

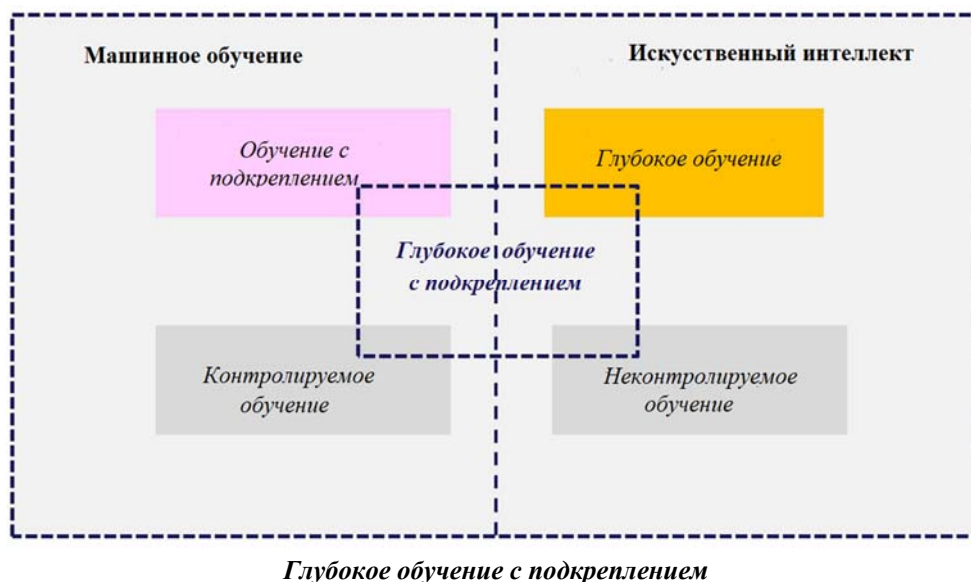
АНАЛИЗ ТЕХНОЛОГИЙ ГЛУБОКОГО ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ ДЛЯ СИСТЕМ МАШИННОГО ЗРЕНИЯ

И. Б. Широков, С.В. Колесова, В.А. Кучеренко, М.Ю. Серебряков

В последние годы особый метод машинного обучения под названием глубокое обучение приобрел огромную популярность, поскольку он позволил получить поразительные результаты в таких областях применения, как распознавание образов, распознавание речи, компьютерное зрение и обработка естественного языка. Недавние исследования также показали, что методы глубокого обучения могут быть объединены с методами обучения с подкреплением и адаптированы для задач с высокой размерностью исходных данных. В данной статье рассматривается происхождение глубокого обучения с подкреплением, его достоинства, недостатки и перспективы развития.

Ключевые слова: глубокое обучение, обучение с подкреплением, глубокое обучение с подкреплением.

Как следует из названия, глубокое обучение с подкреплением (DRL) возникло в результате совмещения обучения с подкреплением и глубокого обучения и может рассматриваться как мост между обычным машинным обучением и настоящим искусственным интеллектом, что продемонстрировано на рисунке. Оно сочетает в себе технику вознаграждения на основе действий из обучения с подкреплением и идею использования нейронной сети для изучения различного вида характеристик из глубокого обучения.



Традиционное обучение с подкреплением ограничено сферами с простым представлением состояния, в то время как DRL позволяет агентам принимать решения на основе многомерных и неструктурированных входных данных, используя нейронные сети для формирования политик. За последние несколько лет проводилось множество активных исследований в области DRL, которые охарактеризовались значительным прогрессом, наряду с возрастающим интересом к глубокому обучению.

Глубокое обучение с подкреплением в жизни простых людей. DRL привлекло много внимания, в особенности благодаря своим известным достижениям в играх. Начиная примерно с 2013 года, DeepMind продемонстрировала впечатляющие результаты обучения в видеоиграх компании Atari на уровне человека, без использования дополнительных функций, закодированных вручную. Другой важной вехой стал 2016 год, когда AlphaGo [1] впервые обыграла профессионального игрока в го, игру, пришедшую из Древнего Китая. Эта компьютерная программа была усовершенствована до AlphaZero [2] в 2017, что повысило ее производительность в игре в шахматы и сёги. В 2019 году Pluribus [3] удачно превзошел профессионалов в многопользовательском покере, а OpenAI [4] превзошел предыдущих чемпионов мира в демонстрационном матче по игре Dota 2.

Вопросами исследования каналов связи РТС занималось большое число научных школ, этим исследованиям посвящено большое количество работ. Определенный обзор этих систем осуществлён в [4]. Там же показано, что известные методы исследования информационных каналов связи РТС не позволяют в полной мере оценить влияние локальных турбулентностей, возникающих в открытых каналах

связи, на параметры электромагнитного поля. Показано, что на сегодняшний день практически все экспериментальные исследования посвящены изучению флуктуаций амплитуды принятого сигнала, в отдельных случаях – амплитуды и углов прихода радиоволн. На основе этих исследований получены энергетические спектры флуктуаций сигналов, сформированы рекомендации по улучшению работы РТС. В работе [4] показано, что указанные исследования являются неполными. Об исследованиях флуктуаций набега фазы электромагнитных волн микроволнового диапазона при прохождении их через турбулентную атмосферу до момента их описания в [4] в мире не известно. Это связано с тем, что традиционные методы измерения фазы не могут быть использованы на протяженных трассах, поскольку для организации таких исследований необходимо на обоих концах измерительной трассы организовывать синфазные микроволновые колебания. Это задача не решена, и не может быть решена в принципе. Вместе с тем, организация подобных исследований представляет большой интерес для науки и целого ряда практических применений. Это даст возможность оценивать характер турбулентности не по косвенным измерениям (амплитуда и угол прихода), а по прямым измерениям амплитуды и набега фазы. Такие исследования позволят расширить знание в вопросах распространения радиоволн, ввести новые модели механизмов распространения, сделать соответствующие прогнозы и дать рекомендации по построению РТС различного назначения.

Существуют и другие примеры влияния DRL на человека. В наши дни многие операции проводятся при помощи компьютерного зрения и манипуляторов, управляемые хирургом. Эта огромная часть прогресса медицинских технологий, которая не только сводит на «нет» фактор человеческой усталости, но также и повышает эффективность самой процедуры. Роботы с DRL могут сильно помочь простым хирургам. Например:

- следить за работой врача и страховать его в случае невнимательности;
- улучшать видимость места оперирования, подсказывать последовательность действий во время процедуры;
- точно и инвазивно разрезать ткани;
- подбирать оптимальную геометрию разреза и накладываемого шва, снижая тем самым уровень боли для пациента.

Помимо хирургии DRL также используется в сфере медицины для диагностики заболевания. Для решения такой задачи модель на вход получает изображение с рентгенологическим или ультразвуковым исследованием пациента, и по ним составляет заключение о наличии болезни. Точность предсказания наличия заболевания превышает 90%. Модели обучаются на специальных тренировочных наборах данных. Большое распространение такие DRL получили в сфере классификации заболеваний легких, предсказании появления опухолей или подборе дозы контраста изображения во время проведения МРТ.

Это лишь всего две сферы деятельности, но, DRL имеет большой потенциал и в других областях, в том числе и в робототехнике, обработке естественного языка, компьютерном зрении, транспортировке объектов, сфере здравоохранения и т.д. Многие потрясающие открытия исследований данной темы были опубликованы множеством крупных компаний, включая Google Brain, DeepMind, Facebook, а также научными лабораториями Беркли, Стэнфорда и университета Карнеги-Меллона, совместно с некоторыми независимыми некоммерческими исследовательскими организациями, такими как OpenAI, и некоторыми промышленно ориентированными компаниями.

Алгоритмы DRL. Наиболее часто используемые алгоритмы DRL можно разделить на методы, основанные на ценности, методы, основанные на градиенте политики, и методы, основанные на моделях.

Методы, основанные на ценности. Методы, основанные на ценности, конструируют функцию полезности для определения политики, которая основывается на методе Q-обучения (Q-learning) [5], используя при этом уравнение Беллмана и его вариант, адаптированное Q-обучение. Алгоритм Deep Q Network, используемый с большим успехом, является представителем этого класса, как и следующие за ним более усовершенствованные версии, такие как двойная глубокая Q-сеть (DQN) [6], Distributional DQN и т.д. Комбинация этих улучшений была изучена при тестировании на наборе игр Atari 2600 как с точки зрения эффективности, так и конечной производительности.

Метод градиента политики. Однако подходы на основе DQN ограничены проблемами с дискретными и низкоразмерными пространствами действий и детерминированными политиками, в то время как методы, основанные на градиенте политики, могут работать с непрерывными пространствами действий, а также могут составлять стохастические политики. Благодаря различным вариантам стохастического градиентного подъема по параметрам политики были разработаны данные методы, целью которых является поиск политики, параметризованной нейронной сетью, для максимизации ожидаемого суммарного вознаграждения. Как и все остальные методы, основанные на политике, данный вариант обычно требует оценки функции полезности для текущей политики и тогда пример эффективного подхода будет заключаться в использовании архитектуры актер-критик, которая может работать с данными вне политики. Алгоритм глубокого детерминированного градиента политики является представителем этого типа методов. В данный момент существуют исследователи, которые пытаются сочетать методы градиента политики и Q-обучение.

Методы, основанные на моделях. Как методы, основанные на ценности, так и методы, основанные на градиенте политики, не используют какую-либо модель среды, за счет чего иногда называются безмодельными методами, что ограничивает их эффективность выборки. При этом в методах, основан-

ных на моделях, модель среды либо явно дана, либо извлечена из опыта с помощью аппроксиматоров функций в сочетании с алгоритмом планирования. Для получения преимуществ с обеих сторон, существует множество доступных исследований, интегрирующих безмодельные и основанные на моделях элементы, которые находятся среди ключевых направлений для будущего развития алгоритмов DRL.

Выводы. В данной работе было описано происхождение глубокого обучения с подкреплением, его связь с искусственным интеллектом, а также приведены примеры использования глубокого обучения с подкреплением в обыденной жизни. Кроме того, были рассмотрены и описаны основные алгоритмы DRL – методы, основанные на ценности; метод градиента политики и методы, основанные на моделях.

Список литературы

1. Silver D., Huang A., Maddison C.J. Mastering the game of Go with deep neural networks and tree search. *Nature* 2016, 529. P. 484–489.
2. Silver D., Hubert T., Schrittwieser J., Antonoglou I. Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *arXiv* 2017, *arXiv:1712.01815*.
3. Popov A.V., Sayarkin K.S., Zhilenkov A.A. Analysis of perspective models of artificial neural networks for control of robotic objects // *Proceedings of the 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2018, St. Petersburg and Moscow, 29 января – 01 2018 года. St. Petersburg and Moscow: Institute of Electrical and Electronics Engineers Inc., 2018. P. 958-961. DOI 10.1109/ElConRus.2018.8317248*.
4. Широков И.Б., Гимпилевич Ю.Б., Сердюк И.В. Исследование характеристик каналов связи: Монография. ИНФРА –М, 2021. 247 с.
5. Жиленьков А.А., Черный С.Г. Извлечение информации из bigdata с помощью нейросетевых архитектур как сетей ассоциаций информационных гранул // *Труды Института системного анализа Российской академии наук. 2022. Т. 72. № 3. С. 81-90*.
6. François-Lavet V., Henderson P. An introduction to deep reinforcement learning. *arXiv* 2018, *arXiv:1811.12560*.

Широков Игорь Борисович, д-р техн. наук, профессор, shirokov@ieee.org, Россия, Севастополь, Севастопольский государственный университет,

Колесова София Васильевна, магистрант, Россия, Санкт-Петербург, Санкт-Петербургский государственный морской технический университет,

Кучеренко Владислав Александрович, ассистент, vlad.kucherenko1@mail.ru, Республика Крым, Керчь, Керченский государственный морской технологический университет,

Серебряков Михаил Юрьевич, преподаватель, zhilenkovanton@gmail.com, Россия, Санкт-Петербург, Санкт-Петербургский государственный морской технический университет

ANALYSIS OF DEEP LEARNING TECHNOLOGIES WITH REINFORCEMENT FOR MACHINE VISION SYSTEMS

I.B. Shirokov, S.V. Kolesova, V.A. Kucherenko, M.Y. Serebryakov

In recent years, a particular machine learning technique called deep learning has gained enormous popularity, as it has produced striking results in applications as wide-ranging as pattern recognition, speech recognition, computer vision, and natural language processing. Recent research has also shown that deep learning methods can be combined with reinforcement learning techniques to teach useful representations for tasks with high input dimensionality. This paper discusses the origin of deep reinforcement learning, its advantages, disadvantages, and prospects for development.

Key words: deep learning, reinforcement learning, deep reinforcement learning.

Shirokov Igor Borisovich, doctor of technical sciences, professor, shirokov@ieee.org, Russia, Sevastopol, Sevastopol State University,

Kolesova Sofia Vasilievna, master, Russia, St. Petersburg, St. Petersburg State Maritime Technical University,

Kucherenko Vladislav Aleksandrovich, assistant, vlad.kucherenko1@mail.ru, Republic of Crimea, Kerch, Kerch State Marine Technological University,

Serebryakov Mikhail Yurievich, teacher, zhilenkovanton@gmail.com, Russia, St. Petersburg, St. Petersburg State Maritime Technical University