Министерство науки и высшего образования Российский Федерации
Московский физико-технический институт
(государственный университет)

А. И. Панов

Методы и алгоритмы машинного обучения с подкреплением

Учебно-методическое пособие

Москва
МФТИ
2018

В пособии рассмотрены основные

Оглавление

Введение

Обучение с подкрепление (reinforcement learning, RL) является разделом машинного обучения, активно развивающимся направлением в искусственном интеллекте. Несмотря на то, что формально обучение с подкреплением относится к разделу приобретения знаний, оно кардинально отличается от таких методов, как обучение с учителем или без учителя. В первую очередь, здесь явно выделен субъект приобретения знаний (агент), который принимает решения и некоторым образом влияет на источник анализируемых данных (среду). Эта агентная постановка очень близка по своей методологии к одному из определений искусственного интеллекта, который давали одни из основоположников искусственного интеллекта Рассел и Норвиг [11]:

> *Искусственный интеллект — это наука об «интеллектуальных агентах», т.е. о некотором устройстве или программе, которая воспринимает свою среду и выполняет действия, которые максимизируют ее шансы на успех при достижении какой-то цели.*

Наличие у агента некоторого набора возможных способов воздействия на среду (действий) и его стремления достигнуть некоторой поставленной заранее цели в этой среде позволяют естественным образом применять обучение с подкреплением в более сложных интеллектуальных системах, которые разрабатываются для синтеза целенаправленного поведения: в интеллектуальных динамических системах и в частности в робототехнике [10]. В обучении с подкреплением наиболее тесно переплетаются методы планирования поведения, представления и приобретения знаний. В настоящее время именно подсистемы, реализующие методы обучения с подкреплением, становятся центральными элементами комплексных систем управления поведением автономных объектов, взамен ранее занимавших главенствующую позицию подсистем представления знаний. Таким образом, наблюдается переход от более статичных когнитивных архитектур (например, Soar [3]), к более активным обучающимся архитектурам (например, знаковым [9]).

Успехи в развитии методов обучения с подкреплением возродили интерес к разработке агентов, действующих в искусственных средах, в том числе игровых. Были разработаны обучающиеся агенты, демонстрирующие иногда результаты, превосходящие уровень человека, для сред компьютерных игр (Atari [5], DOOM [1], Starcraft [6], Minecraft [2]), так и для более серьезных, приближенных к условиям, в которых действуют люди в реальной жизни (Go [4], OpenAI Universe[1]). Демонстрируемые успехи, понятные и знакомые даже далеким от искусственного интеллекта людям, породили большую волну новых исследований и сейчас секции по обучению с подкреплением занимают самую большую часть научных конференций (ICML, NIPS, IJCAI).

[1] `https://blog.openai.com/universe/`

Настоящее учебное пособие призвано дать краткий обзор современных подходов в обучении с подкреплением и является сжатым описанием основных алгоритмов, в том числе тех, которые появились буквально в последние несколько лет. Все подходы распределены на группы (обучения с моделью, по стратегиям и т.п.), каждой выделена отдельная глава. Границы между этими группами зачастую условным и некоторые подходы могут быть отнесены сразу к нескольким направлениям. Так как обучение с подкреплением развивается очень динамично и каждый день появляются новые работы, настоящий обзор не может быть полным. Особое внимание уделено перспективному обучению с подкреплением и некоторым нейрофизиологическим и психологическим обоснованиям.

В подготовке этого пособия были использованы материалы больших монографий по обучению с подкреплением, которые могут служить основной дополнительной лит литературой: книга одних из основоположников этого направления Саттона и Барто [12], краткий обзор Жепешвари [7] и ряд Интернет ресурсов [2].

В качестве короткой исторической справки необходимо отметить, что идеи, лежащие в основе современной теории обучения с подкреплением, высказывались еще на первых этапах становления искусственного интеллекта с 60-х гг. XX в. (отечественные работы по автоматам Цетлина и Стефанюка [13], зарубежные инженерные работы Вальц и Фу и др.[8]) и были заимствованы из области психологии, где еще с начала XX в. существовало понятие обусловленности поведения и условных рефлексов (Павлов, Скиннер).

[2]блог Массимиалано Патачола `https://mpatacchiola.github.io/blog/2016/12/09/dissecting-reinforcement-learning.html`, курс Школа Яндекса `https://github.com/yandexdataschool/Practical_RL`, блог Мустафы Алзантота `https://medium.com/@m.alzantot/deep-reinforcement-learning-demystified-episode-0-2198c05a6124`.

Глава 1

Основные понятия

1.1 Марковский процесс принятия решений

Агент, среда, подкреплением, марковский процесс.

1.2 Динамическое программирование

1.3 Методы Монте-Карло

1.4 Q-обучение

Глава 2

Приближенные методы

2.1 Предсказание с изменением стратегии

2.2 Предсказание без изменения стратегии

2.3 Нейронные сети как аппроксиматоры

Глава 3

Перспективные направления

3.1 Иерархическое обучение с подкреплением

Иерархия действий: Options

Иерархия автоматов: HAM

Оптимизация функции оценки: MaxQ

Автоматическое формирование иерархий

3.2 Внутренняя мотивация

Глава 4

Обучение с подкреплением и другие науки

4.1 Психология

4.2 Нейрофизиология

4.3 Робототехника

Заключение

Немного о целях

Список литературы

1. Clyde: A deep reinforcement learning DOOM playing agent / N. For [et al.] // What's Next For AI In Games. — 2017.
2. Control of Memory, Active Perception, and Action in Minecraft / J. Oh [et al.]. — 2016. — arXiv: 1605.09128.
3. *Laird J. E.* The Soar Cognitive Architecture. — MIT Press, 2012. — P. 374.
4. Mastering the game of Go with deep neural networks and tree search / D. Silver [et al.] // Nature. — 2016. — Vol. 529, no. 7587. — P. 484–489.
5. Playing Atari with Deep Reinforcement Learning / V. Mnih [et al.] // arXiv: 1312.5602. — 2013. — P. 1–9. — arXiv: 1312.5602.
6. StarCraft II: A New Challenge for Reinforcement Learning / O. Vinyals [et al.]. — 2017. — arXiv: 1708.04782.
7. *Szepesvári C.* Algorithms for Reinforcement Learning. Vol. 4. — 2010. — P. 1–103.
8. *Waltz M., Fu K.* A heuristic approach to reinforcement learning control systems // Automatic Control, IEEE Transactions on. — 1965. — Vol. AC–10, no. 4. — P. 390–398.
9. Знаковая картина мира субъекта поведения / Г. С. Осипов [и др.]. — М. : Физматлит, 2018. — С. 264.
10. *Осипов Г. С.* Методы искусственного интеллекта. — М. : ФИЗМАТЛИТ, 2011. — С. 297.
11. *Рассел С., Норвиг П.* Искусственный интеллект: современный подход. — 2-е. — М. : Издательский дом "Вильямс", 2006. — С. 1408.
12. *Саттон Р., Барто Э. Г.* Обучение с подкреплением. — 2-е. — М. : БИНОМ. Лаборатория знаний, 2011. — С. 399.
13. *Стефанюк В. Л.* Локальная организация интеллектуальных систем. — М. : ФИЗМАТЛИТ, 2004. — С. 328.