ИПМ им.М.В.Келдыша РАН

Будущее-2021 • Труды конференции

В.С. Смолин

Революция в искусственном интеллекте: Достижения и перспективы

Рекомендуемая форма библиографической ссылки

Смолин В.С. Революция в искусственном интеллекте: Достижения и перспективы // Проектирование будущего. Проблемы цифровой реальности: труды 4-й Международной конференции (4-5 февраля 2021 г., Москва). — М.: ИПМ им. М.В.Келдыша, 2021. — С. 147-155. — https://keldysh.ru/future/2021/13.pdf https://doi.org/10.20948/future-2021-13

Размещено также видео выступления

Революция в искусственном интеллекте: Достижения и перспективы

В.С. Смолин

Институт прикладной математики им. М.В. Келдыша РАН

Аннотация. Коммерчески успешное применение нейросетевых алгоритмов в системах и устройствах искусственного интеллекта (ИИ) после 2010 г. значительно ускорило процесс достижения новых успехов в решении «интеллектуальных» задач. Дальнейшее развитие работ по ИИ повлияет не только на технологический уклад, но и на социальные отношения в человеческом обществе, и о возможных последствиях такого влияния необходимо задумываться уже сейчас.

Ключевые слова: искусственный интеллект, нейровычисления, формальный нейрон, цивилизация

The artificial intelligence revolution: Achievements and prospects

V.S. Smolin

RAS Keldysh Institute of Applied Mathematics

Abstract. The commercially successful application of neural network algorithms in artificial intelligence (AI) systems and devices after 2010 has significantly accelerated the process of achieving new successes in solving “intellectual" problems. Further development of work on AI will affect not only the technological order, but also social relations in human society, and it is necessary to think about the possible consequences of such an influence right now.

Keywords: artificial intelligence, neurocomputing, formal neuron, civilization

1. Введение

Развитие вычислительной техники и информационных сетей позволяет с каждым днём делать технические системы и устройства всё более «умными», решать без участия человека задачи, которые ранее было принято рассматривать как «интеллектуальные».

Уже сейчас имеются значительные успехи в обработке изображений, устной и письменной речи, управлении беспилотными дронами и автомобилями, в логических и компьютерных играх и ряде других задач. Полученные практические результаты позволяют не только выпускать коммерчески успешные продукты, но и приближаться к более глубокому пониманию природы решения «интеллектуальных» задач. До 2010 г. центральным направлением развития систем ИИ было создание эвристических алгоритмов, имитирующих «интеллектуальные» действия, а так называемые «нейросетевые» алгоритмы рассматривались как перспективная ветвь, не имевшая серьёзных практических применений.

В 2010-12 гг. ситуация существенно изменилась: начиная с этого времени каждый год, месяц, а теперь уже и неделя приносят сообщения о новых успешных решениях всё более сложных и интересных «интеллектуальных» задач с использованием «нейросетевых» алгоритмов.

Почему переход на «нейросетевые» алгоритмы при решении «интеллектуальных» задач принято называть революцией в ИИ? Ведь все «нейросетевые» алгоритмы точно так же придуманы разработчиками, как и другие эвристические программы? Отличия всё-таки есть: «нейросетевыми» принято называть алгоритмы, допускающие массовое распараллеливание вычислений и выполняемые ими преобразования строятся больше на анализе входных данных, чем на опыте разработчиков.

2. Алгоритмические достижения ИИ

Главным достижением нейросетевой революции в ИИ является увеличение числа настраиваемых параметров с десятков и сотен до миллионов и миллиардов (в 2020 г.GPT-3 – $175 \cdot 10^9$ параметров, в 2021 г. счёт пошёл на триллионы).

В настоящее время в подавляющем большинстве «нейросетевых» моделей используются алгоритмы на идее градиентного спуска и методе быстрого вычисления градиентов, получившим название «обратного распространения ошибки» (BPE, back propagation error). BPE позволяет эффективно вычислять градиенты функции ошибки по вектору параметров для многослойных нейросетей. Описание BPE вошло в большинство современных учебников по глубоким нейросетям [1,2]. «Глубокими» принято называть «нейросетевые» структуры со скрытыми слоями, сейчас используются сети со многими десятками и сотнями скрытых слоёв.

«Скрытыми» (hidden) называют все слои кроме входного и выходного слоёв, см. рис. 1. Активность входного слоя определяется поданным сигналом $\vec{X}$, активность выходного слоя $\vec{Y}$ можно наблюдать и сравнивать с желаемой.

Можно наблюдать и за активностью скрытых слоёв. Но в этом нет необходимости, поскольку BPE обеспечит настройку параметров (т.е. весов связей между элементами) автоматически, без контроля человека.

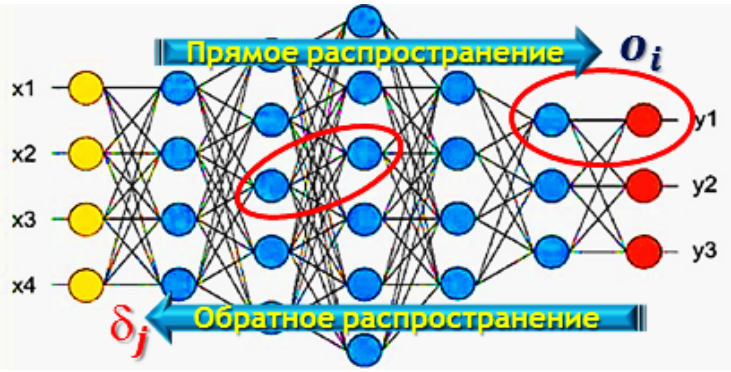

Рис. 1. «Нейросетевая» структура преобразования $\vec{X} \to \vec{Y}$
и «прямое» распространение выходной (output) активности o_i
и «обратное» распространение «ошибки» δ_j

«Нейросетевая» структура образована из формальных нейронов, выполняющих преобразование, состоящее из линейной активации элементов a_i и нелинейного выхода o_i:

$$a_i = \vec{O}^l * \vec{W}_i - b_i;\ o_i = \varphi(a_i), \tag{1}$$

где $\vec{O}^l = \{o_k^l\}$ – вектор выходной активности элементов предыдущего слоя, $\vec{W}_i = \{w_{ki}\}$ – вектор весов связей, b_i – параметр сдвига, а $\varphi(a_i)$ – монотонно возрастающая нелинейная функция.

Выполняемое структурой преобразование $\vec{X} \to \vec{Y}$ настраивается таким образом, чтобы аппроксимировать некоторое эталонное преобразование $\vec{X} \to \vec{T}$ на основе набора примеров его выполнения $\{\vec{X}_m \to \vec{T}_m\}$, причём цель настройки – построение аппроксимации эталонного преобразования, которое может быть использовано для входных сигналов $\vec{X}$, не вошедших в $\{\vec{X}_m \to \vec{T}_m\}$.

Задается функция ошибки, которая позволяет количественно рассчитать точность полученной аппроксимации:

$$E = \tfrac{1}{2} \sum_{m=1}^{P} \sum_{i=1}^{N} (t_i^m - y_i^m)^2, \tag{2}$$

где суммирование производится по номерам эталонов m из набора примеров $\{\vec{X}_m \to \vec{T}_m\}$ и компонентам i выходного вектора $\vec{Y}$.

Настройка миллионов параметров w_{ki} на основе миллиардов примеров $\{\vec{X}_m \to \vec{T}_m\}$ выглядит очень сложной задачей. Ситуация облегчается тем, что если мы знаем, как менять линейную активацию – того элемента структуры, (то есть сумели определить значение $\delta_i = \partial E / \partial a_i$), то легко считаются все значения

$$\frac{\partial E}{\partial w_{ki}} = \frac{\partial E}{\partial a_i}\frac{\partial a_i}{\partial w_{ki}} = \delta_i o_k. \quad (3)$$

Для выходного слоя δ_i равна сумме разностей между значениями эталонной t_i^m и полученной y_i^m компонентами выходного сигнала для всех используемых для обучения примеров, умноженных на производную $\varphi()$ при текущем значении a_i^m:

$$\delta_i = \sum_{m=1}^{P} \delta_i^m = \sum_{m=1}^{P} \frac{\partial E}{\partial a_i^m} = \sum_{m=1}^{P} \frac{\partial E}{\partial y_i^m}\frac{\partial y_i^m}{\partial a_i^m} = \sum_{m=1}^{P} (t_i^m - y_i^m)\frac{\partial \varphi(a_i^m)}{\partial a_i^m} \quad (4)$$

Линейное суммирование возможно, поскольку суммируются производные по линейной активации. Для скрытых слоёв δ_j можно подсчитать на основе δ_i следующего за рассматриваемым слоя.

$$\delta_j = \sum_{m=1}^{P} \delta_j^m = \sum_{m=1}^{P} \frac{\partial E}{\partial a_j^m} = \sum_{m=1}^{P} \left(\sum_{i=1}^{N} \delta_i w_{ji}\right)\frac{d\varphi(a_j^m)}{d a_j^m}. \quad (5)$$

Поскольку для последнего (выходного) слоя δ_i считаются по (4), таким образом можно рассчитать δ_j для всех слоёв, и на основе (3) задать приращения всех весов связей:

$$\Delta w_{kj} = -\alpha \delta_j \sum_{m=1}^{P} o_k^m, \quad (6)$$

где коэффициент скорости обучения $\alpha \ll 1$.

Формулы (1)–(6) описывают алгоритм настройки параметров w_{kj} для структур с прямым, не имеющим обратных связей, распространением сигнала. Но подход обобщается и на рекуррентные сети, которые сейчас получили широкое распространение. Для этого используется эквивалентная замена рекуррентной структуры на сеть с прямым распространением сигналов (рис. 2).

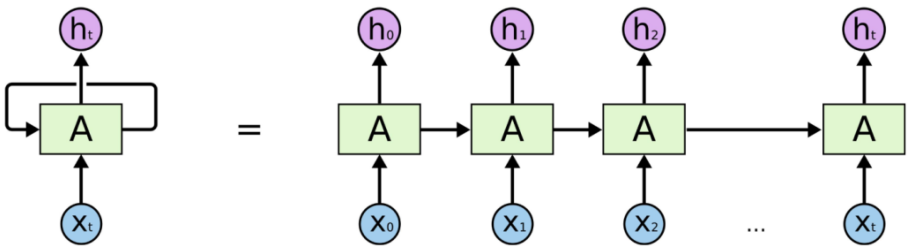

Рис. 2. Эквивалентная замена рекуррентной структуры

Ни градиентный спуск, ни расчёт производных на графе вычислений не являются новыми идеями. Более того, приведённый выше формализм ВРЕ был разработан ещё в 1980-е гг. Почему же нейросетевая революция в ИИ произошла только десять лет назад? Дело в том, что при простом

применении к многослойным сетям метод BPE становится неустойчивым: и выходная активность o_i и «обратное» распространение «ошибки» δ_j имеют тенденцию либо «взрываться», либо затухать до нуля. И простая нормализация обеих величин возможна, но ведёт к потере эффективности.

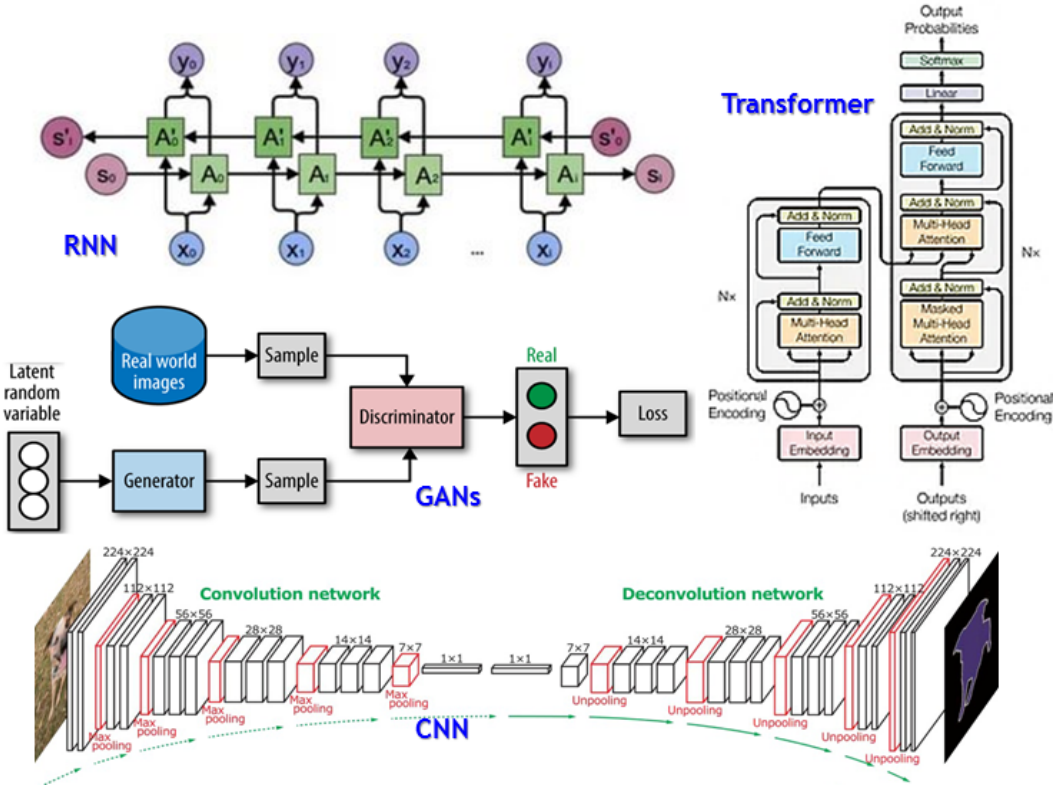

Рис. 3. Основные современные «нейросетевые» модели

К 2010 г. научились осуществлять тонкие настройки параметров, позволявшие выполнять BPE на структурах с небольшим, порядка десяти, числом слоёв. Только к 2014 г. был разработан метод BatchNorm, позволивший нормализовать o_i и δ_j без заметной потери эффективности выполняемых преобразований. Это позволило увеличить число слоев сетей до многих десятков и сотен и сделало их действительно «глубокими».

Было и много других алгоритмических находок. Заметно расширился набор используемых структур «нейросетевых» моделей. Но именно успехи в решении проблемы устойчивости послужили основой нейросетевой революции в ИИ. А применение больших данных и мощных векторных (графических) процессоров способствовало дальнейшему прогрессу.

3. Зимы ИИ – не будет!

Волна энтузиазма по поводу нейросетевых вычислений – это уже третья «весна» ИИ. Но предыдущие два раза за весной следовала «зима» – интерес к нейросетевой тематике спадал, финансирование урезали. Хотя эвристические алгоритмы для ИИ от «сезонов» не страдали: победа

шахматного суперкомпьютера Deep Blue (без «нейросетевых» алгоритмов) над Каспаровым состоялась в мае 1997 г, в разгар последней «зимы».

И сейчас высказываются мнения, что нейрохайп пройдёт, а эвристические алгоритмы останутся. Понятно, что подпитывают такие мнения как раз специалисты по эвристическим алгоритмам ИИ, которые десятилетиями их разрабатывали, добились ряда успехов, признания, возглавили лаборатории и кафедры.

Но объективные показатели говорят об обратном – «зимы» ИИ для нейровычислений больше не будет! Основные причины придерживаться такого мнения собраны в табл.

Наличие успехов нейровычислений в разные периоды времени

Показатель	1973	1997	2021
Более успешное решение «интеллектуальных» задач по сравнению с эвристическими алгоритмами	нет	нет	да
Коммерчески успешные продукты на нейросетевых методах	нет	нет	да
Крупносерийное производство средств вычислительной техники, предназначенных для нейровычислений	нет	нет	да
Государственные программы развития ИИ, широкое распространение учебных курсов по нейровычислениям	нет	нет	да

Нейросетевые алгоритмы всё более широко рассматривается как «новое программирование»: вместо написания эвристических алгоритмов на основе анализа больших данных человеком нейросетевые модели строятся для обработки больших данных и автоматической настройки алгоритмов их преобразования.

Преимущества «нового программирования» проявились, например, в работе программы «AlphaZero», которой потребовалось всего несколько часов, чтобы достичь уровня игры в шахматы, сёги и го, превосходящей уровень любого игрока, включая все написанные ранее программы, на разработку которых учёными и программистами ушли месяцы и годы [3] (рис. 4).

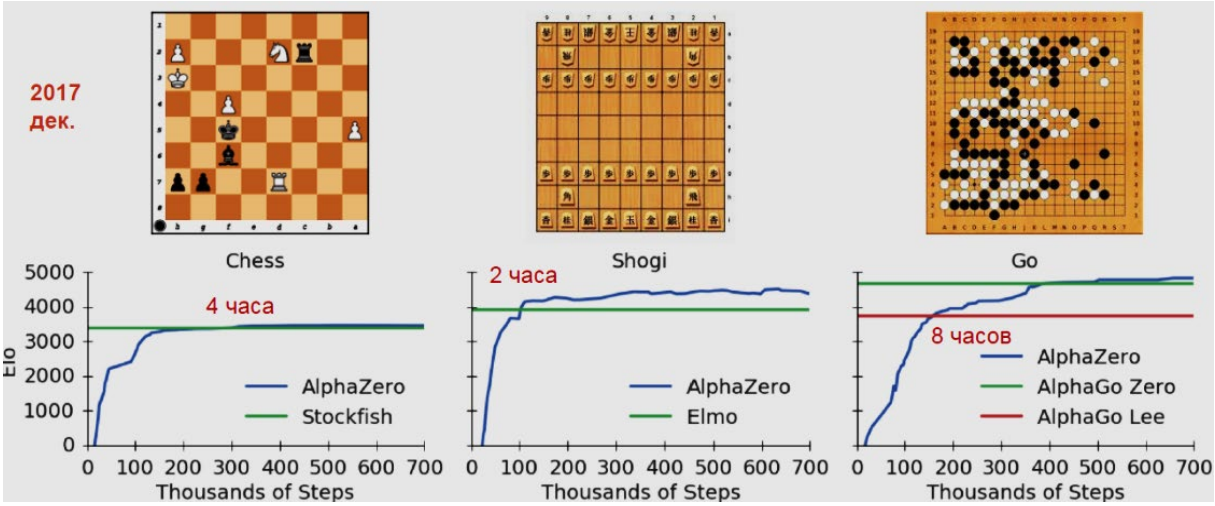

Рис. 3. Затраты времени «AlphaZero» на превышение уровня коэффициента Эло лучших программ для настольных логических игр

4. Социальная революция

Так же, как появление паровых машин, электричества и затем информационных технологий последовательно изменяли не только технологии производства, но и социальную жизнь людей, так и широкое внедрение систем и устройств ИИ приведёт к существенным изменениям. Многими признаётся, что «информация – это нефть XXI века», а ИИ – средства для промышленной обработки информации. Понимание важности развития ИИ привело к тому, что большинство развитых стран мира (включая Россию [4]) приняли государственные программы развития ИИ, а В.В. Путин, выступая 1 сентября 2017 г. перед ярославскими школьниками, сказал, что «Страна, добившаяся лидерства в создании искусственного интеллекта, будет властелином мира» [5].

Современный ИИ основан на узкоспециализированных нейросетях, очень глубоких, требующих длительного обучения и не приспособленных к применению чужого опыта. Он преимущественно используется для выполнения задач, решение которых известно, но требует внимания и использования значительных объёмов данных. «Интеллектуальность» таких задач можно оценивать по-разному, но в ближайшие годы узкий ИИ позволит замещать всё большее число людей техническими устройствами.

Высвобождаемые от монотонной работы люди могут быть использованы в творческих профессиях и управлении. Но для этого необходимо повышать уровень образования населения и делиться властными полномочиями. Это требует не только некоторых затрат, но и передачи части властных полномочий управляющими кастами. Западные управленцы не настроены на это и предпочитают вывести значительную часть населения из экономики путём выплаты им безусловного обязательного дохода (БОД), средства для которого могут быть получены за счёт повышения производительности труда при использовании ИИ.

Россия, отставшая в технологиях производства элементной базы информационной техники, могла бы пойти другим путём: развивать алгоритмические идеи совершенствования ИИ на основе сохранившегося в стране научного потенциала и готовить новые кадры для разработки и внедрения перспективных подходов к ИИ. Это дало бы стране определённые преимущества перед Западом, не считающим имеющийся у них человеческий потенциал большой ценностью.

Но в национальной программе «О развитии искусственного интеллекта в РФ» больший упор делается на формирование крупных проектов в рамках госкорпораций, а не на создание условий для вовлечения населения в работы по развитию ИИ. К тому же высший менеджмент страны ориентирован на следование западному курсу в социальной политике, что делает возможность РФ пойти своим путём маловероятной.

5. Сильный ИИ (AGI)

Значительные успехи в разработке узкого ИИ делают более обоснованным материалистический взгляд на создание сильного ИИ (artificial general intelligence, AGI), способного решать любые, доступные человеку интеллектуальные задачи. В то же время понятно, что просто количественное увеличение суммы решаемых узким ИИ задач не приведёт к созданию AGI. Необходимо преодолеть ряд сложностей описания реального мира:

- Сложность описания простых объектов и явлений;
- Сложность декомпозиции:
- Комбинаторная сложность выбора действий;
- Сложность построения целей;
- Сложность задания потребностей.

Наличие потребностей у агентов AGI необходимо для решения остальных сложностей, но важно, чтобы потребности можно было совмещать с целями сообщества, в котором агенты AGI будут жить. У человека потребности наследуются в структурах, называемых центрами удовольствия и совместимость с социумом обеспечивается отбором.

Аналогичный тонкий отбор должен быть организован для агентов AGI, хотя некоторые требования к закладываемым в AGI потребностям могут быть проанализированы и на теоретическом уровне. В любом случае AGI должен создаваться в виде агентов, совместимость которых с цивилизацией может обеспечиваться ограничением их возможностей, при которых они будут должны считаться с мнением социума по поводу выполняемых ими действий (или бездействия).

6. Выводы

Всё идёт к тому, что в ближайшие 5-10 лет будут созданы первые агенты AGI и межгосударственная (а также межкорпоративная…) конкуренция вступит в новую фазу. Если раньше соревновались сообщества из близких по возможностям субъектов, то включение агентов AGI может значительно усилить неравенство между сообществами.

Для предотвращения катастрофических последствий создания AGI следует не только заботиться о технических путях обеспечения дружественности отдельных агентов AGI к отдельным людям, но и о достижении понимания человечеством необходимости сохранения межгосударственной и межкорпоративной конкуренций в правовом поле в условиях появления агентов AGI. Только наличие конкуренции и ограничение власти отдельных субъектов позволит поддерживать выживание и прогресс цивилизации.

Работа выполнена при поддержке РФФИ (проекты 19-01-00602 и 20-511-00003).

Литература

1. *Гудфеллоу Я., Бенджио И., Курвилль А.* Глубокое обучение/ Пер. с англ. А.А.Слинкина/ 2-е изд., испр. – М.: ДМКПресс, 2018.
2. *Николенко С., Кадурин А., Архангельская Е.* Глубокое обучение. Погружение в мир нейронных сетей. – Питер. 2018.
3. *Silver D., Hubert T. et al.* A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play // Science. 2018. V.362, Is.6419, pp.1140-1144.
4. Указ Президента Российской Федерации от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации». URL: http://publication.pravo.gov.ru/Document/View/0001201910110003
5. https://tass.ru/obschestvo/4524746