1. Введение в машинное обучение

...

Алгоритмы обучения с подкреплением

Обучение с подкреплением (Reinforcement learning, RL) представляет собой уникальную парадигму в **машинном обучении**, при которой агент обучается путем взаимодействия с окружением. В отличие от **обучения с учителем**, которое опирается на размеченные данные, или **обучения без учителя**, при котором исследуются неразмеченные данные, **обучение с подкреплением** сфокусировано на обучении методом проб и ошибок, которое управляется обратной связью в виде вознаграждений или штрафов. Этот подход имитирует то, как люди учатся на собственном опыте, что делает **обучение с подкреплением** особенно подходящим для задач, связанных с последовательным принятием решений в динамических окружениях.

Представьте это как дрессировку собаки. Вы не даете собаке явные инструкции в виде команд “сидеть”, “стоять” или “апорт”. Вместо этого вы поощряете ее лакомствами и похвалой, когда она выполняет нужные действия, и корректируете, когда она этого не делает. Собака учится связывать определенные действия с положительными результатами посредством проб, ошибок и обратной связи.

Как работает обучение с подкреплением

В **обучении с подкреплением** агент взаимодействует с окружением, выполняя действия и наблюдая за последствиями. Окружение предоставляет обратную связь в виде вознаграждений или штрафов, направляя агента к обучению оптимальной политике. **Политика** - это стратегия выбора действий, которая максимизирует суммарное вознаграждение с течением времени.

Алгоритмы **обучения с подкреплением** можно в общем виде разделить на:

1. **Обучение с подкреплением на основе модели**: агент изучает модель окружения, которую он использует для предсказания будущих состояний и планирования своих действий. Этот подход аналогичен наличию карты лабиринта перед его прохождением. Агент может использовать эту карту для планирования наиболее эффективного пути к цели, снижая необходимость в пробах и ошибках.

2. **Обучение с подкреплением без модели**: агент обучается напрямую из опыта без явного моделирования окружения. Это похоже на прохождение лабиринта без карты, где агент полагается исключительно на пробы и ошибки и обратную связь от окружения, чтобы определить наилучшие действия. Агент постепенно улучшает свою политику, исследуя различные пути и обучаясь на полученных вознаграждениях или штрафах.

Основные концепции обучения с подкреплением

Чтобы разобраться с тем, как работает **обучение с подкреплением (RL),** необходимо понять его основные концепции. Эти концепции формируют основу понимания того, как агенты обучаются и взаимодействуют с окружением для достижения своих целей.

Агент

Агент (agent) - это обучающаяся и принимающая решения сущность в системе **обучения с подкреплением**. Он взаимодействует с окружением, выполняя действия и наблюдая за последствиями. Цель агента - обучиться оптимальной политике, которая максимизирует суммарное вознаграждение с течением времени.

Можете представить, что **агент** - это робот, перемещающийся по лабиринту, программа, играющая в игру, или беспилотный автомобиль, движущийся в потоке транспорта. В каждом случае **агент** принимает решения и учится на основе своего опыта.

Окружение

Окружение (environment) - это внешняя система или контекст, в котором действует агент. Оно включает все, что находится вне агента, включая физический мир, симулированный мир (смоделированное окружение) или даже игровое поле. **Окружение** реагирует на действия агента и предоставляет обратную связь в виде вознаграждений или штрафов.

В примере с лабиринтом **окружение** - это сам лабиринт с его стенами, проходами и расположением цели. В сценарии игры **окружение** - это сама игра с ее правилами и действиями противника.

Состояние

Состояние (state) представляет текущее положение или состояние окружения. Оно предоставляет снимок релевантной информации,

необходимой агенту для принятия обоснованных решений. **Состояние** может включать различные аспекты окружения, такие как положение агента, положения других объектов и любые другие релевантные переменные.

Состояние робота, перемещающегося по лабиринту, может включать его текущее расположение и окружающие стены. В шахматной игре **состояние** - это текущая конфигурация на шахматной доске.

...