УДК 004.8

ЭВОЛЮЦИЯ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ: ОТ ALPHAZERO К ALPHAPROOF И ИХ ПРИМЕНЕНИЕ В РЕШЕНИИ МАТЕМАТИЧЕСКИХ ЗАДАЧ

Курновский Р.М.

Джи Пи Морган Банк, Самара, e-mail: r.kurnovskii@gmail.com

В данной статье подробно рассматриваются алгоритмы машинного обучения, заложенные в основу системы AlphaZero, а также их применение для решения сложных математических задач в системе AlphaProof. Целью работы являлось определение математических правил работы алгоритмов, благодаря которым они столь эффективны. В статье также рассматриваются перспективы и вызовы, связанные с применением нейронных сетей в научных исследованиях, особенно в области математических доказательств. Для этого был проведен всесторонний обзор научных источников и систематизация данных исследований. Было выявлено, что эффективность моделей была связана с оптимизациями алгоритмов поиска по дереву Монте-Карло и разработки новых методов. AlphaProof использует методы обучения с подкреплением, разработанные на базе AlphaZero, которая изначально применялась для игр, таких как шахматы и го. Эти методы позволяют системе справляться с математическими задачами высокой сложности. Путем преобразования более миллиона задач из различных областей, включая алгебру, теорию чисел и геометрию, в формальные языки, такие как Lean, AlphaProof эффективно генерирует и проверяет решения, что делает ее мощным инструментом для математических исследований.

Ключевые слова: машинное обучение, нейронные сети, математические задачи, AlphaZero, AlphaProof

EVOLUTION OF MACHINE LEARNING METHODS: FROM ALPHAZERO TO ALPHAPROOF AND THEIR APPLICATION IN SOLVING MATHEMATICAL PROBLEMS

Kurnovskiy R.M.

J.P. Morgan, Samara, e-mail: r.kurnovskii@gmail.com

This paper provides a detailed account of the machine learning algorithms that underpin the AlphaZero system, together with an analysis of their application to the resolution of complex mathematical problems in the AlphaProof system. The objective of this paper is to identify the mathematical principles underlying the algorithms that make them so effective. Furthermore, the paper investigates the potential and obstacles to utilizing neural networks in scientific enquiry, particularly within the domain of mathematical proofs. To this end, a comprehensive review of the scientific literature and systematic organization of the research data were carried out. It was determined that the efficacy of the models was contingent upon optimizations of Monte Carlo tree search algorithms and the development of novel methodologies. AlphaProof employs reinforcement learning techniques derived from AlphaZero, which was initially deployed in games such as chess and Go. These techniques enable the system to address mathematical problems of considerable complexity. By transforming over a million problems from diverse domains, including algebra, number theory, and geometry, into formal languages like Lean, AlphaProof can efficiently generate and verify solutions, making it a valuable tool for mathematical research.

Keywords: machine learning, neural networks, mathematical problems, Alphazero, Alphaproof

Введение

25 июля 2024 г. команда Research компании Google DeepMind, занимающаяся разработкой и применением методов машинного обучения для решения математических задач, объявила о том, что их последние модели AlphaProof и AlphaGeometry 2 смогли решить задания сложнейшей международной математической олимпиады (65th International Mathematical Olympiad, IMO 2024) на уровне серебряного медалиста, отстав от порога для золотой медали на 1 балл [1]. Стоит учесть, что, в отличие от реальных участников олимпиады, решающих задачи 4,5 часа, нейросети справились лишь за 3 дня, но, несмотря на это, в скором времени ожидается многократное ускорение работы AlphaProof. Бурный рост нейронных сетей в самых разных прикладных и фундаментальных областях за последние несколько лет привлекает внимание многих специалистов. Одним из важных вопросов является вопрос эволюции и развития методов обучения.

Цель работы заключается в определении ключевых технических и математических особенностей работы алгоритмов AlphaZero, AlphaProof и похожих моделей, благодаря которым они столь успешно решают математические задачи.

Материалы и методы исследования

Для проведения исследования была осуществлена систематическая оценка и ана-

лиз научных публикаций, посвященных эволюции методов машинного обучения с акцентом на разработку и применение моделей, таких как AlphaZero и AlphaProof, в решении математических задач. Основным методом исследования стал литературный обзор, включающий поиск, отбор, классификацию и критический анализ научных статей, опубликованных в период с 2012 по 2024 г. Материалы для анализа были получены из международных научных журналов, включая Nature, IEEE, Science, а также публикаций Google DeepMind. Ключевыми словами поиска стали: AlphaZero, AlphaProof, «машинное обучение», «решение математических задач» «глубокое обучение». Были рассмотрены статьи, охватывающие как технические аспекты алгоритмов, так и их применимость в математике и смежных областях. Критериями отбора статей служили: актуальность исследований в контексте применения методов машинного обучения к математическим задачам, детальное описание архитектур моделей AlphaZero и AlphaProof. Для структурирования данных применялась методология PRISMA, которая позволила систематизировать процесс поиска, исключения дублирующихся данных и анализа релевантных источников.

Результаты исследования и их обсуждение

Все используемые обозначения и символы даны в таблице.

Используемые обозначения

p	вектор, каждая компонента $\boldsymbol{p}_a$ которого – это вероятность принять данное положение при данном действии
v	скаляр оценки результата
Θ	гиперпараметры нейросети
f_Θ	функция нейросети при данных гиперпараметрах
s	данное положение на доске (на примере игры в го)
a	произведенный переход по дереву (действие)
$Pr(a\|s)$	вероятность занять данное положение на доске при данном действии
z	скаляр, характеризующий результат игры
π_a	вектор вероятности произвести данный переход по дереву из данного начального положения
v_t	скаляр оценки результата игры на данном шаге
l	функция потерь в методе градиентного спуска
T	скаляр, характеризующий конечную позицию на дереве
c	регулирующий параметр
c_{puct}	скаляр, определяющий уровень исследованности данной ветви дерева
x	параметр ошибки функции потерь, отличающий алгоритм PBT от алгоритма AlphaZero

Все решения семейства Alpha компании DeepMind, включающие системы AlphaFold 2, AlphaZero, AlphaGo, AlphaStar, AlphaTensor, AlphaCode и др., направлены на решение вычислительных прикладных и научных задач. Рассматриваемая система AlphaProof представляет собой предварительно обученную на большой выборке математических задач и их решений модель с подкреплением AlphaZero, схему работы которой можно проиллюстрировать рис. 1. Более миллиона математических рукописных задач из всех областей геометрии, алгебры, теории вероятностей и теории чисел, математического анализа и других были переведены на формальный язык Lean с помощью Gemini [1]. Это решило большинство проблем с обработкой естественного языка, нейросетевых галлюцинаций и ошибок. Для решения геометрических заданий была значительно улучшена модель AlphaGeometry за счет большего количества задач для обучения.

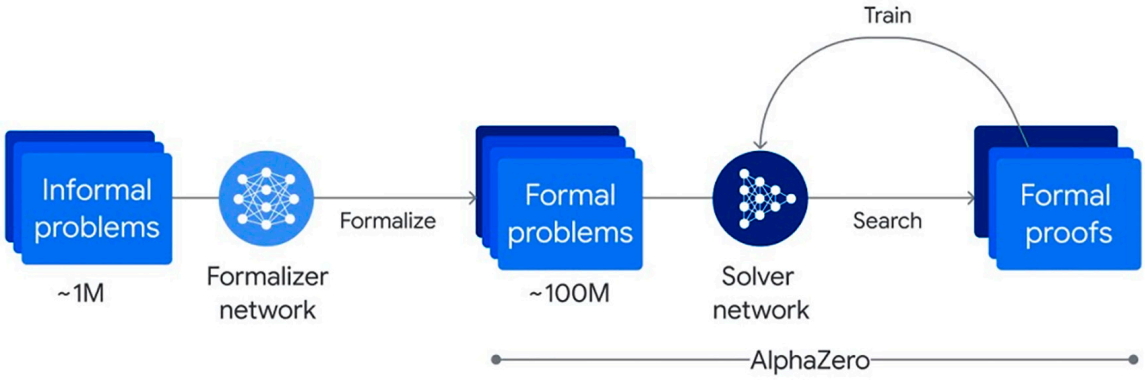

Рис. 1. Схематичное изображение процесса формализации математических задач, обучения и их решения с помощью AlphaZero в системе AlphaProof

AlphaZero – это нейронная сеть, которая обучается за счет соревнования сама с собой в течение многих миллионов попыток с подкреплением. Сначала процесс обучения случаен, но довольно быстро нейросеть учится корректировать свои параметры, причем намного более успешно, чем модели, обученные на заранее подготовленных данных. AlphaZero использует эвристический алгоритм поиска по дереву Монте-Карло (Monte Carlo Tree Search, MCTS) с оценкой НОД функциями на основе deep learning. Именно нейросетевой оценкой НОД этот алгоритм отличается от классического MCTS. Эта нейросеть тренируется предсказывать по прошлым данным дальнейшие данные (SL-policy network), потом тренируется играть сама с собой (RL-policy network), а далее тренируется предсказывать шансы на выигрыш [2]. В основе этого метода все еще лежат математические методы теории принятия решений (марковские процессы принятия решений и их расширения при частичных наблюдениях), теории игр и комбинаторика, метод Монте-Карло и искусственный интеллект в настоящих играх.

Алгоритм MCTS итеративно строит дерево поиска решения до достижения какого-то ограничения по памяти, времени, точности и т.п. Как и у множества других таких алгоритмов, итерации алгоритма производятся в четыре шага: выбор дочерних НОД, расширение количества НОД, моделирование и обновление статистики ошибок [3]. Иллюстрация алгоритма представлена на рис. 2.

Рассмотрим математические основы работы алгоритмов работы AlphaZero. Отметим, что выделенные полужирным начертанием символы – это векторные величины, если не указано иное.

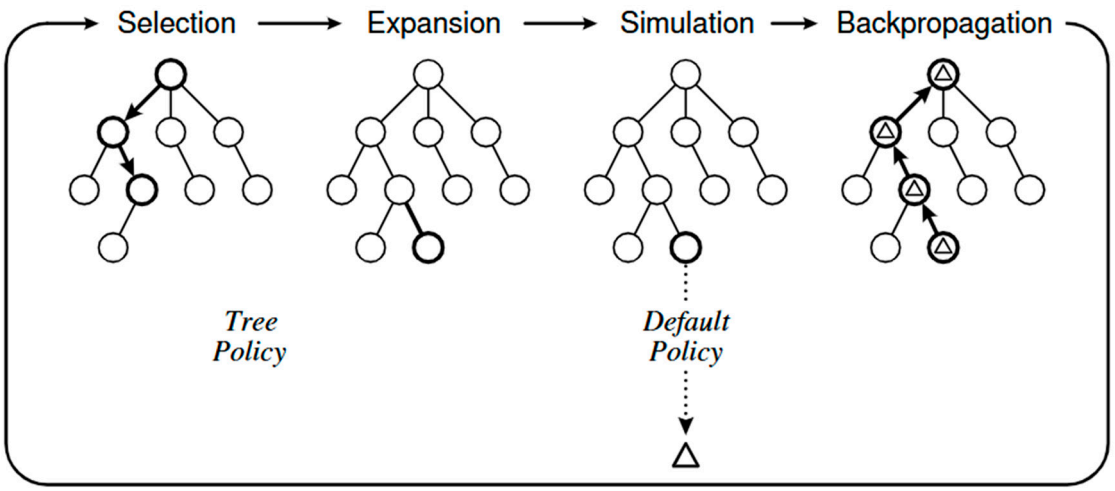

Рис. 2. Схематичное изображение основных шагов алгоритма Monte Carlo Tree Search

Авторы AlphaZero в работе [4] описывают используемую нейросеть с глубоким обучением как функцию $(\boldsymbol{p},v) = f_\Theta(s)$ с параметрами Θ. На примере с обучением игры в игру го, нейросеть $f_\Theta(s)$ принимает данное положение на доске s, а на выходе предоставляет вектор вероятности $\boldsymbol{p}$ с компонентами $\boldsymbol{p}_a = Pr(a|s)$ для каждого действия a и скаляр оценки v ожидаемого результата игры z из отношения $v \approx \mathbb{E}\left[z \middle| s \right]$. Параметры Θ подбираются при обучении с играми со случайно подобранными Θ благодаря подкреплению. В каждой игре с начальным положением в ходе поиска по ветви возвращается вектор $\boldsymbol{\pi}_a = Pr(a|s_0)$. Параметры нейронной сети постоянно обновляются, чтобы минимизировать разницу между величиной предсказанного результата игры v_t с реальным результатом z. Для этого параметры Θ корректируются путем градиентного спуска по функции потерь l:

$$l = (z - v)^2 - \pi^T log\left(\boldsymbol{p}\right) + c \,||\, \theta \,||^2, \quad (1)$$

где T – это конечная позиция на дереве, c – параметр, который контролирует регуляризацию этой функции.

В AlphaZero, так же как и в AlphaGo Zero, используется байесовская оптимизация гиперпараметров, но они, как и настройки сети и всего алгоритма, не изменяются от игры в игру [5]. Каждое ребро (s,a) в дереве поиска хранит набор статистических данных:

$$\left\{N\left(s,a\right), W\left(s,a\right), Q\left(s,a\right), P\left(s,a\right)\right\}, \quad (2)$$

где априорная вероятность $P(s,a)$, количество посещений $N(s,a)$, значение действия $Q(s,a)$, $W(s,a)$ – суммарное значение действия на ветви дерева. Каждое моделирование начинается с начального состояния s_0 и итеративно выбирает ходы, которые максимизируют верхнюю доверительную границу вида

$$a_t = argmax\left(Q\left(s_t,a\right) + U\left(s_t,a\right)\right), \quad (3)$$

где $U\left(s,a\right) \infty \left(s,a\right) / \left(1 + N\left(s,a\right)\right)$. В частности, в алгоритме PUCT предлагается такой вид функции $U(s,a)$:

$$U\left(s,a\right) = c_{puct} P\left(s,a\right) \frac{\sqrt{\sum_b N\left(s,b\right)}}{1 + N\left(s,a\right)}, \quad (4)$$

где c_{puct} – это константа, определяющая уровень исследованности данной ветви.

При прохождении по ребру (s,a) обновляется значение счетчика $N(s,a)$, а $Q(s,a)$ обновляется по правилу

$$Q\left(s,a\right) = \frac{1}{N\left(s,a\right) \sum_{s' | s,a \to s'} V\left(s'\right)}. \quad (5)$$

Здесь $s,a \to s'$ показывает, что симуляция достигла данного s' при выполнении хода a из состояния s.

Помимо приведенного выше облегченного математического объяснения работы алгоритма работы AlphaZero, полная математическая модель содержит большое количество вероятностных поправок и небольших, но многочисленных шагов корректировок.

После публикации статей AlphaZero несколько исследовательских команд, как в самой компании DeepMind, так и другие, проводили изучение способов и методов оптимизации алгоритмов AlphaZero. Так, исследователи из работы [6] в 2020 г. предложили собственный разработанный «популяционный метод» (PBT). Они использовали подходы, применяемые в машинном обучении для обработки генетических данных, а именно несколько нейросетей со случайными начальными параметрами Θ. Все сети объединяют информацию для улучшения гиперпараметров, а в случае, если они недостаточно точные, то происходит их прямая замена на лучшие гиперпараметры другой нейросети. Также следует отметить, что алгоритм предусматривает возможность ручного изменения гиперпараметров в научно-исследовательских целях.

В данном исследовании использовалось 16 нейросетей для оценок, причем функция потерь теперь также зависит и от параметра ошибки x между z и v:

$$l = x(z - v)^2 - \pi^T log\left(\boldsymbol{p}\right) + c \,||\, \theta \,||^2 . \quad (6)$$

В результате экспериментов авторы обнаружили, что, используя метод PBT к игре в го на поле 19×19, процент побед над другой нейросетью Facebook's ELF OpenGo v2, которая наиболее близка по мощности со свободными реализациями AlphaZero, составил 74 %. Это говорит о том, что оригинальное применение различных подходов в нейросетевых методах обучения имеет перспективу значительно увеличить способности таких соревновательных систем.

AlphaProof использует заранее обработанные Gemini данные, которые представляют собой формализованные с помощью Lean математические задачи. Это одна из слабых точек технологии и недостаток обучения моделей на естественном языке, из-за чего на настоящий момент невозможно избежать многочисленных искажений. Для решения конкретной математической задачи, AlphaProof необходимо формализовать ее, в соответствии с алгоритмом AlpaZero, описанным выше, генерируются сотни варианты решения этой задачи, а после про-

исходит проверка этого решения с помощью Lean, если оно идентифицируется как неверное, то начинается проверка следующего решения и т.д., до достижения правильного решения. Это возможно благодаря тому, что Lean – это функциональный язык программирования с зависимыми типами на основе CoC (Calculus of Constructions) и CiC (Calculus of Inductive Constructions) [7]. Версия Lean 4 поддерживает высокопроизводительные технологии управления памятью, что может значительно упростить процесс обучения таких сложных систем, как AlpaProof.

На данный момент еще нет публикаций с тестами результатов работы алгоритмов системы AlphaProof, неясно, есть ли отличия точности и скорости ее работы в зависимости от области математики задач, компания Google DeepMind не представила подробные данные о том, какими были промежуточные этапы обучения, о скорости обучения на различных задачах, требовались ли корректировки метода обучения. Эти данные ожидаются в ближайшем будущем с выходом второй версии системы, но уже сейчас становится понятным, что такое использование нейросетей будет большой и важной частью будущего развития математики, так как позволит с высокой точностью и скоростью строго проверять сложнейшие и объемные теоремы, также как и теоремы из областей математики, в которых большое количество абстрактных конструкций и идей – сейчас такие задачи непосильны существующим моделям искусственного интеллекта. Помимо доказательства теорем, это также будет большим прорывом не только для решения сложнейших задач, но и для их составления. О том, как такие нейросети внесут большой вклад в будущее математики на The Oxford Mathematics Public Lectures, рассказал ведущий мировой математик Теренс Тао, который видит в них незаменимый инструмент и роль коллеги при математических исследованиях.

Заключение

Все вышеизложенное позволяет заключить, что значительный прогресс в методах машинного обучения, улучшения и распространения применения AlphaZero в прикладных и фундаментальных исследованиях будет только увеличиваться. AlphaProof, в свою очередь, показывает перспективы в качестве сильного инструмента для ученого, что показывает прогресс в работах по оптимизации и улучшению алгоритмов таких систем.

Список литературы

1. AI achieves silver-medal standard solving International Mathematical Olympiad problems // Google DeepMind. [Электронный ресурс]. URL: https://deepmind.google/discover/blog/ai-solves-imo-problems-at-silver-medal-level/ (дата обращения: 04.09.2024).

2. David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, Demis Hassabis. Mastering the game of Go with deep neural networks and tree search // Nature. 2016. Vol. 529, Is. 7587. P. 484–489.

3. Cameron B. Browne, Edward Powley, Daniel Whitehouse, Simon M. Lucas, Peter I. Cowling, Philipp Rohlfshagen. A Survey of Monte Carlo Tree Search Methods // IEEE Trans. Comput. Intell. AI Games. 2012. Vol. 4, Is. 1. P. 1–43.

4. David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, Demis Hassabis. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play // Science. 2018. Vol. 362, Is. 6419. P. 1140–1144.

5. David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel & Demis Hassabis. Mastering the game of Go without human knowledge // Nature. 2017. Vol. 550, Is. 7676. P. 354–359.

6. Ti-Rong Wu, Ting-Han Wei, I-Chen Wu Accelerating and Improving AlphaZero Using Population Based Training. The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20). 2020. P. 1046–1053.

7. About Lean // Lean. [Электронный ресурс]. URL: https://lean-lang.org/about/ (дата обращения: 04.09.2024).