

*Гордєєв Артем,
здобувач другого (магістерського) рівня вищої освіти
фізико-математичного факультету
Науковий керівник: Мосіюк Олександр,
кандидат педагогічних наук, доцент,
доцент кафедри комп'ютерних наук та інформаційних технологій,
Житомирський державний університет імені Івана Франка
м. Житомир, Україна*

ОСНОВНІ ПОНЯТТЯ МАШИННОГО НАВЧАННЯ З ПІДКРІПЛЕННЯМ

Актуальність. В умовах стрімкого розвитку технології машинного навчання та систем штучного інтелекту одним з важливих напрямів наукових досліджень у цій сфері є розробка алгоритмів, здатних до самостійного навчання і прийняття рішення. Одним із перспективних напрямів цього напрямку є навчання з підкріпленням, яке дає змогу певним агентам знаходити оптимальні рішення для максимізації деякого уявлення про сукупну винагороду. Методи навчання з підкріпленням дозволяють створювати систему, здатну приймати адаптивні рішення в умовах невизначеності й складного завчасно невідомого середовища. Задача навігації в лабіринті є класичною проблемою для навчання з підкріпленням, оскільки вимагає ефективної взаємодії з навколишнім простором, вибору стратегії та знаходження оптимального маршруту. Такі дослідження важливі, як з теоретичної так і з практичної точки зору. Зокрема, сама оптимізація алгоритмів пошуку шляхів допомагає підвищити ефективність використання обчислювальних ресурсів і дозволяє розробляти систему, яка є більш стійкою до непередбачуваних змін [5].

Секція 4. Технології розробки інформаційних систем

У цьому контексті **метою** є розкриття ключових понять, характерних для навчання з підкріплення.

Виклад основного матеріалу. Перш ніж перейти до розкриття досвід у створення алгоритмів машинного навчання з підкріпленням варто розкрити саме поняття. Типово, навчання з підкріпленням (або англійською мовою reinforcement learning) – це галузь машинного навчання, орієнтована на прийняття рішень для максимізації кумулятивної винагороди в певній ситуації [4]. Іншими словами це напрямок штучного інтелекту, для якого характерним є навчання на власних помилках. Певна програм у вигляді «агенту» діє в певному середовищі, де за кожну дію він отримує оцінку: позитивну чи негативну. З часом «агент» розуміє, які дії призводять до кращих результатів і починає їх повторювати.

Щоб краще зрозуміти сутність процесу варто навести приклад дитини, яка навчається ходити. Спочатку вона робить випадкові рухи, але з часом починає розуміти, які з них приводять до успіху, а які навпаки – до падіння. І так, поступово, дитина вдосконалює свої навички і починає ходити впевнено. Аналогічно і певна нейронна мережа навчається. Спочатку певна модель штучного інтелекту пробує різні дії, де за правильні вона отримує нагороду, а за неправильні покарання. Саме таким чином нейромережа поступово здобуває досвід та прагне виконувати завдання з максимальною ефективністю.

Цей підхід відрізняється від традиційних методів машинного навчання тим, що «агент» не отримує готових відповідей, а сам експериментує, взаємодіє з певними даними й середовищем. Тобто навчається на власних помилках [1].

Тепер розкриємо основні терміни, які визначають базу понять для навчання з підкріпленням [3].

Агент – спеціалізована програма, яка безпосередньо взаємодіє з навколишнім середовищем.

Навколишнє середовище – певним чином формалізовано описані умови, в яких діє агент.

Стан – це набір змінних, що описують навколишнє середовище в будь-якій точці. Завдяки взаємодії «агенту» з навколишнім середовищем, він отримує значення стану й інтерпретує надану інформацію, щоб визначити ступінь нагороди та ухвалити рішення про наступну дію.

Простір станів являє собою діапазон значень для всіх змінних стану, які повністю описують навколишнє середовище.

Дією вважається певна одиниця взаємодії з навколишнім середовищем, яка доступна агенту в певному стані.

Простір дій – це усі можливі дії, доступні агенту.

Завдяки *винагороді* реалізується зворотний зв'язок від навколишнього середовища, що дозволяє оцінити наслідки виконання певних дій.

Політика, набір визначених правил – це, власне, адаптація поведінки, певна стратегія дій моделі вироблена на підставі її навчання, яка дасть змогу їй досягти мети й отримати найвищу винагороду [3].

Секція 4. Технології розробки інформаційних систем

Розуміючи, що представляє собою навчання з підкріпленням та його базову термінологію, розкриємо яким чином такі алгоритми розрізняють.

Загалом існує декілька підходів до класифікації алгоритмів машинного навчання з підкріпленням. Перший підхід передбачає їх поділ два наступні основні типи.

- Алгоритми з моделлю. Такі алгоритми дозволяють «агенту» створювати внутрішню модель середовища, щоб передбачити наслідки своїх дій. Це як грати в шахи, знаючи всі можливі ходи. Такі алгоритми ідеально підходять для статичних середовищ, де результат кожної дії чітко визначений.

- Алгоритми без моделі. В такому алгоритмі «агент» навчається без попереднього знання про середовище, вивчаючи його шляхом проб і помилок. Він безпосередньо взаємодіє з середовищем, отримуючи винагороди або покарання за свої дії та поступово формує стратегію, що максимізує винагороду [2].

Класифікувати також доцільно і за типом підкріплення.

Позитивне підкріплення – це коли подія, яка відбувається внаслідок реакції на певну дію «агенту», збільшує силу і частоту цієї дії. Іншими словами, воно позитивно впливає на поведінку. До ключових переваг такого типу навчання з підкріпленням відносять такі:

- максимізація продуктивності;
- підтримує зміни у продовж тривалого періоду часу.

У той же час варто розуміти, що занадто багато підкріплення може призвести до перевантаження станів, а отже привести до зменшення ефективності вирішувати конкретну задачу.

Негативне підкріплення визначається як посилення поведінки через те, що «агент» навчається шляхом припинення дії певних штрафів або ж певних негативних факторів. Позитивним для такого підходу є те, що це:

- дозволяє посилити поведінку;
- забезпечує достатню кількість підкріплень, щоб відповідати мінімальному рівню поведінки [4].

Висновок. Навчання з підкріпленням пропонує дещо відмінний підхід до навчання штучного інтелекту. Моделі штучного інтелекту формуються на основі взаємодії з навколишнім середовищем, отримуючи винагороди або покарання за свої дії. Такий підхід знаходить широке застосування в різних галузях, починаючи від розробки ігрових агентів до створення автономних роботів.

Перспективи подальших дослідження пов'язуються, у першу чергу, з дослідженнями алгоритмів, які дозволяють вибудувати алгоритм та навчити модель, яка зможе виконати пошуку шляху в лабіринті.

Список використаних джерел та літератури

1. Навчання з підкріпленням у машинному навчанні URL: <https://evergreens.com.ua/ua/articles/reinforcement-learning.html> (дата звернення: 7.11.2024р.)

Секція 4. Технології розробки інформаційних систем

2. Gillis A. S., Hashemi-Pour C., Carew Joseph M. What is reinforcement learning? *TechTarget*. URL: <https://www.techtarget.com/searchenterpriseai/definition/reinforcement-learning> (дата звернення: 7.11.2024р.)
3. Myronenko R. Deep Reinforcement Learning 101. *DOU.UA*. URL: <https://dou.ua/forums/topic/35882/> (дата звернення: 7.11.2024р.)
4. Reinforcement learning. URL: <https://www.geeksforgeeks.org/what-is-reinforcement-learning/> (дата звернення: 7.11.2024р.)
5. Sutton R. S. & Barto A. G. *Reinforcement Learning, second edition: An Introduction*. Bradford Book Publ., 2018. 552 p.