

АЛГОРИТМИ ОБРОБКИ ЗНАНЬ У МАШИННОМУ НАВЧАННІ

Юскович-Жуковська Валентина Іванівна

кандидат технічних наук,

доцент кафедри інформаційних систем,

Лотюк Юрій Георгійович

кандидат педагогічних наук,

доцент кафедри математичного моделювання,

Соловей Людмила Ярославівна

старший викладач кафедри інформаційних систем

Міжнародний економіко-гуманітарний університет імені

академіка С. Дем'янчука

Анотація. Машинне навчання, як наука, досліджує по-перше, методи побудови алгоритмів, здатних навчатися, по-друге, методи, що дозволяють комп'ютерам покращувати свої характеристики на основі отриманого досвіду. В даний час ведуться розробки новітніх систем машинного навчання, призначених для використання в таких технологіях майбутнього, як Інтернет речей, в концепції «розумне» місто, при створенні смарт – технологій, технологіях блокчейн тощо.

Ключові слова: машинне навчання, алгоритми, нейронні мережі.

Проблеми набуття знань з даних, розробка даних та процедури аналізу і прогнозування в галузі машинного навчання потребують сучасних досліджень. Розрізняють два типи машинного навчання: навчання по прецедентах (індуктивне навчання) і дедуктивне навчання. Класичне машинне навчання (classical machine learning) побудоване на класичних статистичних алгоритмах і розв'язує задачі, пов'язані з прийняттям рішень на основі аналізу даних. Для машинного навчання використовуються різні технології та алгоритми. Зокрема, можуть застосовуватися дискримінантний аналіз, байєсовські класифікатори та інші математичні методи, штучні нейронні мережі (ANN). Нейронні мережі моделюють роботу людського мозку і самонавчаються, враховуючи попередній досвід. Чим більше даних завантажено в

систему, тим більше вибірка і тим точніше працюють алгоритми обробки знань. Під кожен конкретну задачу підбирається свій алгоритм.

Для розв'язання задач в галузі машинного навчання існує багато підходів. Оскільки у більшості випадків алгоритми є наперед невідомими, тому краще використовувати алгоритми пошуку і прийняття рішень. Алгоритми оптимального вибору оперують інтерпретованим цифровим простором, зібраними структурованими або неструктурованими даними, враховують поставлену мету та обирають найкращі шляхи.

Для задач прогнозування знання даних набувають шляхом послідовностей, що задані поміченими графами. Задачі моделювання міркувань є формалізованими, але можуть не мати алгоритмів розв'язку взагалі. Основні алгоритми машинного навчання засновані на математичних моделях, аналізі даних, пошуці оптимальних параметрів за заданими критеріями. Розпізнавання ознак може здійснюватись на ПК, мобільних гаджетах, мікропроцесорах в роботах, в серверах комп'ютерних мереж.

Згідно блогу українських розробників програмного забезпечення DOU машинне навчання - це застосування алгоритмів для автоматичного знаходження закономірностей в даних і використання їх для прийняття великої кількості однотипних рішень, для яких певний відсоток помилок є допустимими [1].

На практиці виокремлюються такі класи застосування machine learning:

- навчання з учителем, тобто кероване навчання (supervised learning);
- навчання без учителя, тобто спонтанне навчання (unsupervised learning);
- активне навчання, тобто тренувальне навчання (active learning);
- напівавтоматичне навчання, тобто неповне тренувальне навчання (semi-supervised learning);
- навчання із підкріпленням, тобто у віртуальному середовищі (reinforcement learning);
- перенесення навчання (transfer learning) [2].

За допомогою навчання з учителем вирішуються, зокрема задачі класифікації та регресії. Машинне навчання без учителя включає в себе наступні типи: кластеризація, узагальнення, пошук правил. Ці алгоритми часто застосовуються в Data Mining і їх можна розглядати як частину Data Science [3]. Одними з найбільш популярних задач машинного навчання є задачі класифікації. Завданням класифікації є передбачення категорій об'єктів та їх поділ за наперед визначеними і заданими ознаками. Система класифікації допомагає вирішувати задачі регресії. Кластерний аналіз (data clustering) використовує алгоритм, який групує набір даних (об'єктів) і визначає чи існує взаємозв'язок між даними (об'єктами). На підставі цього машина вчиться сама. Завданням пошуку асоціативних правил є знаходження закономірностей в потоці даних.

Якщо необхідно вирішити складну обчислювальну задачу і жоден алгоритм не підходить ідеально, то використовують ансамблі. Ансамблі – це поєднання відразу декількох алгоритмів, які навчаються одночасно і при обробці знань виправляють помилки один одного. На сьогоднішній день саме вони дають найточніші результати, тому саме їх найчастіше використовують усі великі ІТ-компанії, для яких важлива швидка обробка великої кількості даних [4].

Найкращий результат виходить, коли алгоритми в ансамблях максимально різні. Наприклад, регресія (regression) і дерева рішень (decision trees) легко поєднуються між собою. Для збору ансамблів застосовують такі алгоритми як Proof of Stake (PoS). Ансамблеві методи StackEnsemble, VoteEnsemble та BestLearner можуть розглядатися як мета-ансамблі, оскільки вони містять інших учнів, включаючи інші ансамблі. При випадковій вибірці даних спочатку навчають кілька алгоритмів, після цього результати обробки знань показують останньому алгоритму. Саме він і приймає остаточне рішення.

При послідовному навчанні алгоритмів спочатку навчаємо перший і відзначаємо місце, де він помилився. Потім навчаємо другий, при цьому особлива увага приділяється багам, на яких помилявся перший і т.д. Таким чином робляться вибірки даних, проте вже не за випадковою ознакою. Тепер кожна наступна вибірка складається з тих даних, на яких помилився попередній алгоритм. Таким чином досягається

найбільш якісний результат, але тільки в тих випадках, де не потрібні паралельні методи.

В компанії Google вважають, що скоро її продукти «перестануть бути результатом традиційного програмування - в їх основу буде покладено машинне навчання». Тому компанії Google, Facebook, Apple, Amazon, Microsoft запрошують фахівців у сфері штучного інтелекту [5]. Ринок машинного навчання швидко зростає. Згідно прогнозів до 2025 року він може збільшитися до \$39,98 млрд. Оскільки методи розв'язання кібернетичних завдань потребують заміни людського інтелекту, то алгоритми обробки знань у машинному навчанні постійно вдосконалюються і переводяться на вищий рівень, на автономне машинне навчання, що потребує нових наукових досліджень.

Література

1. Вступ до Machine Learning: знайомство з моделями. URL: <https://dou.ua/lenta/articles/introduction-machine-learning-1/> (дата звернення: 15.05.2021).

2. Вступ до MachineLearning: чи потрібен вам ML і як правильно поставити йому завдання. URL: <https://dou.ua/lenta/articles/master-of-machines-1/> (дата звернення: 15.05.2021).

3. Класичне машинне навчання: завдання класифікації, узагальнення, кластеризації даних. URL: <https://evergreens.com.ua/ua/articles/classical-machine-learning.html> (дата звернення: 15.05.2021).

4. Ансамблі моделей машинного навчання URL: <https://evergreens.com.ua/ua/articles/ensembles.html> (дата звернення: 15.05.2021).

5. Технології та інновації. IT-enterprise URL: <https://www.it.ua/knowledge-base/technology-innovation/machine-learning> (дата звернення: 15.05.2021).