

МЕТОД ІНТЕЛЕКТУАЛЬНОГО МОНІТОРИНГУ ТА ОПТИМІЗАЦІЇ РЕСУРСІВ ІТ-ІНФРАСТРУКТУРИ НА ОСНОВІ МАШИННОГО НАВЧАННЯ

Самуськов О. Д.¹ [ORCID], Островська К. Ю.² [ORCID]

¹Український державний університет науки і технологій, аспірант, Україна

²Український державний університет науки і технологій, к.т.н., доцент, Україна

Анотація. У роботі проведено аналіз сучасних підходів до моніторингу та управління навантаженням в ІТ-інфраструктурах. Розглянуто обмеження традиційних реактивних методів та обґрунтовано доцільність переходу до інтелектуальних систем класу AIOps. Запропоновано концепцію методу моніторингу, що базується на використанні алгоритмів машинного навчання (зокрема моделей LSTM та нейро-нечітких мереж) для проактивного прогнозування часових рядів навантаження. Окрему увагу приділено інтелектуальному балансуванню ресурсів у хмарних та мікросервісних середовищах, що дозволяє мінімізувати затримки та оптимізувати витрати на інфраструктуру. Отримані результати можуть бути використані при розробці адаптивних систем автоматичного масштабування.

Ключові слова: ІТ-інфраструктура, моніторинг, машинне навчання, прогнозування навантаження, AIOps, балансування ресурсів, нейро-нечіткі системи.

Стрімкий розвиток хмарних технологій, мікросервісної архітектури та концепції Edge Computing висуває нові вимоги до надійності та продуктивності ІТ-інфраструктур [1, 8]. Сучасні корпоративні системи генерують величезні обсяги телеметрії, обробка якої в ручному режимі стає неможливою. Традиційні методи управління ресурсами переважно базуються на реактивному підході: система реагує на проблему (наприклад, перевантаження CPU або брак пам'яті) лише після перетину певного статичного порогу. Це призводить до часових затримок у наданні послуг та фінансових втрат через неефективне використання обчислювальних потужностей.

Традиційні інструменти моніторингу, такі як Prometheus, Zabbix та Nagios, фокусуються на зборі метрик у реальному часі та візуалізації поточного стану системи. Основною проблемою таких підходів є «алерт-фатиг» (надмірна

кількість сповіщень) та нездатність виявляти складні, неявні залежності між компонентами системи [2].

Вирішенням цих проблем є впровадження концепції AIOps (Artificial Intelligence for IT Operations) [3]. Використання методів машинного навчання дозволяє автоматизувати аналіз логів та метрик [3, 6]. Наприклад, алгоритми кластеризації, такі як K-means або DBSCAN, дозволяють виявляти аномалії в поведінці системи, які неможливо зафіксувати звичайними пороговими методами. Це забезпечує перехід від простого спостереження до глибокої аналітики (observability), що є критично важливим для розподілених середовищ [2].

Важливим аспектом сучасного моніторингу є кореляція подій. Замість сотень окремих сповіщень інтелектуальна система здатна ідентифікувати першопричину збою (Root Cause Analysis), аналізуючи взаємозв'язки між різними рівнями інфраструктури — від мережевого обладнання до прикладного програмного забезпечення.

Ефективна оптимізація IT-інфраструктури неможлива без точного прогнозування майбутніх потреб у ресурсах. Традиційно для цього використовувалися статистичні методи, такі як моделі авторегресії та інтегрованого рухомого середнього (ARIMA/SARIMA) [4]. Ці методи добре працюють із лінійними даними та чітко вираженою сезонністю, проте вони демонструють низьку точність у високодинамічних середовищах, де навантаження змінюється нелінійно та залежить від багатьох зовнішніх факторів.

На сучасному етапі перевага надається методам машинного навчання (ML) та глибинного навчання (DL) [5, 6]. Серед них варто виділити наступні підходи:

- Регресійні моделі (Random Forest, Gradient Boosting): дозволяють враховувати велику кількість вхідних ознак та виявляти нелінійні зв'язки, що значно підвищує точність у порівнянні зі статистичними методами [6].

- Рекурентні нейронні мережі (RNN): зокрема архітектура LSTM (Long Short-Term Memory), яка спеціально спроектована для роботи з часовими рядами [5]. LSTM здатна «запам'ятовувати» довгострокові залежності в даних,

що дозволяє прогнозувати пікові навантаження на сервери з урахуванням історичних трендів за тривалі періоди.

- Трансформери (Transformer-based models): останнє слово в обробці послідовностей, що забезпечує високу паралелізацію обчислень та ще вищу точність для надскладних систем.

Вибір конкретного підходу залежить від обчислювальних можливостей системи та вимог до точності. Порівняльний аналіз розглянутих методів представлено в табл. 1.

Таблиця 1

Порівняння методів інтелектуального аналізу метрик

Метод	Придатність для Real-time	Точність прогнозу	Складність реалізації
ARIMA	Висока	Середня	Низька
Random Forest	Середня	Висока	Середня
LSTM	Середня	Дуже висока	Висока
Fuzzy Logic	Висока	—	Середня

На основі наведених даних, для реалізації проактивного моніторингу в роботі обрано поєднання архітектури LSTM для прогнозування та методів нечіткої логіки для прийняття рішень, що забезпечує баланс між точністю та швидкістю адаптації інфраструктури [5, 6].

Прогнозування навантаження є лише першим етапом; наступним критично важливим кроком є прийняття рішень щодо перерозподілу ресурсів. Класичні алгоритми балансування, такі як Round Robin або Least Connections, працюють на основі спрощених правил і не враховують поточну «вагу» кожного завдання та динаміку зміни стану всієї інфраструктури [8].

Для вирішення завдань у хмарних та мікросервісних середовищах доцільно використовувати методи інтелектуального балансування:

- Нечітка логіка (Fuzzy Logic): дозволяє системі приймати рішення в умовах невизначеності, коли вхідні дані (наприклад, «завантаження CPU високе», «затримка мережі середня») не мають чітких меж. Це ідеально підходить для створення правил автоматичного масштабування (Autoscaling).

- Нейро-нечіткі системи (Neuro-Fuzzy Systems): поєднують у собі здатність нейронних мереж до навчання та прозорість нечітких правил. Такий гібридний підхід дозволяє системі постійно адаптуватися до нових типів навантажень, зберігаючи при цьому інтерпретованість логіки управління.

- Навчання з підкріпленням (Reinforcement Learning): агент системи балансування навчається методом спроб та помилок, отримуючи «винагороду» за стабільну роботу інфраструктури та мінімізацію витрат.

Використання таких підходів у поєднанні з сучасними платформами оркестрації, як Kubernetes, дозволяє реалізувати інтелектуальні контролери автоскейлінгу [7, 9]. Це забезпечує динамічне виділення ресурсів саме в той момент, коли це необхідно, що є ключовим фактором для забезпечення стабільної роботи систем з мікросервісною архітектурою в умовах непередбачуваного трафіку.

Висновки. Аналіз сучасних методів моніторингу та управління навантаженням підтверджує необхідність переходу від статичних реактивних підходів до інтелектуальних проактивних систем. Запропоноване поєднання методів глибинного навчання (LSTM) для прогнозування часових рядів та нейро-нечітких моделей для прийняття рішень дозволяє створити високоефективний метод оптимізації ІТ-інфраструктури [3, 9]. Впровадження таких рішень забезпечує підвищення надійності систем, зменшення часу відгуку та суттєву економію хмарних ресурсів.

ЛІТЕРАТУРА / REFERENCE

1. Designing Data-Intensive Applications / M. Kleppmann. — O'Reilly Media, 2017.
2. Observability Engineering / C. Majors et al. — O'Reilly, 2022.
3. AIOps: Real-World Use Cases and Platforms / S. Dang et al. — Springer, 2021.
4. Time Series Analysis: Forecasting and Control / G. Box, G. Jenkins. — Wiley, 2015.
5. Deep Learning / I. Goodfellow, Y. Bengio, A. Courville. — MIT Press, 2016.
6. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow / A. Géron. — O'Reilly, 2022.
7. Kubernetes: Up and Running / K. Burns, B. Grant. — O'Reilly, 2022.
8. Cloud Native Patterns / C. Richardson. — Manning, 2018.
9. An SLO-Driven and Cost-Aware Autoscaling Framework for Kubernetes / V. Punniyamoorthy, B. Kumar, S. Saha, L. Butra, M. Palanigounder, A. K. Agarwal, K. Kannan — arXiv, 2025. DOI: <https://doi.org/10.48550/arXiv.2512.23415>

INTELLIGENT METHOD FOR MONITORING AND RESOURCE OPTIMIZATION OF IT INFRASTRUCTURE BASED ON MACHINE LEARNING

Oleksandr Samuskov, Kateryna Ostrovska

Abstract. *The paper analyzes modern approaches to monitoring and load management in IT infrastructures. The limitations of traditional reactive methods are considered and the expediency of moving to intelligent AIOps systems is substantiated. The authors proposed a concept of a monitoring method based on the use of machine learning algorithms (in particular, LSTM models and neuro-fuzzy networks) for proactive forecasting of load time series. Particular attention is paid to intelligent resource balancing in cloud and microservice environments, which allows minimizing latency and optimizing infrastructure costs. The results can be used in the development of adaptive autoscaling systems.*

Keywords: *IT infrastructure, monitoring, machine learning, load forecasting, AIOps, resource balancing, neuro-fuzzy systems.*