

Володимир Паньків, Олександр Сторожук*, Наталія Гіссовська*

¹ НЛТУ України, Львів, Україна, ORCID: 0009-0006-2720-0864,
pankiv.volodymyr@nltu.lviv.ua

² НЛТУ України, Львів, Україна. ORCID: 0000-0001-6566-5271,
storozhuk@nltu.edu.ua

³ НЛТУ України, Львів, Україна. gissovskan@nltu.edu.ua

Інноваційні підходи до розробки алгоритмів машинного навчання для оброблення даних у реальному часі в системах Big Data

Анотація. У статті розглядаються сучасні інноваційні підходи до розробки алгоритмів машинного навчання для оброблення даних у реальному часі в системах Big Data. Проаналізовано основні проблеми масштабування, продуктивності та складності алгоритмів машинного навчання. Запропоновано нові методи та архітектури, які сприяють ефективній обробці даних у реальному часі, зокрема потокове навчання, розподілені обчислення та використання гібридних моделей. Результати дослідження можуть бути застосовані у сферах фінансів, телекомунікацій, охорони здоров'я та інших галузях, де важлива швидка та точна обробка великих обсягів інформації.

Ключові слова – Big Data, машинне навчання, обробка в реальному часі, алгоритми, масштабування, потокові дані, розподілені обчислення, глибоке навчання.

У сучасному інформаційному суспільстві обсяги даних зростають експоненційно, створюючи безпрецедентні виклики для традиційних методів оброблення та аналізу. Згідно з прогнозами, до 2025 року глобальний обсяг даних досягне 175 зетабайт [1]. Системи Big Data стали критично важливими для бізнесу, науки та урядових організацій, проте стандартні алгоритми машинного навчання часто не здатні ефективно працювати з такими обсягами інформації в реальному часі. Це зумовлює потребу в інноваційних підходах до розробки алгоритмів, які можуть забезпечити швидку, масштабовану та ефективну обробку даних.

Реальний час оброблення даних стає ключовим фактором у багатьох галузях, таких як фінанси, де рішення повинні прийматися миттєво, або охорона

здоров'я, де життєво важлива інформація потребує негайної оброблення. У зв'язку з цим виникає необхідність у розробці нових алгоритмів та архітектур, які здатні обробляти великі обсяги даних з мінімальною затримкою[2].

Основні проблеми та виклики. Однією з головних проблем при обробці великих обсягів даних у реальному часі є масштабованість алгоритмів машинного навчання. Традиційні алгоритми часто не пристосовані до роботи з поточними даними та не можуть забезпечити необхідну продуктивність. Окрім цього, складність даних, їх нерівномірність та гетерогенність додають додаткових труднощів. Забезпечення низької затримки, висока пропускна здатність та стійкість до відмов є критичними вимогами для систем реального часу.

Інша проблема полягає у необхідності адаптації алгоритмів до динамічно змінюваних даних. У багатьох випадках дані можуть змінюватися з часом, що вимагає від алгоритмів здатності до навчання "на льоту" без втрати точності.

Інноваційні підходи до розробки алгоритмів. Одним із перспективних напрямків є розвиток потокового машинного навчання, яке дає змогу обробляти дані в міру їх надходження. Алгоритми потокового навчання можуть постійно оновлювати свої моделі, адаптуючись до нових даних. Це особливо важливо в таких сферах, як соціальні мережі, де дані генеруються користувачами постійно і в великих обсягах.

Розподілені обчислення є ще одним ключовим компонентом у вирішенні проблем масштабування. Використання кластерів та хмарних обчислень дає змогу розподіляти обробку даних між багатьма вузлами, забезпечуючи високу продуктивність. Платформи такі як Apache Spark та Apache Flink стали де-факто стандартом для розподіленої оброблення великих обсягів даних.

Глибоке навчання з використанням нейронних мереж також відіграє важливу роль. Сучасні архітектури, такі як згорткові та рекурентні нейронні мережі, здатні ефективно працювати з великими наборами даних, а використання графічних процесорів (GPU) та тензорних процесорів (TPU) дає змогу прискорити обчислення[3].

Запропоновані рішення та методи. Для подолання викликів, що виникають під час оброблення великих обсягів даних у реальному часі, можна застосувати кілька сучасних методів, які демонструють високу ефективність в умовах динамічної роботи з інформацією. Наприклад, гібридні моделі навчання стають популярним рішенням завдяки поєднанню переваг батчового та потокового навчання. Спочатку модель навчається на повному наборі даних, що забезпечує її точність, а потім оновлюється в реальному часі за допомогою нових даних. Це особливо важливо для сфер, де необхідно реагувати на зміну параметрів системи швидко і без втрати якості.

У фінансовій сфері алгоритми машинного навчання для оброблення даних у реальному часі активно застосовуються у виявленні шахрайства під час оброблення платіжних транзакцій. Компанії, такі як Visa і Mastercard, використовують поточкові моделі, які аналізують кожну транзакцію на предмет підозрілої поведінки, враховуючи історичні дані та вхідні параметри в реальному часі. Це дає змогу не лише вчасно блокувати шахрайські операції, але й знижувати ризик хибних спрацювань.

В охороні здоров'я алгоритми оброблення даних у реальному часі широко застосовуються для моніторингу життєвих показників пацієнтів. Пристрої для безперервного спостереження за серцевим ритмом або рівнем кисню в крові використовують адаптивні алгоритми, що автоматично оновлюють параметри аналізу залежно від стану пацієнта. Такі системи забезпечують оперативну передачу даних медичному персоналу та негайну реакцію на відхилення від норми.

У сфері телекомунікацій можна відзначити впровадження розподілених алгоритмів машинного навчання для оптимізації роботи мереж. Технології компанії Ericsson дають змогу у реальному часі прогнозувати пікові навантаження на мережу та адаптувати її конфігурацію для забезпечення стабільного з'єднання. Це досягається за допомогою аналізу поточкових даних з базових станцій та мобільних пристроїв.

Розподілені алгоритми машинного навчання також мають значний потенціал, адже дають змогу ефективно обробляти дані паралельно на багатьох вузлах. Такий підхід забезпечує не лише високу швидкість оброблення, а й оптимізацію використання ресурсів, що є критично важливим у великих системах, як-от телекомунікаційні мережі чи хмарні платформи. З іншого боку, адаптивні алгоритми виявляються незамінними там, де дані можуть бути нерівномірними або гетерогенними. Ці моделі автоматично налаштовуються відповідно до змінних характеристик даних, що дає змогу зберігати ефективність оброблення навіть у непередбачуваних умовах.

Іншим інноваційним підходом є автоматизація процесів машинного навчання (AutoML), що значно скорочує час та зусилля, необхідні для розробки моделей. AutoML дає змогу швидко впроваджувати моделі у системи реального часу, що особливо важливо для задач, де затримки неприпустимі, наприклад, в аналізі мережевого трафіку або виявленні кіберзагроз. Ще одним важливим напрямком є розроблення моделей з малим споживанням ресурсів, які залишаються високоточною альтернативою для пристроїв із обмеженими обчислювальними можливостями. Це актуально для Інтернету речей (IoT), де ефективність часто залежить від здатності працювати на обмежених ресурсах.

Застосування таких методів вже сьогодні змінює різні галузі. У фінансовому секторі вони використовуються для оброблення транзакцій і швидкого виявлення шахрайства. В охороні здоров'я ці підходи допомагають в реальному часі моніторити стан пацієнтів і вчасно реагувати на критичні зміни. У телекомунікаціях розподілені алгоритми сприяють оптимізації роботи мережі шляхом аналізу величезних потоків даних, а в сфері IoT забезпечують швидку обробку інформації від тисяч сенсорів. Отже, інтеграція цих інновацій не лише підвищує ефективність оброблення даних, але й відкриває нові можливості для розвитку технологій у різних секторах економіки.

Висновки. Інноваційні підходи до розробки алгоритмів машинного навчання для оброблення даних у реальному часі в системах Big Data є ключовим фактором успіху в сучасному інформаційному суспільстві. Запропоновані мето-

ди та архітектури можуть значно підвищити ефективність та продуктивність систем Big Data. Подальші дослідження в цій галузі сприятимуть розвитку нових технологій та їх широкому застосуванню в різних сферах економіки та науки.

Список використаних літературних джерел

- [1] Reinsel, D., Gantz, J., Rydning, J. The Digitization of the World: From Edge to Core. *International Data Corporation*, 2018.
- [2] Zaharia, M., et al. Apache Spark: A Unified Engine for Big Data Processing. *Communications of the ACM*, 59(11), 56–65, 2016.
- [3] Dean, J., et al. Large Scale Distributed Deep Networks. In *Advances in Neural Information Processing Systems*, 1223–1231, 2012.

Innovative approaches to developing real-time machine learning algorithms for processing data in Big Data systems

Volodymyr Pankiv, Oleksandr Storozhuk, Natalya Hissovska

The article examines modern innovative approaches to the development of real-time machine learning algorithms for processing data in Big Data systems. Key issues related to the scalability, performance, and complexity of machine learning algorithms are analyzed. New methods and architectures, including streaming learning, distributed computing, and hybrid models, are proposed to enhance data processing efficiency in real-time. The research outcomes can be applied in various fields such as finance, telecommunications, and healthcare, where rapid and accurate data processing is essential.