

С.Є. Малієнко, Т.В. Селівьорстова

ОГЛЯД МАТЕМАТИЧНИХ МОДЕЛЕЙ ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ БІЗНЕС АНАЛІЗУ ВЕЛИКИХ WEB-ДАНИХ

Анотація. У сучасному бізнесі, особливо у сфері інтернет технологій, існує величезна кількість даних, які постійно надходять та накопичуються. Для ухвалення ефективних рішень необхідно проводити аналіз цих даних. Однак, обробка великих обсягів даних потребує спеціальних математичних моделей та інформаційних технологій. У зв'язку з цим дослідження математичних моделей та інформаційних технологій для аналізу великих web даних є актуальною темою.

Проблема полягає в тому, що для аналізу великих обсягів даних, отриманих зі сфери інтернет технологій, потрібні ефективні методи обробки та аналізу. Існуючі методи аналізу не завжди дозволяють отримати актуальну та точну інформацію, яка може бути використана для прийняття рішень.

Метою дослідження є огляд та аналіз існуючих математичних моделей та інформаційних технологій, що використовуються для аналізу великих web даних. Для досягнення цієї мети були використані методи аналізу літератури, порівняльний аналіз методів та засобів аналізу даних.

В результаті дослідження були виявлені основні математичні моделі та інформаційні технології, які широко використовуються для аналізу великих web даних. Було проведено аналіз та порівняння існуючих методів, виявлено їх переваги та недоліки.

Аналіз даних є важливим інструментом прийняття ефективних рішень у сфері інтернет технологій. Використання ефективних математичних моделей та інформаційних технологій дозволяє отримати точну та актуальну інформацію з великих web даних. Результати дослідження можуть бути використані для розробки нових методів та засобів аналізу даних, що дозволить покращити якість прийнятих рішень.

Ключові слова: web, бізнес аналіз, нейронні мережі, кластеризація, регресійний аналіз, NLP, машинне навчання, e commerce.

В даний час великі веб-дані є основним джерелом інформації для бізнесу. Ці дані містять у собі величезну кількість інформації про поведінку споживачів, тенденції в галузях, маркетингових кампаніях та багато іншого. Але для того, щоб використовувати ці дані в бізнес-аналізі, потрібні математичні моделі та

інформаційні технології. У цій статті буде розглянуто огляд математичних моделей та інформаційних технологій бізнес-аналізу великих web-даних.

У сучасному світі величезна кількість даних генерується та зберігається в інтернеті. Великі компанії, у тому числі й ті, що займаються веб-дизайном, маркетингом та рекламою, використовують ці дані для ухвалення бізнес-рішень. Однак, обсяги цих даних стають настільки великими, що їх аналіз та обробка стають трудомісткими та дорогими. Крім того, деякі параметри, наприклад, відгуки клієнтів можуть бути неструктурованими, що додатково збільшує складність їх аналізу.

Аналіз останніх досліджень і публікацій. В останні роки було проведено безліч досліджень у галузі аналізу великих веб-даних. Деякі були присвячені аналізу тексту, наприклад, аналізу тональності тексту і класифікації тексту. У роботі "A Comparative Study on Sentiment Analysis Techniques" [1] було проведено огляд різних методів аналізу тональності тексту та проведено їх порівняння з урахуванням метрик якості. У роботі "A review of text mining techniques and applications" [2] були розглянуті різні методи та технології, що використовуються для аналізу тексту, включаючи тематичне моделювання, кластеризацію, аналіз настроїв та інші.

Інші дослідження були присвячені аналізу зображень та відео, наприклад, розпізнаванню об'єктів та класифікації зображень. У роботі "Image Classification with Deep Convolutional Neural Networks" [3] була проведена класифікація зображень з використанням глибоких нейронних мереж. У роботі "Video classification with deep convolutional neural networks" [4] була представлена модель глибокої нейронної мережі для класифікації відео.

Крім того, були проведені дослідження в галузі аналізу соціальних мереж та аналізу мережових структур. У роботі "Social media analytics: a survey of techniques, tools and platforms" [5] було проведено огляд різних методів та технологій, що використовуються для аналізу соціальних мереж. У роботі "Graph Analysis: An Overview" [6] було проведено огляд методів та технологій, що використовуються для аналізу графових структур.

Також існують дослідження безпосередньо аналізу великих даних:

У роботі "Big Data Analytics in E-commerce: A Review" [7] автори оглядово розглядають різні методи аналізу великих даних в e-commerce, включаючи аналіз асоціативних правил, кластеризацію, класифікацію, рекомендаційні системи тощо. Також автори відзначають важливість використання глибокого навчання та нейронних мереж для обробки та аналізу великих даних.

У роботі "Big Data Analytics in Supply Chain Management: A Review" [8] автори розглядають застосування аналітичних методів управління ланцюгами поставок. Автори відзначають, що в даний час більшість компаній мають величезну кількість даних, і використання методів аналізу великих даних може допомогти оптимізувати та покращити ефективність управління ланцюгами постачання.

У роботі "Data Mining for Business Applications" [9] автори представляють огляд різних методів та алгоритмів для аналізу даних у бізнесі. Зокрема, автори розглядають методи кластеризації, класифікації, регресії, пошуку асоціативних правил тощо. Крім того, автори наголошують на важливості правильного вибору методу аналізу даних, враховуючи специфіку конкретного бізнесу та цілей аналізу.

В цілому, всі ці дослідження та публікації показують, що існує безліч методів та технологій, які можуть бути використані для аналізу великих веб-даних. Кожен метод та технологія мають свої переваги та недоліки, і вибір найбільш відповідного методу залежить від конкретної задачі аналізу даних.

Метою даного дослідження є огляд математичних моделей та технологій бізнес-аналізу великих web-даних. В рамках даного огляду буде розглянуто декілька методів та технологій, таких як аналіз асоціативних правил, кластеризація, класифікація, регресія та нейронні мережі. Також буде проведено порівняння ефективності даних методів та технологій на прикладі реальних даних.

Викладення основного матеріалу дослідження. Машинне навчання - це область штучного інтелекту, де комп'ютери навчаються з урахуванням даних. Машинне навчання використовується для передбачення поведінки користувачів, виявлення патернів у даних та створення моделей машинного навчання для певних завдань. Одним із найпоширеніших методів машинного навчання є метод наївного Байеса.

Формула методу наївного Байеса:

$$P(A|B) = \frac{\{P(B|A) * P(A)\}}{\{P(B)\}}$$

Метод наївного Байеса використовується для класифікації даних. Він заснований на теоремі Байеса, яка говорить про те, як ймовірність того, що подія А станеться за наявності події В залежить від ймовірності події В за умови, що подія А відбудеться. Метод наївного Байеса дозволяє створювати моделі ма-

шинного навчання, що використовуються для класифікації текстів, фотографій, звукових файлів та інших типів даних.

Кластерний аналіз - це метод, який використовується для виявлення угруповань даних на основі їхньої подібності. Кластерний аналіз може використовуватися для виявлення груп користувачів, схожих з поведінці чи інтересам.

Одним із найпоширеніших методів кластерного аналізу є метод k-середніх.

Формула методу k-середніх:

$$V = \sum_{i=1}^k \sum_{x \in S_i} (x - \mu_i)^2$$

Метод k-середніх полягає в розбитті даних на k груп, де k - це задане число кількості груп, а потім обчислення центру мас для кожної групи. Дані потім перерозподіляються, щоб мінімізувати відстань між кожним елементом та центром мас групи. Цей процес повторюється, поки дані будуть розбиті на k груп.

Кластерний аналіз може бути використаний для аналізу поведінки користувачів на сайті, виявлення схожих груп клієнтів або виявлення схожих товарів чи послуг.

Регресійний аналіз - це метод, який використовується для визначення зв'язку між двома чи більше змінними. Регресійний аналіз може використовуватися для передбачення тенденцій даних і для визначення, які змінні можуть бути пов'язані з певним результатом.

Одним із найпоширеніших методів регресійного аналізу є метод лінійної регресії.

Формула методу лінійної регресії:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

Метод лінійної регресії полягає у визначенні лінійного зв'язку між залежною змінною Y і однією або декількома незалежними змінними X. Коефіцієнти використовуються для визначення сили зв'язку між змінними, а помилка ϵ використовується для вимірювання точності моделі.

Регресійний аналіз може бути використаний для передбачення продажів, визначення факторів, що впливають на утримання клієнтів, і визначення того, які маркетингові кампанії найбільш ефективні.

В таблиці 1 наведені результати порівняльного аналізу основних математичних методів аналізу даних.

Порівняння математичних методів та їх ефективності для аналізу даних

Метод	Опис	Переваги	Недоліки
Машинне навчання	Навчання комп'ютерів на основі даних	Висока точність передбачень, здатність обробляти великі обсяги даних	Вимагає великої кількості даних для навчання
Кластерний аналіз	Метод групування даних на основі подібності	Дозволяє виявляти приховані зв'язки між даними, може використовуватися для створення персоналізованих рекомендацій	Вимагає вибору кількості груп, може бути неефективним при високих обсягах даних
Регресійний аналіз	Визначення зв'язку між змінними	Дозволяє прогнозувати результати на основі даних, визначати вплив факторів на результат	Потребує точних даних, не враховує можливих невідомих факторів

Додаткові методи аналізу даних, які можуть бути використані для аналізу великих web-даних (таблиця 2), включають:

1. Аналіз часових рядів: Цей метод використовується для аналізу даних, що змінюються з часом, таких як дані про продаж, відвідування сайту і т.д. Він може допомогти виявити тренди та сезонні зміни, а також передбачити майбутні значення. Для цього методу можуть бути використані такі інструменти, як ARIMA (авторегресійна інтегрована ковзна модель) та Prophet.

2. Аналіз текстів: Цей метод використовується для аналізу текстової інформації, як-от відгуки клієнтів, коментарі на соціальних мережах і т.д. Він може допомогти виявити теми, настрої та думки. Для цього можуть бути використані такі інструменти, як Natural Language Processing (NLP) і Sentiment Analysis.

3. Аналіз соціальних мереж: Цей метод використовується для аналізу соціальних мереж, таких як Twitter, Facebook і т.д. Він може допомогти виявити тенденції, думки та взаємозв'язки між користувачами. Для цього можуть бути використані такі інструменти, як Social Network Analysis і Graph Theory.

Методи аналізу web-даних

Метод аналізу	Опис	Переваги	Недоліки
Аналіз часових рядів	Аналіз даних, що змінюються з часом	Допомагає виявити тренди та сезонні зміни, передбачити майбутні значення	Вимагає точних даних, може бути неефективним при великих обсягах даних
Аналіз текстів	Аналіз текстової інформації, такий як відгуки клієнтів	Допомагає виявити теми, настрої та думки	Точність аналізу може бути обмежена через різні варіанти вираження сенсу
Аналіз соціальних мереж	Аналіз соціальних мереж, таких як Twitter, Facebook	Допомагає виявити тенденції, думки та взаємозв'язки між користувачами	Не всі дані доступні для аналізу, може бути неефективним при великих обсягах даних

Висновки. У цій статті було проведено огляд математичних моделей та технологій, що використовуються для аналізу великих web-даних. Машинне навчання, кластерний аналіз та регресійний аналіз були розглянуті як основні методи.

Машинне навчання є гнучким та ефективним методом, який може використовуватися для створення різних моделей, у тому числі для класифікації та кластеризації даних. Кластерний аналіз дозволяє виявляти приховані зв'язки між даними, але може бути неефективним при великих обсягах даних. Регресійний аналіз дозволяє прогнозувати результати на основі даних, але потребує точних даних і не враховує можливих невідомих факторів.

Залежно від цілей бізнес-аналізу та доступних даних, вибір конкретного методу може змінюватись. Важливо розуміти, що кожен метод має свої переваги та недоліки, і правильний вибір методу може суттєво вплинути на результати аналізу.

Інформаційні технології, такі як Hadoop та Apache Spark, можуть бути використані для обробки великих обсягів даних та прискорення обчислювальних процесів. Крім того, інструменти для візуалізації даних, такі як Tableau та Power BI, можуть допомогти з легким розумінням результатів аналізу та прийняттям ефективних бізнес-рішень.

Загалом аналіз великих web-даних є важливим компонентом сучасної бізнес-стратегії. Правильний вибір математичних моделей та інформаційних технологій може допомогти компаніям приймати ефективні бізнес-рішення, оптимізувати процеси та збільшити прибуток.

ЛІТЕРАТУРА / REFERENCES

1. Goyal, P. & Kumar, V. (2017). A Comparative Study on Sentiment Analysis Techniques. In Proceedings of the 2nd International Conference on Telecommunications and Networks (TEL-NET) (pp. 73-76). IEEE.
2. Islam, M. R., Islam, M. M., Islam, M. N., Islam, M. R., & Rahman, M. A. (2020). A review of text mining techniques and applications. Journal of King Saud University-Computer and Information Sciences, 32 (3), 257-269.
3. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Image Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems (pp. 1097-1105).
4. Feichtenhofer, C., Pinz, A., & Zisserman, A. (2015). Video classification with deep convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4475-4483).
5. Salehan, M., & Kim, D. J. (2015). Social media analytics: A survey of techniques, tools and platforms. Telematics and Informatics, 32 (4), 575-591.
6. Al Hasan, M., Chaoji, V., Salem, S., & Zaki, M. (2016). Graph Analysis: An Overview. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 6 (1), 1-15.
7. Choudhury, P., & Shankar, R. (2018). Big Data Analytics in E-commerce: A Review. Journal of King Saud University-Computer and Information Sciences, 30 (4), 431-448.
8. Mishra, D., & Gunasekaran, A. (2018). Big Data Analytics in Supply Chain Management: A Review. Journal of Business Research, 83, 36-46.
9. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). Data Mining for Business Applications. AI Magazine, 17 (3), 37-55.

Received 24.04.2023.

Accepted 27.04.2023.

Review of mathematical models and information technologies for business analysis of the big web data

The article provides a comprehensive review of mathematical models and information technologies used for analyzing large amounts of data in web applications. The latest research and publications in the field are analyzed, including a comparative analysis of machine learning methods, text, image, video analysis, social network analysis, and graph algorithms. The goal of this research is to analyze the effectiveness and applicability of mathematical models and information technologies in business analysis of large

web data. The article presents the results of the research and a comparative analysis of the efficiency of methods, which will help business analysts choose the optimal tools for processing and analyzing large amounts of data in web applications.

The article begins with an overview of the problem and the latest research and publications in the field. The article provides a detailed description of various mathematical models and information technologies, including their strengths and weaknesses. A comparative analysis of these methods is presented, with a focus on their effectiveness and applicability in business analysis.

The article also provides a detailed description of the applications of mathematical models and information technologies in various industries, such as e-commerce and supply chain management. The article analyzes the challenges and opportunities associated with the use of these technologies in business analysis and provides recommendations for businesses that want to take advantage of these technologies.

Overall, the article provides a comprehensive overview of mathematical models and information technologies used in business analysis of large web data. The article is a valuable resource for business analysts, data scientists, and researchers who want to learn more about the latest developments in this field.

Keywords: mathematical models, information technologies, business analysis, web data, machine learning, social network analysis, graph algorithms, big data analytics

Малієнко Станіслав Євгенович – аспірант 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.

Maliienko Stanislav – postgraduate student of the Department of Information Technologies and Systems, Faculty of Applied Computer Technologies at Ukrainian State University of Science and Technologies.

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.