

Гирилович Д.С., здобувач

Державний університет «Київський Авіаційний Інститут»

АРХІТЕКТУРА DEEP Q-NETWORK ЯК МЕТОД ГЛИБИННОГО НАВЧАННЯ З ПІДКРІПЛЕННЯМ

Навчання з підкріпленням (reinforcement learning, RL) є одним із ключових напрямів сучасного машинного навчання, що дозволяє агентам самостійно формувати стратегії поведінки в умовах невизначеності. Особливого розвитку цей підхід набув із появою методів глибинного навчання, які забезпечили ефективну апроксимацію складних функцій. Одним із найбільш відомих алгоритмів у цій області є Deep Q-Network (DQN), який поєднує класичний Q-learning з можливостями глибоких нейронних мереж [1].

Задача навчання з підкріпленням формалізується у вигляді марківського процесу прийняття рішень (MDP), що описується п'ятіркою (S, A, P, R, γ) , де S – множина станів середовища, A – множина дій агента, P – ймовірнісна функція переходів між станами, R – функція винагороди, γ – коефіцієнт дисконтування [1]. Основною метою є знаходження оптимальної політики $\pi(a|s)$, яка максимізує очікувану сумарну дисконтовану винагороду.

Алгоритм Q-learning базується на ітераційному оновленні функції цінності дій $Q(s, a)$, яка визначає очікувану вигоду від виконання дії a у стані s [3]. Однак класичний табличний підхід стає непридатним у випадку великого або безперервного простору станів. Саме тому виникла необхідність використання апроксимації Q-функції за допомогою нейронних мереж, що і реалізується в архітектурі DQN [3].

Архітектура Deep Q-Network представляє собою багат шарову нейронну мережу, яка приймає на вхід представлення стану середовища та повертає оцінки Q-значень для всіх можливих дій [1]. У класичному варіанті, запропонованому для задач керування іграми, вхідними даними можуть бути сирі пікселі зображення, які проходять через згорткові шари (convolutional layers) для виділення просторових ознак, після чого обробляються повнозв'язними шарами (fully connected layers), що формують кінцеві оцінки дій [3].

Незважаючи на свою ефективність, базова архітектура DQN має ряд обмежень, серед яких переоцінка Q-значень та повільна збіжність. Для їх усунення було запропоновано ряд модифікацій. Зокрема, Double DQN дозволяє зменшити переоцінку значень шляхом розділення вибору та оцінки дії [3]. Dueling DQN вводить окреме представлення для функції

цінності стану та переваги дії, що покращує якість навчання. Пріоритизований replay (prioritized experience replay) забезпечує більш ефективний відбір важливих прикладів для навчання.

Важливим аспектом побудови DQN є вибір структури винагороди. Неправильно сформована функція винагороди може призвести до небажаної поведінки агента або до відсутності збіжності [2].

Сучасні дослідження у галузі глибинного навчання з підкріпленням демонструють, що DQN та його похідні методи можуть ефективно застосовуватися у широкому спектрі задач: від ігрових середовищ до робототехніки та автономних систем [1]. Водночас подальший розвиток цієї області пов'язаний із підвищенням стабільності навчання, зменшенням вимог до обсягу даних та адаптацією до реальних умов функціонування.

Функція втрат у DQN базується на мінімізації різниці між поточним значенням Q та цільовим значенням, яке визначається за рівнянням Беллмана [3]. Таким чином, процес навчання зводиться до задачі регресії, де нейронна мережа поступово наближає істинну функцію цінності дій. Оптимізація здійснюється за допомогою градієнтних методів, таких як стохастичний градієнтний спуск або його модифікації (Adam, RMSProp) [1].

Таким чином, архітектура Deep Q-Network є важливим етапом еволюції методів навчання з підкріпленням, поєднуючи переваги глибинних нейронних мереж та класичних алгоритмів оптимізації. Її використання відкриває широкі можливості для створення інтелектуальних систем, здатних до самонавчання та адаптації в складних динамічних середовищах.

Список використаних джерел:

1. Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). A brief survey of deep reinforcement learning. *IEEE Signal Processing Magazine*, 34(6), 26–38.
2. García, J., & Fernández, F. (2015). A comprehensive survey on Safe Reinforcement Learning. *Journal of Machine Learning Research*, 16, 1437–1480.
3. Fan, J., Wang, Z., Xie, Y., & Yang, Z. (2020). A Theoretical Analysis of Deep Q-Learning. In *Proceedings of the 2nd Conference on Learning for Dynamics and Control* (PMLR, Vol. 120, pp. 486–489).