної та доповненої реальності (VR/AR), а також використання ігрових рушіїв, таких як Unreal Engine 5, відкривають нові можливості для інтерактивного навчання. Використання графічних інтерактивних ігрових додатків у навчальному процесі дозволяє:

- Моделювати складні технологічні процеси без ризику пошкодження обладнання чи загрози безпеці студентів.
- Створювати реалістичні сценарії виробничих завдань, що сприяють розвитку критичного мислення та вміння приймати рішення.
- Підвищити мотивацію студентів через використання гейміфікації та інтерактивних механік навчання.
- Забезпечити доступ до навчальних матеріалів у будь-який час через інтеграцію з онлайн-платформами та VR-симуляторами.

Крім того, застосування таких технологій дозволяє скоротити витрати на навчання, оскільки зменшується потреба у використанні реального обладнання для початкової підготовки. У контексті цифрової трансформації освіти, розробка і впровадження ін-

терактивних симуляторів є стратегічно важливим напрямком, що сприяє підготовці висококваліфікованих кадрів для металургійної галузі.

Таким чином, дослідження особливостей розробки графічного інтерактивного ігрового додатку для ознайомлення з технологією МБЛЗ є актуальним як з точки зору освіти, так і для підвищення ефективності виробничих процесів у промисловості.

Мета роботи – розробка графічного інтерактивного ігрового додатку для візуалізації та вивчення технологічного процесу машини безперервного лиття заготовок (МБЛЗ) з метою покращення підготовки фахівців у металургійній галузі та підвищення ефективності освітнього процесу.

Огляд існуючих підходів до візуалізації технологічних процесів. Візуалізація технологічних процесів відіграє важливу роль у навчанні та професійній підготовці спеціалістів у різних галузях промисловості. Вона дозволяє студентам та інженерам краще розуміти складні технічні процеси, підвищує ефективність навчання та мінімізує ризики помилок у реальних виробничих умовах.

Існує кілька основних підходів до візуалізації технологічних процесів, які використовуються у навчальних програмах та промислових симуляціях (таблиця 1).

Таблиця 1

Методи візуалізації технологічних процесів:

Метод візуалізації	Опис	Переваги	Недоліки	Приклади використання
Статичні схеми та графіки	Діаграми, креслення, схеми, що пояснюють процеси.	✓ Простота реалізації ✓ Висока доступність	- Відсутність інтерактивності - Низька реалістичність	Навчальні посібники, технічна документація
Відео та анімації	Відеозаписи або комп'ютерні анімовані моделі технологічних процесів.	✓ Наглядність ✓ Динамічне відображення процесу	- Відсутність інтерактивності - Неможливість змінювати параметри в реальному часі	Освітні відео, тренінги для працівників
2D-графічні симуляції	Візуалізація процесів у вигляді двовимірних моделей та схем.	✓ Відносно легка розробка ✓ Може використовуватися у веб-додатках	- Низька деталізація - Менш ефективне занурення	Інтерактивні навчальні платформи, технічні симулятори
3D-модельовання	Реалістичні 3D-моделі технологічного обладнання та процесів.	✓ Висока наочність ✓ Можливість масштабування та деталізації	- Обмежена інтерактивність - Високі вимоги до апаратного забезпе-	AutoCAD, SolidWorks, Simulink

			чення	
VR-симуляції (віртуальна реальність)	Повне занурення у віртуальне середовище для навчання та тренувань.	✓ Висока реалістичність ✓ Інтерактивність та моделювання сценаріїв	- Вимагає потужного обладнання - Висока вартість розробки	Unreal Engine, Unity, OpenVR
AR-симуляції (доповнена реальність)	Накладення цифрової інформації на реальний світ через мобільні пристрої або окуляри AR.	✓ Динамічний інтерактивний досвід ✓ Можливість використання на мобільних пристроях	- Обмежене поле зору - Не завжди точно передає фізичні процеси	AR-програми для навчання, промислові AR-інструменти
Гейміфіковані навчальні симулятори	Інтерактивні ігрові моделі технологічних процесів із можливістю управління параметрами.	✓ Висока залученість ✓ Можливість відпрацювання навичок у безпечному середовищі	- Високі витрати на розробку - Потребує глибокої інтеграції з навчальними програмами	Виробничі тренажери, навчальні VR-ігри

Сучасні методи візуалізації технологічних процесів значно відрізняються за рівнем реалістичності, інтерактивності та складності розробки. У традиційних підходах, таких як статичні схеми та відеоматеріали, основний акцент робиться на доступність та простоту використання, проте вони не забезпечують можливості інтерактивної взаємодії та змінних сценаріїв.

Більш просунуті методи, зокрема 3D-моделювання та цифрові симуляції, дозволяють створювати детальні візуальні представлення процесів, проте вимагають спеціалізованого програмного забезпечення та можуть бути обмежені у взаємодії з користувачем.

Найбільш реалістичні та ефективні підходи, такі як VR/AR-симуляції та гейміфіковані навчальні додатки, забезпечують повне занурення у віртуальне середовище, дозволяючи моделювати складні технологічні процеси та взаємодіяти з ними в режимі реального часу. Проте ці методи потребують значних ресурсів для розробки та потужного апаратного забезпечення.

Таким чином, вибір підходу до візуалізації залежить від цілей навчання:

– Для базового ознайомлення з процесами – достатньо використання схем, відео та 2D-анімацій.

– Для більш глибокого вивчення та інтерактивного навчання – доцільно застосовувати 3D-моделі та цифрові симуляції.

– Для моделювання виробничих ситуацій та практичного відпрацювання навичок – найбільш ефективними є VR/AR-симуляції та гейміфіковані додатки.

У таблиці 2 наведено порівняльний аналіз основних методів візуалізації за ключовими параметрами.

Таблиця 2

Порівняльний аналіз підходів до візуалізації

Підхід	Реалістичність	Інтерактивність	Складність розробки	Доступність
Статичні схеми	Низька	Відсутня	Низька	Висока
Відео та анімації	Середня	Відсутня	Середня	Висока
2D/3D-моделювання	Висока	Обмежена	Висока	Середня
VR/AR-симуляції	Дуже висока	Висока	Дуже висока	Обмежена (вимагає обладнання)
Гейміфіковані симулятори	Дуже висока	Висока	Висока	Середня

Таким чином, традиційні методи (схеми, відео, 3D-моделювання) підходять для базового розуміння технологічних процесів, але не забезпечують повноцінної інтерактивності. VR/AR-симуляції та гейміфіковані додатки на основі Unreal Engine 5 дозволяють максимально наблизити навчальний процес до реальних виробничих умов. Оптимальний підхід – це поєднання інтерактивних VR-симуляцій з гейміфікацією, що дозволить ефективно навчати фахівців у сфері металургії та безпечного управління МБЛЗ.

Огляд технологічного процесу МБЛЗ. Машина безперервного лиття заготовок (МБЛЗ) – це високотехнологічний комплекс обладнання, який використовується для перетворення рідкої сталі у затверділі заготовки (сляби, блюми, заготовки різних форм). Безперервне лиття сталі дозволяє підвищити ефективність виробництва, знизити втрати металу та покращити якість кінцевого продукту.

Слід зазначити, що процес безперервного лиття сталі є складною, багатоступеневою технологією, що включає виплавку, підготовку та затвердіння сталі. Використання МБЛЗ дозволяє суттєво знизити витрати, покращити якість заготовок та підвищити продуктивність виробництва.

Ключові особливості процесу:

- висока швидкість виробництва безперервним потоком,
- мінімізація дефектів за рахунок контрольованого охолодження,
- можливість швидкої зміни складу сталі завдяки вторинній металургії,
- економія матеріалів та зменшення відходів у порівнянні з традиційним розливом у форми.

Розуміння цих етапів є критично важливим для підготовки інженерів та операторів металургійного виробництва, а їх візуалізація у VR-середовищі дозволяє значно підвищити ефективність навчання та контролю якості процесу.

Вибір технологічного стека для розробки додатку. Розробка графічного інтерактивного ігрового додатку для ознайомлення з машиною безперервного лиття заготовок (МБЛЗ) вимагає ретельного підбору технологічного стека (таблиця 3). Основними критеріями вибору є реалістичність графіки, інтерактивність, продуктивність, можливість роботи у VR-середовищі та легкість у масштабуванні.

Таблиця 3

Технологічний стек для розробки інтерактивного додатку

Категорія	Технологія	Призначення
Ігровий рушій	Unreal Engine 5	Розробка інтерактивного середовища, фізичних симуляцій, підтримка VR/AR
3D-модельовання	Blender	Створення 3D-моделей обладнання та середовища
	Houdini	Процедурне модельовання фізичних ефектів (розплавлений метал, охолодження)
Фізичні симуляції	Chaos Physics (UE5)	Реалістична взаємодія твердих тіл та динамічних об'єктів
	Nvidia PhysX	Модельовання фізичних процесів
	Niagara (UE5)	Симуляція частинок (дим, розплавлений метал, охолодження)
VR/AR-інтеграція	OpenXR	Кросплатформна підтримка VR/AR
	SteamVR	Підтримка гарнітур VR (HTC Vive, Oculus, Valve Index)
	Oculus SDK	Оптимізація для пристроїв Meta Quest 2/Pro
Програмування	Blueprints (UE5)	Візуальне створення логіки додатку
	C++ (UE5)	Реалізація складних фізичних симуляцій та обчислень
Користувацький інтерфейс (UI)	UMG (Unreal Motion Graphics)	Розробка графічного інтерфейсу користувача
	Slate UI Framework	Кастомізація UI та інтеграція складних інтерфейсних рішень
Аудіо	FMOD Studio	Інтеграція звуків, просторове аудіо
	Unreal Audio Engine	Вбудоване рішення для обробки звуку

Цей технологічний стек забезпечує реалістичну графіку, точне фізичне модельовання, підтримку VR/AR та ефективну інтерактивну взаємодію, що робить його оптимальним вибором для створення навчального симулятора процесу безперервного лиття сталі.

Модельовання технологічного середовища. Розробка інтерактивної 3D-моделі за допомогою платформи Unreal Engine дозволяє досягти найвищого рівня реалістич-

ності завдяки впровадженню сучасних технологій візуалізації. Серед них особливе місце займають Nanite та Lumen, які докорінно змінюють підхід до створення графіки та освітлення в ігровій та симуляційній індустрії.

Nanite – це інноваційна технологія, що дозволяє працювати з неоптимізованими 3D-активами без шкоди для продуктивності. Вона усуває необхідність створення низькополігональних моделей, карт нормалей та рівнів деталізації (LOD), зберігаючи при цьому кінематографічну якість зображення. Завдяки віртуалізованій геометрії та новим форматам сіток, Nanite підвищує продуктивність, забезпечуючи деталізацію на рівні пікселів та спрощуючи створення складних сцен.

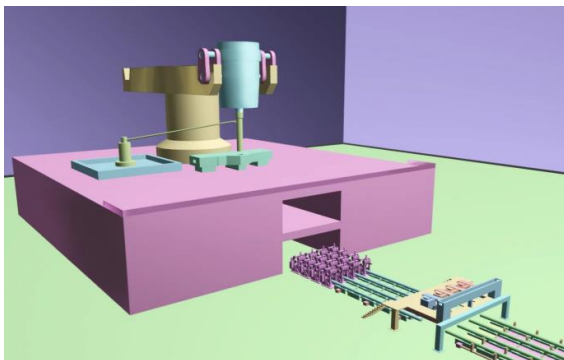
Lumen – це система динамічного глобального освітлення в реальному часі, яка імітує природну поведінку світла, враховуючи його відбиття від поверхонь. Вона створює реалістичне непряме освітлення та динамічні зміни умов освітлення. Система автоматично враховує інтенсивність джерел світла, перешкоди та їх вплив, забезпечуючи якісне освітлення у режимі реального часу.

Таким чином, інтеграція Nanite і Lumen в Unreal Engine 5 значно спрощує процес розробки, дозволяючи створювати вражаючі сцени з високою деталізацією та реалістичним освітленням, що відкриває нові можливості для інтерактивного дизайну та симуляцій [Помилка! Джерело посилання не знайдено.].

Головною метою розробки 3D-моделі МБЛЗ з інтеграцією до віртуального середовища UE5 є візуалізація промислових процесів або створення навчальних симуляцій. Визначившись із метою та об'єктом моделювання можемо поставити такі етапи розробки:

1. розробити деталізовану 3D-модель з необхідними деталями;
2. виконати текстурювання та UV-мапінгу кожного об'єкту;
3. оптимізувати кожний 3D об'єкт МБЛЗ для експорту в UE5;
4. інтегрувати готову модель в Unreal Engine 5 і налаштувати середовище;
5. оцінити результати та потенційну ефективність використання.

Моделювання технологічного середовища є одним із ключових етапів розробки інтерактивного додатку для візуалізації процесів машини безперервного лиття заготовок (МБЛЗ). Воно включає створення реалістичних 3D-моделей обладнання (рисунки 1, 2), налаштування фізичних симуляцій, а також розробку середовища взаємодії користувача з віртуальним простором.



(а)



(б)

Рисунок 1 – 3D-модель МБЛЗ, Blender 3D: а) без текстур, б) з текстурою

Вид сцени від першої особи показано на рисунку 2. Розробка графічного інтерактивного додатку на базі Unreal Engine 5 дозволяє створити реалістичну симуляцію процесу безперервного лиття сталі. Впровадження гейміфікації, VR-інтеграції та фізичних симуляцій забезпечує високу якість навчання та підвищує ефективність підготовки майбутніх фахівців у металургії.



Рисунок 2 – Вид від першої особи, анімація МБЛЗ

Інтерфейс користувача в Unreal Engine 5 також включає 2D-анімації та переходи між ними, що додають динамічності та наочності.

Висновки. У ході дослідження було розглянуто особливості моделювання сталеплавильних процесів у реальному часі з використанням Unreal Engine 5. Використання сучасних графічних технологій та віртуальної реальності у моделюванні технологічних процесів машини безперервного лиття заготовок (МБЛЗ) дозволяє значно підвищити ефективність навчання та оптимізації виробничих процесів.

Основні результати роботи:

1) аналіз сучасних методів візуалізації – розглянуто підходи до комп'ютерного моделювання металургійних процесів та виявлено переваги використання віртуальних симуляторів у навчанні та промисловості.

2) розробка 3D-моделей МБЛЗ – створено реалістичні моделі основних компонентів обладнання, таких як кристалізатор, проміжний ковш, роликовий стан, система вторинного охолодження.

3) впровадження фізичних симуляцій – використано технології Chaos Physics та Niagara в Unreal Engine 5 для моделювання поведінки розплавленої сталі, затвердіння заготовки та впливу температурних режимів.

4) гейміфікація та інтерактивність – розроблено механізми взаємодії користувача з віртуальним середовищем через VR-інтерфейс та систему керування параметрами технологічного процесу.

5) оцінка ефективності – обґрунтовано переваги впровадження VR-симуляцій у навчальних програмах для підготовки фахівців у галузі металургії.

Отримані результати підтверджують, що використання Unreal Engine 5 для віртуального моделювання сталеплавильних процесів дозволяє покращити якість навчання, безпеку та точність аналізу виробничих параметрів. Використання VR- та інтерактивних симуляцій сприяє ефективній підготовці майбутніх спеціалістів у сфері металургії та може бути корисним для автоматизації контролю процесів безперервного лиття сталі.

ЛІТЕРАТУРА / REFERENCES

1. Thomas, B. G. (2004). Continuous Casting. In McGraw-Hill Yearbook of Science and Technology (pp. 1-6). McGraw-Hill.
2. Thomas, B. G. (2002). Modeling of the continuous casting of steel—past, present, and future. Metallurgical and Materials Transactions B, 33(6), 795-812.
3. Jerald, J. (2015). The VR Book: Human-Centered Design for Virtual Reality. ACM Books.
4. Dede, C., Richards, J., & Saxberg, B. B. (2018). Digital Teaching Platforms: Customizing Classroom Learning for Each Student. Teachers College Press.
5. Epic Games. (2023). Unreal Engine 5 Documentation.
6. Shannon, T. (2022). Unreal Engine 5 for Beginners: A Comprehensive Guide to Game Development. Packt Publishing.

Received 20.12.2024.
Accepted 22.12.2024.

Features of real-time modeling of steelmaking processes using unreal engine 5

This paper explores the features of real-time modeling of steelmaking processes using Unreal Engine 5, focusing on the visualization and simulation of continuous casting machine (CCM) operations. The study highlights the advantages of interactive and virtual reality (VR) technologies in the training and optimization of metallurgical processes, providing a safer and more cost-effective alternative to traditional training methods.

A detailed approach to 3D modeling of CCM components is presented, including the tundish, mold, secondary cooling system, and roller sections, with a focus on their realistic visualization and optimization for real-time performance. The implementation of physical simulations using Chaos Physics and Niagara in Unreal Engine 5 allows for an accurate representation of molten steel flow, solidification dynamics, and temperature gradients. Additionally, gamification elements have been integrated to enhance user interaction, enabling students and engineers to explore various operational parameters in a controlled virtual environment.

The research also assesses the efficiency of digital simulation techniques compared to conventional learning methods in metallurgical training programs. The use of VR-based interactive learning modules improves engagement and comprehension by allowing users to manipulate casting parameters and observe the impact on the final product in real-time. The paper concludes that Unreal Engine 5 provides an effective platform for modeling complex

industrial processes, offering high-quality visual representation, dynamic interaction, and immersive training experiences for metallurgical professionals.

Future work will focus on enhancing AI-driven adaptive learning, expanding the simulation scope to include additional steelmaking stages, and integrating augmented reality (AR) tools to bridge the gap between virtual training and real-world applications.

Keywords: Unreal Engine 5, modeling, steelmaking processes, virtual reality, continuous casting machine (CCM), real-time simulation, gamification, interactive training, metallurgical process optimization.

Миргородський Данило Артемович – студент кафедри інформаційних технологій і систем, факультет прикладних комп'ютерних технологій, Український державний університет науки і технологій.

Селівьорстова Тетяна Віталіївна – кандидат технічних наук, доцент, доцент кафедри інформаційних технологій і систем, факультет прикладних комп'ютерних технологій, Український державний університет науки і технологій.

Myrhorodskyi Danylo – student of the Department of Information Technologies and Systems, Faculty of Applied Computer Technologies, Ukrainian State University of Science and Technology.

Selivorstova Tatyana – candidate of technical sciences, associate professor, associate professor department of Information Technologies and Systems, Faculty of Applied Computer Technologies at Ukrainian State University of Science and Technologies.

ЗМІСТ

CONTENTS

Куроп'ятник О.С. Багатокритеріальне моделювання в умовах неоднорідної невизначеності даних у роботах Скалозуба В.В.	3	Kuropiatnyk O. S. Multi-criteria modeling in the conditions of heterogeneous data uncertainty in the V.Skalozub papers	3
Плаксін С.В., Муха А.М., Устименко Д.В., Доскоч В.І., Сандул Т.А., Куліков С.М. Обґрунтування застосування розподіленої мережі фотоелектричних перетворювачів для живлення лінійного двигуна магнітолевітаційного транспорту	10	Plaksin S.V., Mukha A.M., Ustimenko D.V., Doskoch V.I., Sandul T.A., Kulikov S.M. Justification of the application of a distributed network of photoelectric converters to power a linear motor of magnetolevitation transport	10
Глушков О.В. Побудова дерева штейнера за допомогою метода кластеризації	20	Hlushkov O. Construction of a steiner tree using the clustering method	20
Зінченко М.Д., Потап О.Ю., Бурчак А.А. Експериментальне дослідження точності прокатки на стані 800	33	Zinchenko M.D., Potap O.Y., Burchak A.A. Experimental study of rolling accuracy on stand 800	33
Герасимчук П.В., Ліхоузова Т.А., Олійник Ю.О. Застосування методів машинного навчання для аналізу повідомлень в Twitter	43	Herasymchuk P.V., Likhouzova T.A., Oliinyk Y.O. Application of machine learning methods for analysis of Twitter messages	43
Остапець Д.О., Мотиленко В.А. Аналіз підходів до реалізації систем електронного голосування	50	Ostapets D., Motylenko V. Analysis of approaches of electronic voting systems implementation	50
Бур А.О., Ліхоузова Т. А., Олійник Ю. О. Прогнозування популярності музичних треків на платформі Spotify на основі числових метрик	61	Bur A.O., Likhouzova T.A., Oliinyk Y.O. Predicting the popularity of music tracks on Spotify based on numerical metrics	61

Вельгош А.С., Фургала Ю.М. Кореляційні властивості гістограм кольору при зниженні якості зображень	72	Velhosh A., Furgala Yu. Correlation properties of color histograms in the case of image quality decreasing	72
Яворська О.О., Новіков Л.А., Іконніков М.Ю. Зміна концентрації пилу у газопиловій хмарі при вибухових роботах у залізорудному кар'єрі	81	Yavorska O.O., Novikov L.A., Ikonnikov M.Yu. Changes in dust concentration in a gas-dust cloud during blasting in an iron ore pit	81
Інкін О.А., Білозьоров В.Є. Інтегроване нейромережеве моделювання еег для діагностики порушень мозкової діяльності	89	Inkin O.A., Belozyorov V.E. Integrated neuronetwork modeling of EEG for diagnostic disorders of brain activity	89
Панасенко Є.С., Білозьоров В.Є. Застосування спектрального аналізу даних ЕЕГ для ідентифікації станів відкритих і закритих очей	101	Panasenko Ye.S., Belozyorov V.Ye. The application of spectral analysis of EEG data for the identification of open and closed eye states	101
Земляний О.Д., Байбуз О.Г. Алгоритми імпутування пропусків у даних на основі ентропії	116	Zemlianyi O., Baibuz O. Algorithms for data imputation based on entropy	116
Довгаль А.В. Використання нітрид-алюмінієвих керамічних плат для забезпечення температурної стабільності вимірювальних підсилювачів	132	Dovhal A.V. The use of nitride-aluminum ceramic boards to ensure the temperature stability of measuring amplifiers	132
Шинкаренко В.І., Гуда А.І., Саблін О.І., Іванов О.П. Конструктивно-продукційне моделювання системи електропостачання тяги постійного струму	145	Shynkarenko V., Guda A., Sablin O., Ivanov O. Constructive-synthesizing modeling of the direct current traction power supply system	145
Захаров Д.В., Л.І. Книш Л.І. Математичне моделювання теплового режиму в сонячній термофотоелектричній панелі	159	Zakharov D.V., Knysh L.I. Mathematical modeling of the thermal regime in solar photovoltaic thermal panel	159

Малієнко О.О., Турчина В.А. Аналіз впливу списку пріоритетів виконання завдань на можливість уникнення аномалій у задачах упорядкування	167	Maliienko O., Turchyna V. Analysis of the impact of task prioritization lists on the potential for avoiding anomalies in task scheduling	167
Палій В.І. Прогнозування ефективності Agile-проектів за допомогою машинного навчання	175	Palii V.I. Forecasting the efficiency of Agile projects using machine learnin	175
Резніченко О.В., Ляшенко О.А., Архипова В.В. Дослідження моделей баз даних для зберігання big data у сфері моніторингу довкілля	185	Reznichenko O., Liashenko O., Arkhypova V. Research on database models for storing big data in the field of environmental monitoring	185
Турчина В.А., Коваленко Є.О. Умови зменшення довжини паралельних упорядкувань вершин спеціальних орграфів при наявності переривань	196	Turchyna V.A., Kovalenko Y.O. Conditions for length reduction of the parallel sequence of special digraphs' vertices in the presence of interruptions	196
Остапець Д.О., Сухомлин О.О. Аналіз існуючих підходів до формування функціональних профілів захищеності	208	Ostapets D., Sukhomlyn O. Analysis of existing approaches to the formation of functional security profiles	208
Поліський Ю.Д. Дослідження операції ділення на два в системі залишкових класів з усіма парними модулями	218	Polissky Yuriy Study of the division operation by two in the remainder class system with all paired modules	218
Михайленко В.С., Коренкова Г.В., Лещенко В.В., Мартинович Л.Я. Методика розробки нейронечітких систем автоматичного управління (САУ) з функцією ідентифікації параметрів судових паротурбінних установок (СПТУ) в процесі їх експлуатації	223	Mykhailenko V.S., Korenkova H., Leshchenko V., Martynovych L. Methodology for the Development of Neuro-Fuzzy Automatic Control Systems (ACS) with a Function for Identifying the Parameters of Marine Steam Turbine Installations (MSTI) during Their Operation	223

Гук Н.А., Мітіков М.Ю. Математична модель оптимізації пошуку дублікатів об'єктів типа string у знімках пам'яті	236	Huk N, Mitikov N. Mathematical optimisation model for searching duplicate string objects in the memory snapshot	236
Миргородський Д.А., Селіворстова Т.В. Особливості моделювання сталеплавильних процесів у реальному часі за допомогою Unreal Engine 5	250	Myrhorodskyi D., Selivorstova T. Features of real-time modeling of steelmaking processes using Unreal Engine 5	250

РЕФЕРАТИ

УДК 004.6:[656+ 007.5]+519.87

Куроп'ятник О. С. **Багатокритеріальне моделювання в умовах неоднорідної невизначеності даних у роботах Скалозуба В.В.** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.3 – 9.

В роботі виконано узагальнення підходів до вирішення завдань багатокритеріального моделювання в умовах неоднорідної невизначеності даних, у тому числі отриманих при моніторингу з нерівномірними та нечіткими інтервалами. Усі розглянуті підходи були створенні під керівництвом та за безпосередньої участі проф. Скалозуба В. В. Отримані результати можуть бути використані для подальшого розвитку запропонованих ідей та підходів.

Бібл. 11.

УДК 621.313.132

Плаксін С.В., Муха А.М., Устименко Д.В., Доскоч В.І., Сандул Т.А., Куліков С.М. **Обґрунтування застосування розподіленої мережі фотоелектричних перетворювачів для живлення лінійного двигуна магнітолевітаційного транспорту** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.10 – 19.

Сучасний високошвидкісний транспорт є основою сталого економічного та соціального розвитку держави, суспільства з дотримання вимог екологічності. Обґрунтована концепція енергозабезпечення лінійного двигуна магнітолевітаційного транспорту від розподіленої мережі фотоелектричних перетворювачів. Запропонований базовий силовий елемент шляхової енергетичної установки у вигляді завершеного вузла до складу якого входять сонячна панель, накопичувач та інвертор, що працює на навантаження у вигляді «короткої» шляхової котушки. Застосування «короткої» шляхової котушки дозволяє зменшити втрати електричної енергії, оскільки тягове зусилля формується лише в зоні взаємодії із рухомою одиницею, до незадіяних ділянок шляхової структури енергія не передається.

Бібл. 6.

УДК 004.421.2

Глушков О.В. **Побудова дерева Штейнера за допомогою метода кластеризації** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.20 – 32.

У роботі досліджується метод побудови дерева Штейнера для оптимізації мережевих структур у розподілених комп'ютерних системах. Впроваджено вдосконалений алгоритм знаходження точок Штейнера за допомогою методу кластеризації, що знижує обчислювальну складність та забезпечує точність. Дослідження проведено на графах з чотирма, п'ятьма та шістьма вершинами за допомогою програмного забезпечення на Python. Результати показали ефективність методу кластеризації для побудови дерева Штейнера, що може сприяти вдосконаленню методів оптимізації мережевих структур.

Бібл. 8, іл. 24.

УДК 681.532.1, 62-531.1

Зінченко М.Д., Потап О.Ю., Бурчак А.А. **Експериментальне дослідження точності прокатки на стані 800** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.33 – 42.

Виконано експериментальне дослідження процесу прокатки на стані 800 квадратної заготовки 100. При дослідженні процесу прокатки вимірювались розміри поперечного перерізу прокату, довжина розкатів після чистової кліті, струм двигуна головного привода чистової кліті. Визначені коливання розмірів прокату і вплив на розміри прокату змінення міжвалкового зазору в чистової кліті стана. Показано, що коливання довжини розкату, що призводять при порізці на пилах до появи штанг немірної довжини, обумовлені як коливаннями маси злитка, так і коливаннями розмірів прокату після прокатки на блюмингу і порізці прокатоного блюма на парогідравлічному ножі. Визначено, що змінення міжвалкового зазору можна застосовувати для регулювання довжини розкату в межах допуску на поперечні розміри.

Бібл. 2, іл.6, табл.3.

УДК 004.94

Герасимчук П.В., Ліхоузова Т.А., Олійник Ю.О. **Застосування методів машинного навчання для аналізу повідомлень в Twitter** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.43 – 49.

В роботі досліджено проблему бінарної класифікації текстових повідомлень на наявність булінгу. Підготовлено датасет з повідомленнями в Twitter та виконано його попередню обробку, зокрема очищення, токенизацію та лематизацію. Створено 3 набори вхідних даних для класифікаційних моделей: Bag-of-Words, матрицю TF-IDF, матрицю word2vec. Побудовано та протестовано моделі на основі різноманітних методів машинного навчання: логістичної регресії, методу k найближчих сусідів, випадкового лісу, методу опорних векторів, наївного баєсівського класифікатора на кожному із наборів вхідних даних.

За результатами тестування моделей виокремлено логістичну регресію на вхідних даних Bag-of-Words як найефективнішу модель для задачі бінарної класифікації текстових повідомлень із обраного набору.

Бібл. 7, іл. 14

УДК 004.056.5

Остапець Д.О., Мотиленко В.А. **Аналіз підходів до реалізації систем електронного голосування** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.50 – 60.

Розглядаються сучасні підходи для побудови систем електронного голосування, такі як блокчейн, а також криптографічні методи, що включають гомоморфне шифрування, докази нульового розголошення, та сліпий підпис. Виділено список вимог до систем електронного голосування, представлено узагальнені структури основних їх видів.

Проведено порівняльний аналіз видів систем електронного голосування на основі відповідності вимогам. Зроблено вибір підходу для подальшої побудови системи.

Бібл. 15, іл. 7., табл. 1.

УДК 004.94

Бур А.О., Ліхоузова Т. А., Олійник Ю. О. **Прогнозування популярності музичних треків на платформі Spotify на основі числових метрик** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.61 – 71.

В сучасному світі музика відіграє важливу роль у житті мільйонів людей, а платформи для стрімінгу музики, такі як Spotify, стали невід'ємною частиною сучасної культури. Популярність музичних треків має велике значення для музичної індустрії, впливаючи на доходи артистів та тренди в музичному світі. Прогнозування популярності музичних треків може допомогти артистам, продюсерам та платформам краще розуміти вподобання слухачів та оптимізувати свої стратегії.

Для прогнозування популярності музичних треків застосовувались моделі Decision Tree Classifier, KNeighbors Classifier, XGBoost Classifier та Random Forest Classifier. Загальний аналіз показав, що моделі XGBoost та Random Forest є найбільш ефективними для прогнозування популярності музичних треків. Вони демонструють високу точність і стійкість до змін у наборі атрибутів, що робить їх придатними для використання у реальних умовах.

Бібл. 7, іл. 13.

УДК 004.932

Вельгош А.С., Фургала Ю.М. **Кореляційні властивості гістограм кольору при зниженні якості зображень** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.72 – 80.

В роботі досліджено можливість використання гістограм кольору, сформованих у колірних моделях HS* для ідентифікації зображень при накладанні на них спотворень різної природи. Розглянуто спотворення, які можуть виникати в реальних умовах при отриманні зображень, зокрема, поворот, накладання шуму, розмиття і оптична дисторсія. Порівняння гістограм кольору оригінальних та спотворених зображень здійснювалося за допомогою обчислення коефіцієнту їх взаємної кореляції. Виявлено, що в межах того самого зображення кореляція є досить високою, надивлячись на накладені спотворення, що дозволяє допустити застосовність даного підходу для задач розпізнавання образів.

Бібл. 12.

УДК [622.271.3:622.235.43]:622.807

Яворська О.О., Новіков Л.А., Іконніков М.Ю. **Зміна концентрації пилу у газопиловій хмарі при вибухових роботах у залізорудному кар'єрі** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.81 – 88.

У статті встановлені діапазони зміни висоти підйому і середньомасової концентрації фракції пилу у газопиловій хмарі на динамічній стадії її формування після масово-

го вибуху у залізорудному кар'єрі. Наведено графічна залежність вмісту фракцій пилу від висоти їх підйому та отримано рівняння регресії.

Бібл. 10, іл. 2.

УДК 612.8:004.8

Інкін О.А., Білозьоров В.Є. **Інтегроване нейромережеве моделювання еег для діагностики порушень мозкової діяльності** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.89 – 100.

Розглянуто питання попередньої обробки ЕЕГ сигналу. Визначені пріоритети методів математичного моделювання для аналізу та прогнозування ознак мозкової активності. Запропоновано підхід використання LSTM моделі для прогнозування поведінки ЕЕГ з попередньою процедурою визначення параметрів вхідних ваг та активаційною функцією з хаотичною складовою.

Бібл. 11, іл. 10.

УДК 519.2:004.9

Панасенко Є.С., Білозьоров В.Є. **Застосування спектрального аналізу даних ЕЕГ для ідентифікації станів відкритих і закритих очей** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.101– 115.

У статті досліджується аналіз даних електроенцефалограм (ЕЕГ) для класифікації станів відкритих і закритих очей за допомогою швидкого перетворення Фур'є (ШПФ). Показано, що даний метод демонструє стабільну точність розпізнавання стану очей (открытые или закрытые) на рівні 70-80%, що підтверджує його ефективність у розв'язуванні задач класифікації біомедичних сигналів. Описано загальні відомості про ЕЕГ, точки для їх зчитування, зокрема про "систему 10-10", наведено інформацію про основні типи ритмів мозку. Також проаналізовано сучасні методи для аналізу ЕЕГ даних, з яких виділено 3 основних напрямлення: спектральний аналіз, рекурентний аналіз і методи машинного навчання. Розроблено програмне забезпечення для класифікації інформації, представленої у вигляді часових рядів ЕЕГ, отриманих у стані відкритих і закритих очей. Програма створена на мові Python з використанням бібліотеки PyRQA.

Бібл. 12, іл. 6, табл. 1.

УДК 004.42:519.25

Земляний О.Д., Байбуз О.Г. **Алгоритми імпутування пропусків у даних на основі ентропії** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.116– 131.

У цій роботі представлено два алгоритми імпутування даних, що ґрунтуються на мінімізації умовної ентропії та спрямовані на зменшення невизначеності в наборах даних з відсутніми значеннями. Перший алгоритм виконується за одну ітерацію, другий алгоритм передбачає можливість уточнення імпутацій ітераційно. Проведено три типи тестів, які показали, що алгоритми покращують результати класифікації порівняно з базовими моделями, проте вони досить повільні. Майбутні дослідження будуть присвячені оптимізації обчислень.

Бібл.15, іл.2, табл.7

УДК 620.22

Довгаль А.В. Використання нітрид-алюмінієвих керамічних плат для забезпечення температурної стабільності вимірювальних підсилювачів // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.132– 144.

Розкриваються принципи використання нітрид-алюмінієвих керамічних плат для забезпечення температурної стабільності вимірювальних підсилювачів. Описуються результати синтезу та технологія виготовлення кераміки із нітриду алюмінію. Запропоновано раціональний метод стабілізації параметрів для забезпечення спільного та однорідного температурного поля що буде впливати однаково на всі компоненти шляхом використання високо теплопровідних матеріалів.

Бібл. 18, іл. 10, табл. 1.

УДК 004.94+510.23+510.25+512.567+629.4

Шинкаренко В.І., Гуда А.І., Саблін О.І., Іванов О.П. Конструктивно-продукційне моделювання системи електропостачання тяги постійного струму // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.145– 158.

Розроблена загальна конструктивно-продукційна модель ділянки електропостачання тяги постійного струму. Модель може використовуватись для рішення низки задач, пов'язаних зі зменшенням електроспоживання як на залізничному так і міському громадському електротранспорті. Розроблена модель орієнтована на визначення наявності і номенклатури обладнання тягових підстанцій та раціонального використання енергії рекуперації. Однак може бути застосована і для вирішення інших задач, пов'язаних з конструкціями системи тягового електропостачання. Наведено приклад сформованої схеми лінійної ділянки електропостачання з трьома підстанціями.

Бібл. 12, іл. 9, табл 1.

УДК 519.6, 621.311.243

Захаров Д.В., Л.І. Книш Л.І. Математичне моделювання теплового режиму в сонячній термофотоелектричній панелі // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.159– 166.

В роботі запропоновано дизайн комбінованої сонячної термофотоелектричної панелі, в якій одночасно генерується електрична та тепла енергія. Тепловий режим в шарах термофотоелектричної панелі досліджувався на основі розробленої нестационарної нелінійної математичної моделі та створеного програмного додатку. В результаті комп'ютерного моделювання доведені перспективи використання термофотоелектричної панелі відповідного дизайну для термічної стабілізації сонячних елементів та генерації низькопотенційної теплоти.

Бібл. 13, іл.3.

УДК 519.8:519.178

Малієнко О.О., Турчина В.А. Аналіз впливу списку пріоритетів виконання завдань на можливість уникнення аномалій у задачах упорядкування // Системні те-

хнології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.167– 174.

У даній роботі розглянуто актуальні проблеми, пов'язані з аномальним погіршенням значень цільової функції при спробах покращення початкових параметрів в одній із задач дискретної оптимізації. Основна увага приділена дослідженню умов, за яких можливе запобігання виникненню таких аномалій. Розглядаються сучасні наукові роботи, присвячені оптимізації розкладів і управлінню пріоритетами завдань, зокрема для задач розподілу та розміщення, що виникають в галузях комп'ютерних наук, інженерії та операційних досліджень.

Бібл. 5.

УДК 004.8:519.85

Палій В.І. **Прогнозування ефективності Agile-проектів за допомогою машинного навчання** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.175– 184.

Дослідження спрямоване на вирішення актуальних завдань у розвитку інформаційних систем і інтеграцію штучного інтелекту для оптимізації методологій Agile. Розглядається моделювання та прогнозування ефективності Agile-процесів враховуючи змінні вимоги та випадкові відхилення. Порівнюються різні моделі машинного навчання: SVM, Random Forest, CNN, LSTM. Встановлено, що нейронні мережі, зокрема LSTM та CNN, є найбільш ефективними для прогнозування успішності проектів.

Бібл. 12, іл. 3, табл. 1.

УДК 004.65+502.3/.7

Резніченко О.В., Ляшенко О.А., Архипова В.В. **Дослідження моделей баз даних для зберігання big data у сфері моніторингу довкілля** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.185– 195.

Питання, пов'язані з утворенням великих масивів даних, до кінця не вирішені. Кількість інформації у світі постійно збільшується, що призвело до виникнення проблеми її збереження. Створене для визначення цих даних поняття «big data» включає у собі наступні характеристики: кількість, швидкість обробки, різноманітність, достовірність, змінність та цінність. До цього типу інформації відносяться характеристики навколишнього середовища; дані визначають розподіл відповідних показників на Землі і дають можливість дати прогноз на майбутнє щодо зміни їх у часі і просторі, що є важливим для ведення господарства та сталого розвитку людства. Але інформація щодо ефективної організації збереження і обробки таких даних є недостатньою, що вимагає проведення додаткових досліджень.

Бібл. 16.

УДК 519.8:519.176

Турчина В.А., Коваленко Є.О. **Умови зменшення довжини паралельних упорядкувань вершин спеціальних орграфів при наявності переривань** // Системні тех-

нології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.196– 207.

В роботі проаналізовано випадки впливу переривань на значення цільової функції для однієї задачі паралельного упорядкування вершин орієнтованого графа спеціальної структури. Досліджено можливі способи побудови самого графа, які базуються на застосуванні комбінаторних конфігурацій. Розглянуто методи розподілення вершин в упорядкуванні. Проаналізовано ефективність переривань у випадках, коли вони дозволені, та одержані необхідні умови, за яких значення цільової функції може бути зменшене за їх рахунок. Основні результати сформульовано у вигляді теореми.

Бібл. 8, іл. 6, табл. 0.

УДК 004.056.5

Остапець Д.О., Сухомлин О.О. **Аналіз існуючих підходів до формування функціональних профілів захищеності** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.208– 217.

У роботі аналізуються існуючі підходи до формування функціональних профілів захищеності (ФПЗ) для інформаційних систем. Автори досліджують ключові характеристики різних методик, зокрема їх складність, часові витрати та залежність від кваліфікації експерта. Робота порівнює переваги та недоліки методик і пропонує рекомендації для підвищення ефективності формування ФПЗ шляхом використання автоматизованих систем та інтелектуальних методів.

Бібл. 11, табл. 1

УДК 681.04

Поліський Ю.Д. **Дослідження операції ділення на два в системі залишкових класів з усіма парними модулями** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.218– 222.

Система всіх парних модулів, кожен з яких не є співмножником жодного з інших модулів цієї системи, побудована на основі системи взаємно простих модулів базису шляхом множення кожного модуля базису на парне число - коефіцієнт переходу. Однією зі складних операцій у такій системі є ділення числа на два. Запропонований підхід до розв'язання задачі є наступним. Виконується ділення на два залишків за модулями системи. Складається модульне рівняння, за результатами якого визначаються два значення залишку для кожного модуля, розташованих у різних інтервалах чисел і маючих протилежні парності. Оскільки в парній системі модулів всі залишки або парні, або непарні, формуємо набір з усіх парних залишків і набір з усіх непарних залишків. При діленні на два числа переводяться в нижню половину інтервалу чисел, тому виконуємо порівняння цих наборів на менший з них. Запропонований підхід забезпечує бажаний результат, і є доцільним застосувати його як перспективний напрямок дослідження складних операцій у системі залишкових класів з усіма парними модулями.

Бібл. 6, табл. 1

УДК 629.12.018:681.518.3

Михайленко В.С., Коренкова Г.В., Лещенко В.В., Мартинович Л.Я. **Методика розробки нейронечітких систем автоматичного управління (САУ) з функцією іден-**

тифікації параметрів суднових паротурбінних установок (СПТУ) в процесі їх експлуатації // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.223– 235.

У статті запропоновано нову модель для оптимізації систем автоматичного управління судновими паротурбінними установками. Вона передбачає вирішення задачі без урахування зовнішніх збурень з наступним коригуванням параметрів для компенсації змін у процесі експлуатації. Модель базується на використанні нейронечітких систем, що дозволяє адаптивно ідентифікувати параметри турбін у реальному часі. Це підвищує точність управління та забезпечує стабільність роботи СПТУ в різних умовах експлуатації, що є критично важливим для надійності та ефективності турбін.

Бібл. 9, іл. 6, табл. 2.

УДК. 004.41:519.85

Гук Н.А., Мітіков М.Ю. **Математична модель оптимізації пошуку дублікатів об'єктів типу string у знімках пам'яті** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.236– 249.

Представлено розробку математичної моделі для процесу виявлення дублікатів об'єктів з властивістю незмінності типу String в знімку пам'яті. Проаналізовано промислові системи, які вимагають сотні ГБ оперативної пам'яті для роботи та містять мільйони об'єктів в оперативній пам'яті.

Бібл. 14.

УДК 004.9:669

Миргородський Д.А., Селівьорстова Т.В. **Особливості моделювання сталеплавильних процесів у реальному часі за допомогою Unreal Engine 5** // Системні технології. Регіональний міжвузівський збірник наукових праць. - Випуск 6(155). – Дніпро, 2024. – С.250– 258.

У статті розглядаються особливості моделювання сталеплавильних процесів у реальному часі за допомогою Unreal Engine 5. Аналізується доцільність використання віртуальних симуляцій та інтерактивних технологій для візуалізації та оптимізації технологічних процесів машин безперервного лиття заготовок (МБЛЗ). Описано підхід до створення 3D-моделей обладнання, впровадження фізичних симуляцій процесів охолодження та твердіння сталі, а також інтеграції гейміфікованих навчальних сценаріїв у VR-середовищі. Визначено переваги використання реалістичних графічних рушіїв для навчання та підготовки спеціалістів у металургійній галузі.

Unreal Engine 5, моделювання, сталеплавильні процеси, віртуальна реальність, МБЛЗ, симуляція в реальному часі, гейміфікація.

Бібл. 6.

UDC 004.6:[656+ 007.5]+519.87

Kuropiatnyk O. S. **Multi-criteria modeling in the conditions of heterogeneous data uncertainty in the V.Skalozub papers** // System technologies. N 6(155) - Dnipro, 2024.- P.3 – 9.

The work summarizes the approaches to solving multi-criteria modeling tasks in conditions of heterogeneous data uncertainty, including cases with data obtained during monitoring with uneven and unclear intervals. All the considered approaches were created under the guidance and with the direct participation of prof. Skalozub V. V. The obtained results can be used for further development of the proposed ideas and approaches.

Bible 11.

UDC 621.313.132

Plaksin S.V., Mukha A.M., Ustimenko D.V., Doskoch V.I., Sandul T.A., Kulikov S.M.

Justification of the application of a distributed network of photoelectric converters to power a linear motor of magnetolevitation transport // System technologies. N 6(155) - Dnipro, 2024.- P.10 – 19.

Modern high-speed transport is the basis of sustainable economic and social development of the state, society in compliance with environmental requirements. The concept of power supply of a linear motor of magneto-levitation transport from a distributed network of photovoltaic converters is substantiated. The basic power element of a track power plant is proposed in the form of a completed unit consisting of a solar panel, a storage device and an inverter, which operates on a load in the form of a "short" track coil. The use of a "short" track coil allows reducing electrical energy losses, since the traction force is formed only in the zone of interaction with the rolling stock, energy is not transmitted to unused sections of the track structure.

Bible 6.

UDC 004.421.2

Hlushkov O. **Construction of a steiner tree using the clustering method** // System technologies. N 6(155) - Dnipro, 2024.- P.20 – 32.

This study investigates the method of constructing a Steiner tree for optimizing network structures in distributed computer systems. An advanced algorithm for finding Steiner points using the clustering method was implemented, reducing computational complexity while ensuring accuracy. The research was conducted on graphs with four, five, and six vertices using software written in Python. Results demonstrated the effectiveness of the clustering method for constructing Steiner trees, which can contribute to the improvement of network structure optimization methods.

Ref. 8, Fig. 24.

UDC 681.532.1, 62-531.1

Zinchenko M.D., Potap O.Y., Burchak A.A. **Experimental study of rolling accuracy on stand 800** // System technologies. N 6(155) - Dnipro, 2024.- P.33 – 42.

An experimental study of the rolling process was carried out on rolling mill 800 of a square blank of 100. During the study of the rolling process, the dimensions of the cross-section of the rolled product, the length of the rolls after the finishing stand, and the motor

current of the main drive of the finishing stand were measured. Fluctuations in the dimensions of rolled products and the effect on the dimensions of rolled products of changes in the inter-roll gap in the finishing stand of the mill are determined. It is shown that the fluctuations in the roll length, which lead to the appearance of rods of excessive length during cutting on saws, are caused both by fluctuations in the mass of the ingot and by fluctuations in the dimensions of the rolled product after rolling on a bloom and cutting the rolled bloom on a steam-hydraulic knife. It was determined that changing the inter-roll gap can be used to adjust the roll length within the tolerance on transverse dimensions.

Bibl. 2, ill. 6, table. 3.

UDC 004.94

Herasymchuk P.V., Likhouzova T.A., Oliinyk Y.O. **Application of machine learning methods for analysis of Twitter messages** // System technologies. N 6(155) - Dnipro, 2024.- P.43 – 49.

The paper investigates the problem of binary classification of text messages for the presence of bullying. A dataset with Twitter messages was prepared and pre-processed, including cleaning, tokenization, and lemmatization. 3 sets of input data for classification models were created: Bag-of-Words, TF-IDF matrix, word2vec matrix. Models based on various machine learning methods were built and tested: logistic regression, k nearest neighbors method, random forest, support vector method, naive Bayesian classifier on each of the input data sets.

Based on the results of model testing, logistic regression on Bag-of-Words input data was singled out as the most effective model for the task of binary classification of text messages from the selected set.

Bibl. 7, fig. 14

UDC 004.056.5

Ostapets D., Motylenko V. **Analysis of approaches of electronic voting systems implementation** // System technologies. N 6(155) - Dnipro, 2024.- P.50 – 60.

The work considers modern approaches to building electronic voting systems, such as blockchain, as well as cryptographic methods that include homomorphic encryption, zero-knowledge proofs, and blind signatures. A list of requirements for electronic voting systems is highlighted, and generalized structures of their main types are presented. A comparative analysis of the types of electronic voting systems based on their compliance with the requirements is carried out. A choice of approach for further system development is made.

Bibl. 15, il. 7., tab. 1.

UDC 004.94

Bur A.O., Likhouzova T.A., Oliinyk Y.O. **Predicting the popularity of music tracks on Spotify based on numerical metrics** // System technologies. N 6(155) - Dnipro, 2024.- P.61 – 71.

In today's world, music plays an important role in the lives of millions of people, and music streaming platforms such as Spotify have become an integral part of modern culture. The popularity of music tracks is of great importance to the music industry, affecting artists'

incomes and trends in the music world. Predicting the popularity of music tracks is an important task that can help artists, producers, and platforms better understand listener preferences and optimize their strategies.

Decision Tree Classifier, KNeighbors Classifier, XGBoost Classifier and Random Forest Classifier models were used to predict the popularity of music tracks. The overall analysis showed that the XGBoost and Random Forest models are the most effective for predicting the popularity of music tracks. They demonstrate high accuracy and resistance to changes in the set of attributes, which makes them suitable for use in real conditions.

Bibl. 7, fig. 13.

UDC 004.932

Velhosh A., Furgala Yu. **Correlation properties of color histograms in the case of image quality decreasing** // System technologies. N 6(155) - Dnipro, 2024.- P.72 – 80.

In image recognition tasks, objects are identified by examining various features such as texture, color, contour detection, and statistical or semantic descriptions. One widely used approach for extracting image attributes is the analysis of intensity histograms. While the traditional RGB color model is commonly used in digital image processing, it is often more effective to analyze color properties in HS* systems (such as HSL, HSV, and HSI) since these systems more closely resemble the spectral representation of color.

Bibl. 12.

UDS [622.271.3:622.235.43]:622.807

Yavorska O.O., Novikov L.A., Ikonnikov M.Yu. **Changes in dust concentration in a gas-dust cloud during blasting in an iron ore pit** // System technologies. N 6(155) - Dnipro, 2024.- P.81 – 88.

The paper establishes the ranges of changes in the height of ascent and the average mass concentration of dust fractions in the gas-dust cloud at the dynamic stage of its formation after a powerful explosion in an iron ore open pit. The graphical dependence of the dust fractions content on the height of its ascent is presented and the regression equation is obtained.

Bibl. 10, ill. 2.

UDC 612.8:004.8

Inkin O.A., Belozyorov V.E. **Integrated neuronetwork modeling of EEG for diagnostic disorders of brain activity** // System technologies. N 6(155) - Dnipro, 2024.- P.89 – 100.

The issue of pre-processing of the EEG signal is considered. The advantages of mathematical modeling methods for analyzing and predicting signs of brain activity are determined. An approach of using the LSTM model for predicting EEG behavior with a preliminary procedure for determining the parameters of the input weights and an activation function with a chaotic component is proposed.

Bible 11, ill. 10.

UDC 519.2:004.9

Panasenko Ye.S., Belozyorov V.Ye. **The application of spectral analysis of EEG data for the identification of open and closed eye states** // System technologies. N 6(155) - Dnipro, 2024.- P.101 – 115.

The article examines the analysis of electroencephalogram (EEG) data for the classification of open and closed eye states using the Fast Fourier Transform (FFT). It is shown that this method demonstrates stable recognition accuracy at the level of 70-80% in distinguishing between open- and closed-eye states, demonstrating its effectiveness in classifying biomedical signals. General information about EEG is described, points for their reading, in particular about the “10-10 system”, information about the main types of brain rhythms is given. Modern methods for analyzing EEG data were also reviewed, highlighting three main approaches: spectral analysis, recurrence analysis, and machine learning methods. Software was developed for classification of information presented in the form of EEG time series obtained in the state of open and closed eyes. The software was developed in Python utilizing the PyRQA library.

Bible 12, ill.6, tab. 1.

UDC 004.42:519.25

Zemlianyi O., Baibuz O. **Algorithms for data imputation based on entropy** // System technologies. N 6(155) - Dnipro, 2024.- P.116 – 131.

This paper presents two data imputation algorithms, one based on conditional entropy minimisation, and one aimed at reducing uncertainty in datasets with missing values. The first algorithm is executed in a single iteration, while the second algorithm assumes that the imputation can be refined iteratively. Three types of tests have been carried out, showing that the algorithms improve classification results over the baseline models, but they are quite slow. Future research will focus on computational optimisation.

Bibl. 15, pict. 2, tabl. 7

UDC 620.22

Dovhal A.V. **The use of nitride-aluminum ceramic boards to ensure the temperature stability of measuring amplifiers** // System technologies. N 6(155) - Dnipro, 2024.- P.132 – 144.

The principles of using nitride-aluminum ceramic boards to ensure the temperature stability of measuring amplifiers are disclosed. The results of the synthesis and the manufacturing technology of aluminum nitride ceramics are described. A rational method of parameter stabilization is proposed to ensure a common and uniform temperature field that will affect all components equally through the use of highly heat-conducting materials.

Ref. 18, fig. 10, tab. 1.

УДК 004.94+510.23+510.25+512.567+629.4

Shynkarenko V., Guda A., Sablin O., Ivanov O. **Constructive-synthesizing modeling of the direct current traction power supply system** // System technologies. N 6(155) - Dnipro, 2024.- P.145 – 158.

A general constructive-synthesizing model of the DC traction power supply section has been developed. The model can be used to solve a number of problems related to reducing electricity consumption in both railway and urban public electric transport. The developed model is focused on determining the availability and nomenclature of traction substation equipment and rational use of recuperation energy. However, it can also be applied to solve other problems related to the designs of the traction power supply system. An example of the

formed scheme of a linear power supply section with three substations is given.

Bibl. 12, il. 9, tabl. 1.

UDK 519.6, 621.311.243

Zakharov D.V., Knysh L.I. **Mathematical modeling of the thermal regime in solar photovoltaic thermal panel** // System technologies. N 6(155) - Dnipro, 2024.- P.159 – 166.

Design of a combined solar photovoltaic thermal panel, in which electrical and thermal energy generates simultaneously, was proposed in this study. A thermal regime in layers of the photovoltaic thermal panel was investigated based on the developed non-stationary non-linear mathematical model and created software.

As a result of computer modeling, the prospects of using a thermophotovoltaic panel of the appropriate design for the generation of low-potential heat have been proven.

Bib. 13, il. 3.

UDC 519.8:519.178

Maliienko O., Turchyna V. **Analysis of the impact of task prioritization lists on the potential for avoiding anomalies in task scheduling** // System technologies. N 6(155) - Dnipro, 2024.- P.167 – 174.

This paper addresses relevant issues related to the anomalous deterioration in objective function values when attempting to improve the initial parameters in one of the discrete optimization problems. The primary focus is on investigating the conditions under which it is possible to prevent the occurrence of such anomalies. Contemporary scientific works devoted to schedule optimization and task prioritization management, particularly for location-allocation problems arising in the fields of computer science, engineering, and operations research, are reviewed.

Bib. 5.

UDC 004.8:519.85

Palii V.I. **Forecasting the efficiency of Agile projects using machine learnin** // System technologies. N 6(155) - Dnipro, 2024.- P.175 – 184.

The study addresses current challenges in developing information systems and integrating artificial intelligence to optimize Agile methodologies. It examines modeling and forecasting the efficiency of Agile processes considering changing requirements and random deviations. Various machine learning models are compared: SVM, Random Forest, CNN, LSTM. It is established that neural networks, particularly LSTM and CNN, are the most effective for predicting project success.

Refs. 12, figs. 3, tables 1.

UDC 004.65+502.3/7

Reznichenko O., Liashenko O., Arkhypova V. **Database models for storing big data** // System technologies. N 6(155) - Dnipro, 2024.- P.185 – 195.

The issues related to the formation of large data collections are not fully resolved. The amount of information in the world is constantly increasing, which has resulted in the problem of its storage. The term "big data" created to define this data includes the following characteristics such as quantity, processing speed, variety, reliability, variability and value. This type of information includes environmental characteristics; the data determine the distribution

of relevant indicators on the Earth and make it possible to make a forecast for the future regarding their changes in time and space, which is important for economic management and sustainable development of humanity. However, there is not enough information on the effective organisation of the storage and processing of such data, and further research is needed. Thus, the object of the study is the data obtained at environmental monitoring stations.

Refs. 16.

UDC 519.8:519.176

Turchyna V.A., Kovalenko Y.O. **Conditions for length reduction of the parallel sequence of special digraphs' vertices in the presence of interruptions** // System technologies. N 6(155) - Dnipro, 2024.- P.196 – 207.

In the paper cases with the influence of interruption on the objective function value for one problem of parallel sequencing of a special structure digraph vertices are analyzed. Possible ways of constructing the graph itself, which are based on the use of combinatorial configurations, have been studied. The methods of vertices distribution in sequence are considered. The effectiveness of interruptions in cases where they are allowed and the necessary conditions under which the value of the objective function can be reduced as a result of their application are obtained. The main results are formulated in the form of a theorem.

Bible 8, fig. 6, tab. 0.

UDC 004.056.5

Ostapets D., Sukhomlyn O. **Analysis of existing approaches to the formation of functional security profiles** // System technologies. N 6(155) - Dnipro, 2024.- P.208 – 217.

The work analyzes existing approaches to the formation of functional security profiles (FSPs) for information systems. The authors examine key characteristics of various methodologies, including their complexity, time requirements, and dependence on expert qualifications. The paper compares the advantages and disadvantages of the methods and offers recommendations for improving the efficiency of FSP formation through the use of automated systems and intelligent methods.

Bibl. 11, tab. 1.

UDC 681.04

Polissky Yuriy **Study of the division operation by two in the remainder class system with all paired modules** // System technologies. N 6(155) - Dnipro, 2024.- P.218 – 222.

The system of all even modules, each of which is not a factor of any of the other modules of this system, is built on the basis of a system of relatively simple modules - the basis - by multiplying each basis module by an even number - the transition coefficient. One of the complex operations in such a system is dividing a number by two. The proposed approach to solving the problem is as follows. The remainder is divided into two by modules of the system. A modular equation is compiled, the results of which determine two remainder values for each module, located in different number intervals and having opposite parities. Since in an even system of modules all remainders are either even or odd, we form a set of all even remainders and a set of all odd remainders. Since, when divided by two, numbers are transferred to the lower half of the range of numbers, we compare these sets to the smaller of them.

The proposed approach provides the desired result, and it seems appropriate to apply it as a promising direction for studying complex operations in a system of residual classes with all even modules.

Bibl. 6, tab. 1.

UDC 629.12.018:681.518.3

Mykhailenko V.S., Korenkova Hanna, Leshchenko Valerii, Martynovych Larysa **Methodology for the Development of Neuro-Fuzzy Automatic Control Systems (ACS) with a Function for Identifying the Parameters of Marine Steam Turbine Installations (MSTI) during Their Operation** // System technologies. N 6(155) - Dnipro, 2024.- P.223 – 235.

The article proposes a new model for optimizing automatic control systems for marine steam turbine installations. It suggests solving the optimization problem without considering external disturbances, followed by parameter adjustments to compensate for changes during operation. The model is based on the use of neuro-fuzzy systems, which allows for adaptive identification of turbine parameters in real-time. This approach enhances control accuracy and ensures stable MSTI operation under various conditions, which is crucial for the reliability and efficiency of the turbines.

Ref. 9, fig. 6, tab. 2.

UDC 004.41:519.85

Huk N, Mitikov N. **Mathematical optimisation model for searching duplicate string objects in the memory snapshot** // System technologies. N 6(155) - Dnipro, 2024.- P.236 – 249.

The development of a mathematical model for the process of detecting duplicate objects with the immutability property of String type in a memory snapshot is presented. Industrial systems that require hundreds of GB of RAM to operate and contain millions of objects in RAM are analyzed.

Bibl. 14.

УДК 004.9:669

Myrhorodskiy D., Selivorstova T. **Features of real-time modeling of steelmaking processes using Unreal Engine 5** // System technologies. N 6(155) - Dnipro, 2024.- P.250 – 258.

This paper explores the features of real-time modeling of steelmaking processes using Unreal Engine 5. The study analyzes the feasibility of using virtual simulations and interactive technologies to visualize and optimize the technological processes of continuous casting machines (CCM). The approach to creating 3D models of equipment, implementing physical simulations of steel cooling and solidification processes, and integrating gamified training scenarios in a VR environment is described. The advantages of using realistic graphics engines for training and preparing specialists in the metallurgical industry are identified.

Keywords: Unreal Engine 5, modeling, steelmaking processes, virtual reality, continuous casting machine (CCM), real-time simulation, gamification.

Bibl. 6.