

РЕФЕРАТ

Магістерська дисертація: 102 с., 40 рис., 1 табл., 2 додатки, 62 джерела.

Актуальність теми: засоби машинного навчання використовується там, де звичайні алгоритми не можуть бути застосовані через складність задачі та неможливість вирішення її традиційними методами. Проте, об'єми даних необхідні для навчання невпинно ростуть і все частіше не можуть бути швидко та ефективно опрацьовані одним робочим пристроєм. Вирішенням цієї проблеми є використання розподілених обчислень та застосування таких підходів до задач машинного навчання з використанням розподілених систем з декількома обчислювальними вузлами та мережевою взаємодією між ними. За рахунок розподілення можна не лише пришвидшити навчання, а й збільшити пропускну здатність, використовувати потоки даних, виконувати оптимізації над моделями, навчати різні версії паралельно тощо.

Мета дослідження: прискорення машинного навчання за рахунок методу розподіленого машинного навчання на прикладі вирішення задачі пошуку аномалій з використанням дерев ізоляцій.

Для реалізації поставленої мети були сформульовані **наступні завдання:**

- виконати аналіз існуючих методів та підходів до розподіленого машинного навчання;
- збір навчальних даних та формування наборів для виконання розподілення;
- розробити метод розподіленого машинного навчання на прикладі алгоритму дерев ізоляцій;
- тестування та аналіз ефективності отриманого методу;
- визначення подальшого напрямку досліджень.

Об'єкт дослідження: процеси розподіленого машинного навчання.

Предмет дослідження: методи розподіленого машинного навчання.

Методи дослідження: для розв'язання поставленої задачі використовувались дерева та ліс ізоляцій, розподілені обчислення, файлова система GFS, обчислювальний підхід MapReduce, потоки даних.

Наукова новизна: науковим результатом магістерської дисертації є створення методу розподіленого навчання на основі використання розподілених даних, обчислювальних ресурсів та залучення потокової обробки даних.

Практичне значення отриманих результатів: визначається тим, що запропонований метод дозволяє прискорити навчання моделей з використанням дерев ізоляцій, збільшити відмовостійкість системи, підтримувати прозору масштабованість для користувача.

Зв'язок роботи з науковими програмами, планами, темами: робота виконувалась на кафедрі автоматизованих систем обробки інформації та управління Національного технічного університету України «Київський політехнічний інститут ім. Ігоря Сікорського» в рамках теми «Методи та технології високопродуктивних обчислень та обробки надвеликих масивів даних». Державний реєстраційний номер 0117U000924.

Апробація: основні положення роботи доповідались і обговорювались на IV всеукраїнській науково-практичній конференції молодих вчених та студентів «Інформаційні системи та технології управління» (ІСТУ-2020), а також на XVI міжнародній науковій конференції «Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту (ISDMCI'2020)» результати магістерської дисертації докладались на наукових конференціях.

Ключові слова: МАШИННЕ НАВЧАННЯ, РОЗПОДІЛЕНЕ НАВЧАННЯ, ПОШУК АНОМАЛІЙ, ПОТОКОВА ОБРОБКА, ПОТОКИ ДАНИХ.