

## **ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ ДО НАВЧАННЯ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ З ПІДКРІПЛЕННЯМ**

У сучасному світі більшість спеціалістів у сфері машинного навчання не займаються розробкою таких базових процесів для глибокого навчання як безпосередня реалізація самої моделі мережі зі збереженням усіх нейронів, їх ваг та зв'язків, імплементація процесу розрахунку прогнозу мережі та зворотного поширення помилки тощо. Але той факт, що у вільному доступі існує велика кількість схожих реалізацій, не робить задачу машинного навчання тривіальною. Отриманий результат багато в чому залежить від вибору та препроцесінгу вхідних даних, підбору правильної топології мережі, а також правильного підбору гіперпараметрів. Так як немає найбільш оптимального методу рішення цих підзадач, можна зробити висновок, їх пошук усе ще є актуальним.

Під час навчання з підкріпленням (reinforcement learning) система навчається за допомогою взаємодії з деяким середовищем. Відбувається дослідження того, як нейронна мережа повинна вести себе для того, щоб максимізувати деякий довгочасний результат. Таке навчання відноситься до класу алгоритмів з частковим навчанням, які намагаються знайти правильну стратегію, яка встановлює правильні дії системи для деяких станів середовища [1].

Одним з найпоширеніших методів навчання з підкріпленням є алгоритм Q-навчання (Q-learning). У його основі лежить апроксимація за допомогою нейронної мережі функції прогнозування сумарної дисконтованої нагороди, яка буде отримана на кінцевому сеансі взаємодії з моделлю. На вхід мережі подається поточний стан та дія. Прогноз відбувається для всіх можливих дій і обирається та дія, для якої прогноз найбільший. Отримані нагороди у процесі взаємодії з системою порівнюються з попередніми прогнозами і ця різниця використовується для навчання нейронної мережі. Так як рішення, що приймаються, залежать від прогнозів, а прогнози від прийнятих рішень, можлива стабілізація системи у неоптимальному стані. Для запобігання такої ситуації алгоритм передбачає деяку ймовірність прийняття випадкового рішення, яка поступово зменшується у процесі навчання, що забезпечує можливість дослідження усіх варіантів незалежно від поточних прогнозів [2].

Іншим поширеним алгоритмом навчання з підкріпленням є АЗС (Asynchronous advantage actor-critic). Він багато в чому схожий на метод Q-навчання, але на відміну від останнього, де політика заснована на результатах прогнозування, в АЗС на нейронну мережу покладається і прогноз (заснований лише на поточному стані), і прийняття рішення про наступні дії, як окремі задачі. Рішення про наступні дії виводиться мережею як ймовірність прийняття даних рішень, фактичне рішення приймається на основі цих ймовірностей. Помилка прогнозу розраховується аналогічно помилці в алгоритмі Q-навчання, помилка на виходах, відповідальних за прийняття рішення, розраховується як помилка прогнозу, помножена на логарифм ймовірності прийняття цього рішення[3].

Було вирішено задачу короточасного прогнозування за допомогою штучної нейронних мереж на прикладі комп'ютерної гри «Змійка». Нейронні мережі після навчання можуть керувати діями змійки. Для вирішення задачі було використано два методи навчання з підкріпленням – Q-навчання та АЗС.

У якості вхідних даних було використано розташування перешкод (тіло змійки, границі поля) та нагород.

Було проведено аналіз роботи обох алгоритмів та порівняння їх між собою за допомогою даних критеріїв:

- кількість очок за деяку кількість часу;
- середньоквадратична помилка прогнозу;
- швидкість навчання, оцінена як швидкість досягнення фіксованих показників.

У результаті аналізу було виявлено, що метод АЗС показав кращі результати по всім критеріям, ніж метод Q-навчання.

### **Список використаних джерел**

1. Goodfellow I., Bengio Y., Courville A. Deep Learning / I. Goodfellow, Y. Bengio, A. Courville – Cambridge: «MIT Press», 2016. – 775 с.
2. Melo Francisco S. Convergence of Q-learning: a simple proof / F. S. Melo – Lisboa: «Institute for Systems and Robotics», 2016. – 4 с.
3. Bergdahl Joakim. Asynchronous Advantage ActorCritic with Adam Optimization and a Layer Normalized Recurrent Network / Joakim Bergdahl – Stockholm: «Royal Institute of Technology», 2017. – 62 с.