

ОБЗОР-РЕЦЕНЗИЯ НА СТАТЬЮ "What Can Transformers Learn In-Context? A Case Study of Simple Function Classes"

за авторством

Dimitris Tsipras, Shivam Garg, Percy Liang и
Greg Valiant

Зинов Александр Александрович

21 октября 2022 г.

1 Суть работы

После публикации статьи [1] от DeepMind оказалось, что текстовые трансформеры обладают способностью учиться решать простые задачи, которые не содержатся в обучающей выборке в явном виде, уже на этапе применения, то есть без непосредственной оптимизации. Данный феномен был назван *in-context learning* и с тех пор интересовал многих исследователей в области глубинного обучения. С применением данного метода ранее уже удавалось аппроксимировать байесовские модели [7]. В попытках глубже понять возникновение такого эффекта авторы рецензируемой статьи ставят задачу аппроксимации простых моделей машинного обучения таких как линейная модель или решающее дерево.

2 Контекст публикации

Статья была опубликована 1 августа 2022 года авторами из Стэнфордского университета Dimitris Tsipras (post-doctoral researcher под руководством Percy Liang и Greg Valiant) и Shivam Garg (PhD student под руководством Greg Valiant). До этого они не публиковали совместные статьи, ровно как и работы на тему *in-context learning*, изучая в своих публикациях самые разные области глубинного обучения с некоторым уклоном в сторону процессов оптимизации. Идея статьи совсем не случайна. Сама тема трансформеров в настоящее время заслуженно популярна, как новая архитектура позволяющая достигать SOTA качества на большом числе самых разных задач (а также революционировавшая в свое время область машинного обучения

для работы с текстами). Феномен in-context learning с исследовательской точки зрения является крайне необычным и действительно заслуживает направленного на него внимания.

3 Влияние на статью

Отправной точкой исследования можно назвать уже упомянутую статью [1], в которой предлагается архитектура GPT-3, демонстрирующая способности к in-context learning. В своем исследовании авторы также опираются на статью [6], в которой изучается используемый в рецензируемой статье подход к *meta-learning*, когда на этапе применения модели ей подаются непосредственно примеры решаемой задачи. Стоит также упомянуть одну из множества статей [5], предшествующих рецензируемой, в которых in-context learning исследуется непосредственно, как практический метод решения задач машинного обучения.

4 Цитирования и конкуренты

В силу того, что статья была опубликована совсем недавно, цитирование у нее пока одно [2]. В этой статье рассматриваются аналогичные эксперименты, но проводимые с архитектурой RNN. Авторы достигают значительных успехов и даже формулируют верхнюю оценку на ошибку их модели. В конкурирующей статье [8] феномен in-context learning исследуется более непосредственно, но с другой точки зрения байесовских моделей.

5 Сильные стороны работы

Авторы хорошо выбрали тему для статьи. In-context learning – актуальная, необычная и интересная тема для исследования. Статья сопровождается кодом для воспроизведения результатов, что как всегда добавляет работе ценности как с точки зрения легкости валидации результатов, так и с точки зрения дальнейшего развития темы на основе авторской кодовой базы. Эксперименты грамотно сформулированы и поставлены. Так на примере задачи линейной регрессии в статье демонстрируется, что трансформер вычисляет предсказания со SOTA качеством (least squares против kNN и простого усреднения). В приложении B.7 отдельное внимание уделяется выкладкам, показывающим, что способности трансформера вряд ли основаны на простом запоминании "обучающей" выборки. Различные способы зашумления как выборки и таргетов, так и распределения самих функций позволяют заключить, что предложенная модель действительно демонстрирует in-context learning способности.

6 Слабые стороны работы

Критика работы строится на двух основных пунктах. Как утверждают авторы статьи [3], скорее всего in-context learning не имеет смысла использовать на практике, поскольку альтернативный подход prompt tuning итак не требует больших затрат по времени на обучение, в то время как зачастую достигает лучшего качества при значительно меньшем потреблении ресурсов на этапе применения – нет необходимости заново подавать на вход модели всю обучающую выборку для получения предсказаний. В то же время не до конца ясны (поскольку явно не сформулированы) исследовательские выводы, которые авторы делают из проделанной работы. Сама идея умения трансформеров проводить in-context learning не нова: здесь авторы разве что добавляют к дискуссии формализованные эксперименты, подтверждающие данное эмпирическое наблюдение. А вот связь результатов работы с in-context learning текстовых трансформеров вроде GPT-3 в тексте статьи так нигде и не обсуждается, оставляя заявленную цель глубже разобраться в данном феномене нереализованной.

7 Дальнейшая работа

Авторы предлагают несколько направлений дальнейшей работы, приведем из них наиболее интересные. В первую очередь с помощью варьирования таких параметров, как сложность класса предсказываемых функций, размер модели и длина последовательности примеров применения функции позволяет сделать выводы касательно общих свойств значения ошибки такой модели. Подобное исследование может быть особенно ценно в контексте применения подобного же рассуждения к текстовым трансформерам вроде GPT-3, с целью оценки снизу их общего качества и производительности в in-context задачах. Авторы также отмечают особенно интригующую возможность использования encoder-decoder архитектуры трансформера для того, чтобы понимать непосредственно закодированный в модели алгоритм решения мета задачи. Это открывает будущим исследователям потенциальную возможность обнаружить ранее неизвестные алгоритмы решения привычных задач, которые при этом возможно будут не так интуитивны, но смогут показывать лучшее качество.

В рамках дальнейшей работы я хочу предложить исследование влияния порядка значений функции в каждом объекте. Как известно, данный параметр оказывает значительное влияние не процесс обучения трансформера. Авторы статьи [4] даже демонстрируют, что это влияние может быть решающим для достижения SOTA качества на определенных задачах. Таким образом было бы интересно исследовать не затронутый в рецензируемой статье вопрос: как порядок примеров внутри одного объекта влияет на качество модели? Можно ли таким образом улучшить авторские результаты? Можно ли на основе алгоритма оптимального упорядочивания сделать выводы о феномене in-context learning в целом?

Список литературы

- [1] Tom B. Brown и др. *Language Models are Few-Shot Learners*. 2020. arXiv: 2005.14165 [cs.CL].
- [2] Surbhi Goel и др. *Recurrent Convolutional Neural Networks Learn Succinct Learning Algorithms*. 2022. DOI: 10.48550/ARXIV.2209.00735. URL: <https://arxiv.org/abs/2209.00735>.
- [3] Haokun Liu и др. *Few-Shot Parameter-Efficient Fine-Tuning is Better and Cheaper than In-Context Learning*. 2022. DOI: 10.48550/ARXIV.2205.05638. URL: <https://arxiv.org/abs/2205.05638>.
- [4] Yao Lu и др. *Fantastically Ordered Prompts and Where to Find Them: Overcoming Few-Shot Prompt Order Sensitivity*. 2021. DOI: 10.48550/ARXIV.2104.08786. URL: <https://arxiv.org/abs/2104.08786>.
- [5] Sewon Min и др. *MetaICL: Learning to Learn In Context*. 2021. DOI: 10.48550/ARXIV.2110.15943. URL: <https://arxiv.org/abs/2110.15943>.
- [6] Nikhil Mishra и др. *A Simple Neural Attentive Meta-Learner*. 2017. DOI: 10.48550/ARXIV.1707.03141. URL: <https://arxiv.org/abs/1707.03141>.
- [7] Samuel Müller и др. *Transformers Can Do Bayesian Inference*. 2021. DOI: 10.48550/ARXIV.2112.10510. URL: <https://arxiv.org/abs/2112.10510>.
- [8] Sang Michael Xie и др. *An Explanation of In-context Learning as Implicit Bayesian Inference*. 2021. DOI: 10.48550/ARXIV.2111.02080. URL: <https://arxiv.org/abs/2111.02080>.