ИПМ им.М.В.Келдыша РАН

Будущее-2022 • Труды конференции

И.В.Оселедец

Успехи и проблемы машинного обучения

Рекомендуемая форма библиографической ссылки

Оселедец И.В. Успехи и проблемы машинного обучения // Проектирование будущего. Проблемы цифровой реальности: труды 5-й Международной конференции (3-4 февраля 2022 г., Москва). — М.: ИПМ им. М.В.Келдыша, 2022. — С. 102-108. — https://keldysh.ru/future/2022/9.pdf https://doi.org/10.20948/future-2022-9

Размещено также видео выступления

Успехи и проблемы машинного обучения

И.В. Оселедец

Сколковский институт науки и технологий

Аннотация. В последнее время методы машинного обучения достигли существенных успехов при обработке изображений, анализе текстов, видео, аудио. В статье представлен обзор некоторых интересных результатов в данной области, обсуждены существующие практические и теоретические проблемы, а также дан краткий анализ перспективных приложений.

Ключевые слова: машинное обучение, искусственный интеллект

Successes and problems of machine learning

I.V. Oseledets

Skolkovo Institute of Science and Technology

Abstract. Recently, machine learning techniques have made significant advances in image processing, text analysis, video analysis, audio analysis. The talk will give an overview of some interesting results in this area, discussing existing practical and theoretical problems, as well as a brief analysis of promising applications.

Keywords: machine learning, artificial intelligence

Введение

В последнее время много говорят про искусственный интеллект и его применение. Под термином ИИ, «искусственный интеллект» каждый понимает что-то свое. В русском языке ИИ имеет коннотацию «общего искусственного интеллекта» (general artificial intelligence) – системы, которая способна решать самостоятельно различные задачи так же, как человек, или даже лучше. Однако практики ИИ обычно подразумевают определение, данное Джоном Маккарти в 1956 г.: «ИИ – это наука и технология создания интеллектуальных машин, особенно интеллектуальных компьютерных программ». Сейчас, когда говорят «искусственный интеллект», в 99% случаях подразумевают «машинное обучение», а когда говорят «машинное обучение» – подразумевают глубокие нейросети. В этой статье мы попробуем остановится на основных понятиях (не вдаваясь в технические дета-

ли, которые требуют глубокой математической и технической подготовки) и обсудить, как работают методы машинного обучения, каких успехов уже удалось достичь и какие проблемы стоят перед наукой в данной области.

Как же работает искусственный интеллект?

Как же работают методы машинного обучения, на которых строятся методы работы искусственного интеллекта? Очень показательно высказывание Р. Саттона, который написал: «ИИ строит модель мира вокруг нас». Подавая в модели данные (информацию) о том, что мы видим, слышим или читаем, мы можем научить машину «понимать», как работает мир. Означает ли это, что ИИ может только воспроизводить то, что он видел, и запоминать? Способен ли он заниматься творчеством, находить что-то новое, удивлять, в конце концов? Может ли у него появиться «свобода воли»? Однозначных, прямых ответов на этот вопрос, конечно, нет, однако можно дать некоторые интересные комментарии. Один из ведущих в мире специалистов по машинному обучению и ИИ, И. Суцкевер привел следующее рассуждение. Почему мы думаем, что наш мозг («естественный интеллект») может принимать решения и имеет свободу воли? Объяснение может быть таким: внутри нашего мозга мы имеем некоторую модель, которая предсказывает наши решения. Но принципы работы нашего мозга настолько сложны, что он не может себя моделировать, и это приводит к решениям, которые мы сами не можем предсказать. Перенося эту концепцию на случай искусственного интеллекта, можно предположить, что если удастся создать систему, модель машинного обучения, которая способна не только решать одну задачу, но способна предсказывать свое собственное поведение, то может возникнуть аналогичная ситуация: ИИ не сможет предсказать свое поведение. Конечно, важным условием будет не то, что ИИ не может предсказать свое поведение, но то, что «естественный интеллект», люди, не смогут предсказать результат работы алгоритма (что уже во многих случаях происходит, достаточно вспомнить достижения в игре в го или шахматы).

Если продолжать сравнение мозга и ИИ, важно отметить текущие существенные преимущества человеческого мозга: он гораздо более энергоэффективен. В среднем человеческий мозг потребляет порядка 30 Вт, и постоянно обучается; обучение больших нейросетевых моделей может требовать мегаватты электроэнергии. Одной из основных задач является создание как вычислительных устройств, так и новых моделей, которые могут быстро и энергоэффективно решать различные задачи.

Причины революции глубокого обучения

Алгоритмы, основанные на глубоких нейросетевых моделях, имеют давнюю историю. Понятие «искусственный нейрон» было введено Макка-

локом и Питтсом и такую конструкцию реализовал Розенблатт [8] в 1958 г. Глубокая нейросеть была предложена в работах Ивахненко и Лапы [3]. Первые сверточные сети появились в работах Ле Куна и соавторов [4] в 1980-х гг. Метод обратного распространения ошибки, ключевой для обучения глубоких нейросетей, известен достаточно давно [9]. Однако до середины 2000-х гг. глубокие нейросети не давали преимущества ни в одной практической задаче по распознаванию образов и проигрывали методам, основанным на «специальных» признаках, которые строятся по изображению. Одним из переломных моментов стало соревнование «Assira», в котором стояла очень простая задача: научить ИИ отличать кошек от собак. По состоянию на 2006 г. самые точные методы делали это с точностью 60%. К концу 2014 г., когда это соревнование закончилось, победитель смог достичь точности 99,98% Это колоссальный прогресс. Сейчас обучить такую нейросеть сможет (при небольшом обучении) даже школьник. Так что же произошло, в чем причина так называемой «революции глубокого обучения»? Были ли придуманы новые алгоритмы, которых раньше не было? Алгоритмы действительно появились, однако базовые, основные подходы были давно известны. Изменилось лишь то, что ими научились правильно пользоваться. «Классические» подходы машинного обучения обладают свойством «насыщения» – с какого-то момента добавление дополнительных данных не приводит к повышению качества работы. Методы, основанные на глубоких нейросетевых моделях, имеют гораздо больше параметров, и они не насыщаются: чем больше данных, тем выше качество.

Важную роль сыграло появление больших датасетов. Трудно недооценить роль ImageNet: набор из 10 млн размеченных изображений с 1 тыс. классов, который был собран коллективными усилиями (Google, Принстон) и выложен в открытый доступ. Задача состоит в том, чтобы по изображению определить, что на нем изображено. На этом датасете оказалось, что глубокие нейросетевые модели позволяют получать очень высокое качество распознавания. Более того, «модульность» нейросети (она состоит из последовательных обучаемых блоков) оказалась крайне полезной: обученные признаки (features, на жаргоне – «фичи») являются очень информативными. В частотности, признаки, полученные из нейросети VGG-16 до сих пор считаются одними из эталонных. Таким образом, обучившись на большом размеченном датасете, модель смогла понять, что является на картинке значимым и за чем «нужно смотреть».

Кроме размеченных данных есть еще несколько важных причин успеха. Появились большие открытые программные пакеты для реализации различных нейросетевых архитектур и их обучения: сначала это были разработки академического сообщества (Theano, Caffee) а потом подключились большие компании (Tensorflow, Pytorch). В этих фреймворках сведена к минимуму техническая работа, а также работа по реализации алго-

ритмов обучения. Так как снижена техническая сложность, количество возможностей по реализации и тестированию алгоритмов увеличилось. Второй момент – вычислительные мощности. Появление графических карт и специализированных библиотек для работы на них существенно снизило время обучения, и наличие нескольких таких карт (за сравнительно небольшую стоимость) позволяло обучать современные, эффективные модели. Сейчас «золотой век» закончен, и модели высокого уровня вновь обучаются на вычислительных кластерах с тысячами графических ускорителей, но на начальном этапе доступность вычислительных ресурсов сыграла важнейшую роль в развитии области.

Все эти факторы (данные, фреймворки, железо) в сочетании привели к фокусировке десятков тысяч исследователей в одной узкой области, что и привело к революции глубокого обучения. Таких феноменов в истории науки было не так уж и много. В результате такой фокусировки появилось огромное количество новых подходов и приложений: генеративные (состязательные) нейросетевые модели [6], модели обработки естественного языка, игра в шахматы и го, решение задач дизайна материалов и многое другое. Методы машинного обучения глубокого проникли в нашу жизнь и продолжают активно развиваться. В большинстве электронных сервисов осуществляется распознавание лиц, анализируются социальные сети, рекомендательные системы, финансовые рынки, промышленность… Число разных подходов велико.

Текущие тренды

Какие вопросы представляют сейчас особый интерес? Сразу следует сказать, что область меняется очень быстро, и многие из текущих трендов могут перестать ими быть в течение ближайших лет. Однако некоторые вопросы остаются нерешенными уже достаточно долго, что позволяет говорить о некоторых трендах. Первой задачей, которую успешно решили методы искусственного интеллекта, является распознавание класса изображения по большой, размеченной выборке. Эта задача называется «обучение с учителем». Недостаток состоит в том, чтобы собрать большой датасет требуется много усилий по разметке, а в некоторых задачах, например при анализе медицинских снимков, сделать это практически невозможно.

Поэтому особое внимание привлекают методы работы с данными без разметки. У нас есть сотни тысяч рентгеновских снимков. Можем ли мы выделить признаки, не просмотрев их вручную? Оказывается, что да. Уже упоминавшиеся генеративные сети были одними из первых работ в этой области, за ними последовали другие.

Другим трендом является снижение вычислительной сложности: хорошо работающие модели часто являются «тяжелыми», и требуют много

вычислительных ресурсов. Как снизить требования по памяти, по числу операций, не снижая при этом точности?

Важным направлением являются «нестандартные» приложения машинного обучения. Основные успехи ИИ связаны с обработкой изображений и текстов, а также звуков. Однако в последнее время те же методы начали успешно применяться для задач биологии, химии, дизайна лекарств, оптимизации производственных процессов. Перенос методов из одной области в другой не всегда оказывается успешным.

И последний тренд, про который бы хотелось упомянуть – это мультимодальные задачи, где на вход идут разнородные данные: текст и изображение, текст, видео и звук и так далее. Также особый интерес уделяется многозадачным подходам, когда одна модель может успешно решать не одну, а целый ряд задач. Это приближает нас к созданию более универсального ИИ.

Открытые задачи и проблемы

Есть несколько существенных, открытых проблем, которые ждут своего решения. Одна из главных проблем – а почему методы глубокого обучения вообще работают? Это требует некоторого пояснения. Число параметров, даже в не очень большой модели, составляет сотни миллионов, а самые «тяжелые» модели имеют сотни миллиардов параметров. Несмотря на то, что все говорят о больших данных, число обучающих примеров – десятки, максимум сотни миллионов. Т.е., число примеров меньше числа параметров. Согласно классической теории, должно происходить переобучение – модель должна прекрасно «запомнить» обучающую выборку, однако на тех данных, которых она не видела, давать большую ошибку. Это мы видим на классических моделях машинного обучения, но совершенно не видим для глубоких нейросетей. Поэтому, мы не можем оценить, сколько нужно обучающих примеров. Достаточно ли 100 примеров? 1000 примеров? Можем ли мы гарантировать какую-то точность? Все это открытые вопросы. Лишь недавно начали появляться работы, в первую очередь, работы посвященные феномену «двойного спуска» (double descent) (М. Белкин) [2], которые частично пытаются объяснить различие между теорией и практикой.

Вторая большая проблема, про которую нужно упомянуть – это эффект неустойчивости. В работе Я. Гудфеллоу и соавторов [7] был впервые замечен следующий эффект: при небольшом, невидимом человеческому глазу возмущении картинки, предсказание нейросети переводит распознанный образец в другой класс. Такие возмущения получили название «adversarial» (злонамеренные). Большое количество работ посвящено защите от таких атак и построению более сильных атак, однако, по-видимому, задача построения абсолютно устойчивого к атакам классификатора не имеет решения. Результаты А.Н. Горбаня и его соавторов [1]

теоретически показывают, что любая модель ИИ будет делать ошибки. Однако, эти ошибки достаточно легко исправить, создавая новые модели.

И наконец, последняя крупная проблема – это проблема интерпретируемости. В отличие от классических моделей машинного обучения, нейросеть представляет из себя «черный ящик», который получает на вход данные, перерабатывает их внутри в соответствии со своими «обученными» параметрами, и выдает ответ. Что повлияло на этот ответ? Что нужно поменять, чтобы получить другой ответ? Все это достаточно нетривиальные вопросы. Ведь ИИ руководствуется, в первую очередь, теми данными, которые приходят на вход, и любые предрассудки и ошибки в этих данных «зашиваются» внутрь обученной модели. Если есть несбалансированность в данных по полу, возрасту, социальному статусу – весь этот дисбаланс, если не принимать специальных мер, будет только усиливаться. Для этого необходимо двигаться в направлении интерпретируемых архитектур и подходов.

Риски

Риски развития ИИ не стоит недооценивать. Их можно разделить на два типа: можно «не успеть» что-то сделать, и мы отстанем от мира; можно сделать слишком много, и работа ИИ будет наносить вред. Для работы алгоритмов ИИ нужны данные. У кого больше данных, у того и конкурентное преимущество, и поэтому уже много лет большое внимание уделяется хранению и работы с персональными данными. В Москве достаточно успешно идет проект по хранению и работе с медицинскими данными, что дает нам большое преимущество в этой области. В других областях (например, в сельском хозяйстве) потенциал по сбору данных огромный, но никакой системной работы не ведется.

Еще одним национальным риском является отставание по вычислительным ресурсам. С учетом постоянно вводимых ограничений, здесь может сложиться достаточно напряженная ситуация, и необходимо усиленными темпами развивать собственную микроэлектронную промышленность, а также альтернативные вычислительные системы (нейроморфные и квантовые). Без значительных вычислительных мощностей создание больших нейросетевых моделей является очень затруднительным.

Если говорить о глобальных рисках ИИ, то они, безусловно, есть. Чем больше действий будет передаваться машинам, тем они выше. Машинное обучение хорошо в тех областях, где цена ошибки не высока. С учетом того, что никто не может дать гарантии надежности, это все приводит к большим юридическим и этическим проблемам, которые еще предстоит решить. Очень просто «довериться» алгоритму, однако последствия могут быть очень тяжелыми. Поэтому любое масштабное внедрение ИИ должно быть контролируемым. Однако положительные эффекты все равно существенно больше.

Заключение

Машинное обучение достигло огромных успехов. Пускай до «универсального ИИ» еще далеко, развитие методов работы с мультимодальными данными дало огромное количество инструментов для создания новых решений и сервисов, а также большой простор для теории. Возможно, требуется «новая математика» для того, чтобы решить открытые проблемы ИИ, или аналог знаменитых «проблем Гильберта» начала XX в., чтобы перейти от большого количества эмпирических результатов к стройной и понятной теории. Однако уже сейчас ИИ предсказывает погоду, ставит диагнозы, выигрывает в сложные игры не только у человека, но и у алгоритмов «прошлого поколения», поэтому будем очень внимательно следить за развитием и принимать в этом активное участие.

Литература

1. *Gorban A.N., Makarov V.A., Tyukin I.Y.* High-dimensional brain in a high-dimensional world: Blessing of dimensionality // Entropy 22.1, 82 (2020).
2. *Belkin M., Hsu D., Ma S., Mandal S.* Reconciling modern machine-learning practice and the classical bias–variance trade-off // Proceedings of the National Academy of Sciences 116(32), 15849-15854 (2019).
3. *Ивахненко А.Г., Лапа В.Г.* Кибернетические предсказывающие устройства. – К.: Наукова думка. 1965.
4. *LeCun Y., Boser B. et al.* Backpropagation applied to handwritten zip code recognition // Neural computation 1(4), 541-551 (1989).
5. *LeCun Y., Bengio Y., Hinton G.* Deep learning // Nature 521(7553), 436-444 (2015).
6. *Goodfellow I., Pouget-Abadie J. et al.* Generative adversarial nets // Advances in neural information processing systems. 2014, 27.
7. *Goodfellow I.J., Shlens J., Szegedy C.* Explaining and harnessing adversarial examples. arXiv:1412.6572
8. *Rosenblatt F.* The perceptron: a probabilistic model for information storage and organization in the brain // Psychological review 65(6), 386 (1958).
9. *Hecht-Nielsen R.* Theory of the backpropagation neural network. InNeural networks for perception 1992 Jan 1 (pp. 65-93). Academic Press.