Зеленський Аркадій Анатолійович, аспірант

ДВНЗ «Криворізький національний університет», м. Кривий Ріг

0009-0008-4019-351X

**ЗАСТОСУВАННЯ Q-НАВЧАННЯ ДЛЯ ЗАДАЧ КЕРУВАННЯ У БАГАТОАГЕНТНИХ СИСТЕМАХ**

Алгоритм машинного навчання Q-навчання використовується для прийняття рішень у ситуаціях, де агент повинен взаємодіяти з оточенням та навчатися від результатів цієї взаємодії. Цей алгоритм зазвичай використовується в задачах управління, де агент має приймати рішення для досягнення мети в середовищі зі складними динамічними характеристиками.

Q-навчання має широкий спектр застосувань у задачах управління в багатоагентних системах. Наприклад, у робототехніці Q-навчання використовується для управління рухом роєм роботів. Рій роботів може виконувати завдання, такі як пошук та рятування або дослідження невідомих областей. Q-навчання дозволяє роботам визначити оптимальні шляхи руху та співпрацювати один з одним для досягнення мети. Управління портфелем цінних паперів може бути розглянуте як задача багатоагентної системи, де кожен агент вибирає операції купівлі-продажу для максимізації прибутку. Q-навчання може бути використане для розробки стратегій торгівлі, агенти навчаються на основі історичних даних про ринок та оновлюють свої стратегії для адаптації до змінних умов ринку. Управління потоками транспорту великих міст може бути багатоагентною задачею. Агенти можуть бути автотранспортними засобами, таксі, автобусами тощо. Q-навчання дозволяє агентам вибирати маршрути і час руху для оптимізації потоку транспорту та зменшення навантаження трафіку. Управління системами постачання води та електроенергії великих міст може бути реалізоване як задача багатоагентної системи. Агенти можуть бути різними джерелами водопостачання або електростанціями. Q-навчання може допомогти агентам оптимізувати режими роботи та забезпечити ефективне використання ресурсів. У відеоіграх, де гравці або NPC (некеровані персонажі) взаємодіють у великих віртуальних світах, Q-навчання може використовуватися для розв'язання проблеми прийняття рішень. Наприклад, NPC може використовувати Q-навчання для навчання того, як краще рухатися або вибирати тактику в бойовій сцені. Управління флотом дронів для виконання завдань, таких як доставка або нагляд, також може бути реалізоване за допомогою Q-навчання. Кожен дрон може бути агентом, який навчається оптимальним маршрутам та стратегіям взаємодії з іншими дронами та оточенням. Це лише кілька прикладів застосування Q-навчання в багатоагентних системах. Важливо відзначити, що цей метод може бути досить універсальним і використовується в різних областях для автоматизації та оптимізації прийняття рішень в умовах невизначеності та змінності.

Створення математичних моделей багатоагентних систем для застосування Q-навчання включає в себе кілька етапів і враховує особливості та складні взаємодії між агентами та їх оточенням. Спершу необхідно описати стани середовища: визначити стани, в яких може опинитися система. Це може бути ряд станів, який представляє параметри або характеристики системи, такі як місце розташування агентів, їхні стани, ресурси тощо. Далі слід визначити можливі дії агентів, а саме описати можливі дії, які властиві агентам в кожному стані системи. Ці дії повинні бути пов'язані з завданнями, які агенти повинні виконати, і враховувати обмеження. Наступним етапом є визначення функції винагороди, яка оцінює вигоду агента в кожному стані та після кожної дії. Ця функція вказує, наскільки корисно або некорисно для агента знаходитися в певному стані та виконувати певні дії. Винагорода може бути позитивною або негативною, залежно від цілей завдання. Далі необхідно визначити модель переходів – як система змінює однин стан на інший після виконання дій агентів. Ця модель описує, як система реагує на дії агентів та інші впливи в середовищі. Після цього варто визначити Q-функцію для оцінки очікуваних винагород після виконання кожної дії в кожному стані. Зазвичай для цього використовується формула Беллмана. Вона допомагає агентам навчитися оптимальним стратегіям на основі попередньої інформації. Далі наступає черга навчання агентів: агенти взаємодіють з середовищем, вибираючи дії на основі їхньої поточної стратегії та оновлюючи Q-функції після кожної дії. Для цього варто визначити, як агенти вибирають дії відповідно до своїх Q-функцій. Це може включати в себе ε-жадібну стратегію для балансу між дослідженням і експлуатацією. Процес навчання агентів є ітеративним: де агенти взаємодіють з середовищем і оновлюють свої Q-функції до досягнення стабільної стратегії або іншого критерію зупинки. В решті-решт, після навчання настає етап перевірки результатів, які досягли агенти, використовуючи навчену стратегію. Необхідно виконати аналіз стратегій обраних агентами та оцінити їх ефективність у вирішенні задачі. Цей процес створення математичної моделі багатоагентної системи для Q-навчання допомагає агентам вчитися оптимальним стратегіям у складних інтеракціях з оточенням та іншими агентами.

Існує кілька різних алгоритмів Q-навчання та його варіацій. Кожен з них має свої переваги та обмеження, вибір конкретного алгоритму Q-навчання залежить від контексту задачі, доступних ресурсів та вимог до системи. У деяких випадках комбінування різних методів або розробка власних варіацій може бути найкращим рішенням для досягнення бажаного результату.

Одним із найбільш популярних варіацій алгоритму є децентралізоване Q-навчання. Його перевагами є ізольованість Q-функції для кожного агента, дозволяє кожному агенту навчатися своїй власній функції та не потребує централізованого контролю. Також, він підходить для сценаріїв, де інформація обмінюється обмежено або агенти мають обмежені обчислювальні ресурси. Серед обмежень можна виділити нижчу ефективність у вирішенні глобальних завдань, де потрібна координація всіх агентів. А також цей підхід вимагає більше часу на навчання, оскільки кожен агент навчається окремо.

Наступною варіацією алгоритму є централізоване Q-навчання зі спільною Q-функцією. До переваг цього алгоритму можна віднести можливість досягнення глобальної оптимальності, оскільки всі агенти навчаються на основі спільної Q-функції. До того ж, цей алгоритм ефективний для задач, де координація агентів вкрай важлива. Серед недоліків даного алгоритму можна виділити те, що він вимагає централізованого контролю та обміну інформацією між агентами, що може бути складним на практиці. Також, масштабування до великої кількості агентів може бути проблематичним з обчислювальної точки зору.

Метод нейронних мереж Deep Q-Networks дозволяє навчати складні Q-функції за допомогою нейронних мереж. Ефективний для задач з великою кількістю станів і можливими безліччю дій. Серед недоліків даного алгоритму можна виділити високу залежність від обчислювальних ресурсів та об'єму даних для навчання, а також в простих сценаріях може бути схильним до перенавчання.

Q-навчання може бути корисним в галузі гірничодобувного підприємства, де управління багатоагентними системами може виникати в різних аспектах операцій та процесів. До можливих практичних застосувань Q-навчання у цій галузі можна віднести:

1. Планування та оптимізація видобутку. В гірничодобувних операціях часто є багато різних локацій та обладнання, які взаємодіють між собою. Застосування Q-навчання може допомогти визначити оптимальні стратегії видобутку для різних ділянок, вибирати маршрути руху обладнання, інтервали обслуговування тощо. Агентами можуть бути видобувні машини, конвеєри, робочі бригади тощо.

2. Безпека та навколишнє середовище. Q-навчання може використовуватися для розробки систем безпеки та моніторингу навколишнього середовища на гірничодобувних підприємствах. Агенти можуть бути дронами, роботами або сенсорами, які відслідковують та реагують на потенційні небезпеки, такі як аварії, витоки газу або обвали.

3. Управління енергозбереженням. В гірничодобувних операціях споживання енергії є важливим фактором. Q-навчання може бути використано для розробки стратегій ефективного використання енергії, контролю освітлення, оптимізації вентиляції та інших систем енергозбереження.

4. Управління транспортними засобами. На гірничодобувних підприємствах можуть бути задіяні транспортні засоби, такі як вантажівки, екскаватори, локомотиви і т. п. Q-навчання може допомогти управляти цими засобами, вибираючи оптимальні маршрути для руху, які підвищать ефективність видобутку та транспортування породи.

Перевагою використання Q-навчання у гірничодобувній галузі є можливість автоматизації та оптимізації різних аспектів виробництва та безпеки, що може призвести до збільшення продуктивності та зниження ризиків. Однак важливо враховувати складність впровадження цих систем і необхідність забезпечення безпеки та надійності в гірничодобувному середовищі.

**Література**

1. Chaoxu M. Q-learning Solution for Optimal Consensus Control of Discrete-Time Multiagent Systems Using Reinforcement Learning / M. Chaoxu. // Journal of the Franklin Institute. – 2019. – №356. – С. 13.
2. Q-Learning: Applications and Convergence Rate Optimization. // Highlights in Science Engineering and Technology. – 2023. – №63. – С. 210–215.
3. Simon M. A multi-agent based cooperative approach to scheduling and routing [Електронний ресурс] / Martin Simon // European Journal of Operational Research. – 2016. – Режим доступу до ресурсу: https://www.sciencedirect.com/science/article/pii/S0377221716300984.
4. KA C. VOLTTRON: Tech-to-Market BestPractices Guide for Small- and Medium-Sized Commercial Buildings [Електронний ресурс] / Cort KA // U.S. Department of evergy. – 2016. – Режим доступу до ресурсу: https://www.pnnl.gov/main/publications/external/technical\_reports/PNNL-25405.pdf.