

Глубокое обучение с подкреплением в задаче двуногой ходьбы

Гречишкин Никита Андреевич

Студент (специалист)

Московский государственный университет имени М.В.Ломоносова,
Механико-математический факультет, Кафедра теоретической механики и мехатроники,
Москва, Россия

E-mail: grechnikita@yandex.ru

Данная работа посвящена исследованию применения алгоритмов обучения с подкреплением в задаче двуногой ходьбы. Основная идея такого обучения в том, что автономный агент взаимодействует с окружающей средой, предпринимая некоторые действия, а информацию из окружающей среды он получает в форме «вознаграждения», имеющего количественное значение. Агент продолжает действовать, пытаясь максимизировать своё «вознаграждение», характеризующее переход из одного состояния в другое, и таким образом ищет оптимальную стратегию взаимодействия со средой. В стандартной постановке данный алгоритм требует для оптимальности многократное использование всех возможных состояний, что невозможно в непрерывных задачах. Для решения этой проблемы рассматриваются различные модификации алгоритма, ускоряющие процесс обучения с подкреплением для применения в практических задачах. В качестве одной из таких модификаций были выбраны нейронные сети.

Для примера применения данного алгоритма в двуногой ходьбе использовалась задача построения устойчивого движения трехмерной модели человекоподобного робота NAO в среде для моделирования роботов V-REP. Основная цель, которая ставилась модели - пройти определённое расстояние, и при этом не упасть в процессе движения. При этом на модель не было наложено никаких дополнительных ограничений, кроме физических из симулятора.

В работе описаны эксперименты, проведённые с помощью программы, написанной на языке Python, которая симулирует движение робота в заданной среде. Применялись различные варианты алгоритмов для решения данной задачи, описаны преимущества и недостатки различных подходов. Исследовались различные параметры, настройки и архитектуры нейронных сетей для выявления оптимальных параметров рассматриваемых алгоритмов. Произведён анализ полученных результатов. Построены различные варианты движения робота, которые можно применять в практических задачах. В реальной сложной ситуации робот с помощью данного алгоритма обучения может автономно подстраиваться под внешнюю среду, исследует её, улучшает своё управление и достигает заданные цели. При этом механизм не нуждается во внешнем управлении даже в условиях изменяющейся внешней среды, подстраиваясь под неё самостоятельно.