1. Навчання з підкріпленням (англ. reinforcement learning) - один із способів машинного навчання, в ході якого випробувана система (агент) навчається, взаємодіючи з деяким середовищем. З погляду кібернетики, одна із видів кібернетичного експерименту. Відгуком середовища (а не спеціальної системи управління підкріпленням, як це відбувається у навчанні з вчителем) на прийняті рішення є сигнали підкріплення, тому таке навчання є окремим випадком навчання з вчителем, але вчителем є середовище або його модель. Також потрібно мати на увазі, що деякі правила підкріплення базуються на неявних учителях, наприклад, у разі штучного нейронного середовища, одночасної активності формальних нейронів, через що їх можна віднести до навчання без вчителя.

Система підкріплення та її види

Крім класичного методу навчання перцептрона - методу корекції помилки, який можна віднести до навчання з учителем, Розенблатт також увів поняття про навчання без вчителя, запропонувавши кілька способів навчання:

Альфа-системою підкріплення називається система підкріплення, при якій ваги всіх активних зв'язків, що ведуть до елемента, змінюються на однакову величину r, а ваги неактивних зв'язків цей час не змінюються.

Гамма-системою підкріплення називається таке правило зміни вагових коефіцієнтів деякого елемента, при якому ваги всіх активних зв'язків спочатку змінюються на рівну величину, а потім з усіх ваг зв'язків віднімається інша величина, рівна повній зміні ваг всіх активних зв'язків, поділеному на число всіх зв'язків. Ця система має властивість консервативності щодо терезів, так як у неї повна сума терезів всіх зв'язків не може ні зростати, ні зменшуватися.

1. DQN: Q-Learning, але з глибокою нейронною мережею як апроксиматор функцій.

Використання нелінійної глибокої нейронної мережі це потужно, але навчання нестабільне, якщо ми застосовуємо його наївно.

Уловка 1 - (S, A, R, S\_next) Відтворення досвіду: Збережіть досвід у буфері відтворення та відберіть з нього міні-батчі для навчання мережі. Це декорелює дані та призводить до підвищення ефективності даних. Спочатку буфер відтворення заповнюється випадковим чином.

Уловка 2 - Цільова мережа: використовуйте окрему мережу для оцінки мети TD. Ця цільова мережа має таку ж архітектуру, як і апроксиматор функції, але з фіксованими параметрами. Кожні кроки T (гіперпараметр) параметри з мережі Q копіюються в цільову мережу. Це призводить до стабільнішого навчання, оскільки зберігає фіксовану цільову функцію (на деякий час).

Використовуючи згорткову нейронну мережу як апроксиматор функції на необроблених пікселях ігор Atari, де оцінка є нагородою, ми можемо навчитися грати в багато з цих ігор з продуктивністю, подібною до людської.

Подвійний DQN: як і в звичайному Q-Learning, DQN має тенденцію до завищення значень через свою максимальну операцію, що застосовується як для вибору, так і для оцінки дій. Ми обходимо це, використовуючи мережу Q для вибору та цільову мережу для оцінки при внесенні оновлень.

Глибокі агенти Q-мережі

Алгоритм глибокої мережі Q (DQN) є методом навчання з підкріпленням поза політикою без моделей. Агент DQN є заснованим на значенні агентом навчання з підкріпленням, який навчає критика оцінювати повернення чи майбутні винагороди. DQN є варіантом Q-вивчення.

Під час навчання, агент:

Оновлює характеристики критика на кожному тимчасовому етапі під час вивчення.

Досліджує прогалину дії за допомогою епсілон-жадібного дослідження. Під час кожного інтервалу управління агент вибирає випадкову дію з ймовірністю ϵ, інакше це вибирає дію жадібно щодо функції значення з ймовірністю 1-ϵ. Ця жадібна дія є дією, для якої функція значення є найбільшою.

Попередній досвід сховищ за допомогою кругового буфера досвіду. Агент оновлює критика на основі міні-пакету подій, випадково вироблених від буфера.

Створення агента

Створити агента DQN:

Створіть об'єкт уявлення критика.

Задайте опції агента за допомогою функції rlDQNAgentOptions.

Створити агента за допомогою функції rlDQNAgent.