

Полная программа

Reinforcement Learning

Reinforcement Learning

Длительность курса: 80 часов

Модуль 1. Введение в Reinforcement Learning

Тема 1

Знакомство с Reinforcement Learning

Цель занятия

рассмотреть обучение с подкреплением, включая его значение, области применения и потенциальное влияние.

Краткое содержание

история обучения с подкреплением; различие между обучением с подкреплением и другими методами машинного обучения; обзор применения обучения с подкреплением.

Тема 2

Ключевые понятия RL: агент, среда, награда, политика. Построение среды

Цель занятия

погрузиться в специфику обучения с подкреплением, охватывающую основные термины и компоненты; понять, что такое среда и как ее создать; понять, что такое среда и как ее построить под свои требования.

Краткое содержание

агент обучения с подкреплением; окружение; состояние; действие; вознаграждение; стратегия; функция оценки; MDP.

Тема 3

Основные алгоритмы RL: Value based

Цель занятия

познакомиться с основными алгоритмами, используемыми в обучении с подкреплением; понять, что такое v-функция и q-функция; понять уравнение Беллмана; алгоритмы группы Value-based: SARSA, Q-learning.

Краткое содержание

v-функция и q-функция; уравнение Беллмана; алгоритмы группы Value-based: SARSA, Q-learning.

Тема 4

Основные алгоритмы RL: Policy based

Цель занятия

понять, что такое политика и как ее использовать для обучения агента; понять и реализовать алгоритм Policy Gradient, Natural Policy Gradient.

Краткое содержание

обучение агентов обучения с подкреплением; понятие политики; алгоритм PG, NPG.

Домашние задания

Разработка агента для решения задачи с конечным набором состояний и действий

Цель
В этом ДЗ вы реализуете один из алгоритмов RL и научитесь взаимодействовать с Gymnasium.

Тренируемые навыки:
- отладка агента RL;
- отработка MDP на практике.

Модуль 2. Deep Reinforcement Learning

Тема 1

Введение в Deep Reinforcement Learning

Цель занятия

познакомиться с концепцией глубокого обучения с подкреплением; рассмотреть значение глубокого обучения с подкреплением в создании сложных моделей; рассмотреть концепцию частично-наблюдаемой среды и понять, какие изменения в основные алгоритмы это вносит.

Краткое содержание

основные концепции глубокого обучения; интеграция глубокого обучения с обучением с подкреплением; преимущества и вызовы глубокого обучения с подкреплением; концепция частично-наблюдаемой среды.

Тема 2

Deep Q-Network (DQN) алгоритм

Цель занятия

изучить, как работают DQN и какие компоненты участвуют в его работе.

Краткое содержание

Q-обучение; нейронные сети; разведка против эксплуатации в DQN; повторение опыта, фиксированные Q-цели.

Домашние задания

Задача с конечным набором дискретных действий

Цель
В процессе выполнения задания вы проведёте первичный анализ поставленной задачи и выберете подходящий алгоритм.

Для успешного выполнения задания вам потребуются:
- понимание отличий между средами с дискретными и непрерывным набором действий;
- навык оценки задачи и выбора подходящего алгоритма;
- навык разработки несложных алгоритмов на Python;
- навык реализации нейронных сетей с использованием фреймворка PyTorch;
- навык отладки алгоритма и выбора оптимальных гиперпараметров.

Тема 3

Deep Policy Gradient (PG) алгоритм

Цель занятия

рассмотреть алгоритмы на основе политик; изучить, как работают алгоритмы на основе политик, и чем отличаются от методов на основе значений q.

Краткое содержание

Policy Gradient; алгоритм REINFORCE.

Тема 4

Actor-Critic алгоритм

Цель занятия

изучить методы actor-critic, которые являются гибридом методов на основе политик и на основе значений; рассмотреть основные преимущества и недостатки этих алгоритмов.

Краткое содержание

Actor-Critic; A2C; A3C.

Домашние задания

Задача с непрерывным спектром действий в непрерывной среде

Цель
В этом ДЗ вам нужно решить задачу управления с непрерывным пространством действий.

Тема 5

TRPO -> PPO

Цель занятия

понять дальнейшее развитие алгоритма Policy Gradient до алгоритма PPO.

Краткое содержание

алгоритм Trusted Region Policy Gradient; алгоритм Proximal Policy Optimization.

Тема 6

DDPG -> TD3 -> LSTM-DD3

Цель занятия

рассмотреть дальнейшее развитие алгоритмов до Deep Deterministic Policy Gradient и Twin Delayed Deep Deterministic Policy Gradient; понять, как применять LSTM блоки для работы с временными последовательностями.

Краткое содержание

алгоритм Deep Deterministic Policy Gradient; алгоритм Twin Delayed Deep Deterministic Policy Gradient; модификация LSTM Twin Delayed Deep Deterministic Policy Gradient.

Модуль 3. Advanced Reinforcement Learning

Тема 1

Обучение с использованием модели среды (model-based rl)

Цель занятия

познакомиться с концепцией обучения с подкреплением на основе модели среды и подходами её использования для выработки более эффективного поведения агента.

Краткое содержание

моделирование динамики среды; алгоритмы планирования; внутренняя и внешняя награда; обучение на основе любопытства.

Тема 2

Model-based, часть 2

Цель занятия

познакомиться с концепцией обучения с подкреплением на основе модели среды и подходами её использования для выработки более эффективного поведения агента.

Краткое содержание

моделирование динамики среды; алгоритмы планирования; внутренняя и внешняя награда; обучение на основе любопытства.

Домашние задания

Обучение Dupa-Q агента для стохастической среды

Цель
1. Реализовать класс, моделирующий стохастическую среду Frozen Lake Vx8.
2. Обучить агента Dupa-Q, использующего стохастическую модель среды и сравнить с простым Q-агентом.

Тема 3

Иерархическое обучение с подкреплением

Цель занятия

познакомиться с идеями обучения агента путём формирования иерархии навыков.

Краткое содержание

декомпозиция глобальной цели на подзадачи; план поэтапного обучения агента; формирование у агента отдельных навыков; перенос обучения; принцип феодализма.

Тема 4

Выбор темы и организация проектной работы

Цель занятия

выбрать и обсудить тему проектной работы; спланировать работу над проектом; ознакомиться с регламентом работы над проектом.

Краткое содержание

правила работы над проектом и специфика проведения итоговой защиты; требования к результату проекта и итоговой документации.

Тема 5

Многоагентное обучение и кооперация агентов

Цель занятия

узнать об особенностях обучения группы агентов со схожими или противоположными целями.

Краткое содержание

взгляд на задачу с позиции теории игр; кооперативные, конкурентные и смешанные среды; централизованное и децентрализованное обучение; дизайн общей функции вознаграждения; коммуникация агентов.

Тема 6

Трансформеры в RL: decision transformers и action transformers

Цель занятия

познакомиться с применением современных трансформерных архитектур к задачам обучения с подкреплением.

Краткое содержание

рассмотрим разницу между подходами decision transformers и action transformers; поймем особенности применения обоих подходов; научимся применять предобученные трансформерные модели.

Модуль 4. Применение RL в реальных задачах

Тема 1

Применение RL в игровой индустрии

Цель занятия

рассмотреть применение RL в игровой индустрии.

Краткое содержание

RL в игровой индустрии.

Тема 2

Применение RL в робототехнике

Цель занятия

рассмотреть применение RL в робототехнике.

Краткое содержание

RL в робототехнике.

Тема 3

RL в рекомендательных системах

Цель занятия

научиться разрабатывать рекомендательные системы по концепции serendipity.

Краткое содержание

рассмотрим современные ключевые проблемы разработки и развития рекомендательных систем применительно к разным видам бизнеса (медиа, ритейл, телеком, финансы); научимся решать задачи поиска похожих элементов и ранжирования по концепции serendipity, используя стэкинг технологий глубокого обучения с подкреплением с классическими supervised и unsupervised подходами.

Тема 4

RL в задаче скоринга

Цель занятия

научиться создавать систему оценки кредитных рисков заёмщика (модель скоринга).

Краткое содержание

оценка заёмщика; сложности разделения хороших/плохих кредитов; преимущества и недостатки скоринговой оценки.

Тема 5

Применение RL в управлении финансовым портфелем

Цель занятия

рассмотреть применение RL в управлении финансовым портфелем.

Краткое содержание

RL в управлении финансовым портфелем.

Модуль 5. Проектная работа

Тема 1

Консультация по проектам и домашним заданиям

Цель занятия

получить ответы на вопросы по проекту, ДЗ и по курсу.

Краткое содержание

вопросы по улучшению и оптимизации работы над проектом; затруднения при выполнении ДЗ; вопросы по программе.

Домашние задания

Курсовой проект

Цель
Сформулировать тему, направленную на решение какой-то практической задачи методами RL.

Тема 2

Предзащита

Цель занятия

представить подготовленную работу и получить обратную связь от руководителя курса для подготовки к защите проектной работы.

Краткое содержание

порядок защиты; состав материалов для защиты; практические советы по подготовке выступления на защите.

Тема 3

Защита проектных работ. Проведение итогов курса

Цель занятия

защитить проект и получить рекомендации экспертов.

Краткое содержание

презентация проектов перед комиссией; вопросы и комментарии по проектам.