Research report and proposal - Gideon Stein

Research title:
"Generative models for action sequence generation in reinforcement learning settings"

"Генеративные модели для синтеза последовательностей действий в постановке обучения с подкреплением"

UDK: 263851

Keywords: Reinforcement Learning, Sequence Generation, Transformer, NLP

Introduction

The goals of this research project are the development and the evaluation of a new reinforcement learning agent action model architecture based on sequence generation models. Recent developments in NLP suggest, that sequence generation models based on attention mechanisms can outperform RNN alternatives on various tasks. The text translation architecture "Transformer" and the recently published promising results of the model GPT2 [1] are only two examples of these kinds of models. Classically, RNN models are often used in reinforcement learning agent action modeling. Based on these facts, the goal of this research will be the deployment of sequence generation models initially used in NLP into a reinforcement learning setup. Furthermore, the newly created model is hoped to have the ability to outperform RNN and other state-of-the-art solutions in multiple environments. Especially, the recently published GPT2 model [1] suggests, that the structure is very flexible for many different NLP tasks, which suggests that it is effective for other sequence generation tasks as well. While the goal of this research will be novel, there is still some basic literature that suggests, that the development proposal will be successful [2,3]. This general goal can be split into multiple sub goals. Firstly, well performing sequence generation models are commonly used in NLP. There is a need to transform them slightly for them to work in a different environment. Especially, the size of these models might need to be changed. Since GPT2 has approximately 1.5Billion weights, a smaller model, that includes the basic structure of GPT2 is worth looking into. Secondly, a finished model must be evaluated on its performance and whether it can outperform state-of-the art models. Finally, the results should be formulated in the form of a scientific paper.

Content of the research

The transformer architecture which was released in 2017 [4] brought new SOTA results to NLP in multiple domains. Based on this initial success, multiple variations of the transformer were developed which work for even more domains in NLP. Currently, transformer-based models such as GPT2[1], BERT [5], TransformerXL[9] and also the classic transformer hold almost all SOTA results in NLP. Transformer architectures have multiple advantages over the previously used RNN alternatives. Firstly, their use of the attention mechanism allows them to understand long term dependencies better than classical RNN alternatives. Additionally, since they feature no recurrent units, they can skip classical RNN problems like vanishing gradient etc. Another important fact about transformer-based models is, that they generate tokens one after another and refeed the generated sentence as the new input for the next generation which seems a very flexible approach.

In the context of deep reinforcement learning, more specifically deep reinforcement learning, the currently used SOTA architectures almost always use RNNs in order to process problems that have a time or sequence dimension. While this usage is intuitive**, it is estimated that replacing the RNN component with a transformer-based architecture can lead to significant performance improvements of reinforcement learning agents.** During this research, this thesis will be evaluated.

Since the requirements for the architecture are only that they attend to the past and generate a new action for an agent, multiple transformer-based architectures can be treated as candidates, excluding only encoder-based architectures like BERT [5]. **The evaluation of the potential of different transformer-based models will therefore be a core feature of this research.**

While the research is only at the beginning, certain starting steps were conducted to prepare further development. Firstly, research on the relevant topics that are connected to the aim of the project was conducted. These included the transformer architecture [4]. GPT2[1] and multiple developments in the field of state-of-the-art general-purpose text generation models [1,5,6,7] as well as reinforcement learning basics [8] and state-of-the-art model architectures [10]. Secondly, the most popular transformer-based models (Transformer, GPT2) were implemented and experimented with, using multiple libraries. Due to the complexity of the transformer models, there are various possibilities for bugs. Additionally, the training of these models is non-trivial. It requires specific learning rate schedules and Hyperparameter settings in order to achieve good results. This is one of the reasons why transformers are currently not popular for reinforcement learning. Experience on training the models were collected, which will prove essential for further steps.

Based on this, the next step will be to deploy the transformer-based models to a reinforcement learning setup to test their abilities as agent action generators. These steps are currently conducted.

Conclusions

While the project is still going on, the preparations for the creation and testing of a new model were successfully finished. With this baseline in place, future goals were set for the ongoing project, including the creation and the testing of multiple transformer-based models in different environments. Concerning the completed project, it is anticipated to achieve relevant findings in the field of reinforcement learning and specifically in the field of deep reinforcement learning agent action generation. This area is a popular field of research currently, which has a lot to discover. A contribution to this venture in a relevant manner will be achieved. All these findings will be summed up in a paper that will be the final product of this research.

TRANSLATION (Needs further improvement)

Вступление

Целями данного исследовательского проекта является разработка и оценка новой модели - модели действий обучающего агента подкрепления, основанной на моделях генерации последовательностей. Недавние разработки в NLP предполагают, что модели генерации последовательностей, основанные на механизмах внимания, могут превзойти альтернативы RNN в различных задачах. Архитектура перевода текста «Трансформер» и недавно опубликованные многообещающие результаты модели GPT2 [1] являются лишь двумя примерами таких моделей. Классически, модели RNN часто используются в моделировании действий обучающего агента подкрепления. Основываясь на этих фактах, целью данного исследования будет развертывание моделей генерации последовательностей, первоначально использованных в НЛП, в обучающей установке подкрепления. Кроме того, надеется, что вновь созданная модель сможет превзойти RNN и другие современные решения в различных средах. В частности, недавно опубликованная модель GPT2 [1] предполагает, что структура очень гибкая для многих различных задач НЛП, что говорит о ее эффективности и для других задач генерации последовательности. Хотя цель этого исследования будет новой, все еще имеется некоторая базовая литература, которая предполагает, что предложение о разработке будет успешным [2,3]. Эта общая цель может быть разделена на несколько подцелей. Во-первых, в НЛП обычно используются хорошо работающие модели генерации последовательностей. Необходимо немного их трансформировать, чтобы они работали в другой среде. Особенно размер этих моделей возможно потребуется изменить. Поскольку вес GPT2 составляет примерно 1,5 миллиарда, стоит рассмотреть модель меньшего размера, включающую базовую структуру GPT2. Во-вторых, готовая модель должна оцениваться с точки зрения ее характеристик и того, может ли она превзойти современные модели. По итогу, эти результаты находятся в процессе.

Содержание исследования

Архитектура трансформатора, выпущенная в 2017 году [4], принесла новые результаты SOTA для NLP в нескольких областях. Основываясь на этом первоначальном успехе, было разработано несколько вариантов трансформатора, которые работают для еще большего количества доменов в НЛП. В настоящее время модели на основе трансформаторов, такие как GPT2 [1], BERT [5], TransformerXL [9], а также классический трансформатор, содержат почти все результаты SOTA в НЛП. Архитектуры трансформаторов имеют множество преимуществ по сравнению с ранее использованными альтернативами RNN. Во-первых, использование ими механизма внимания позволяет им лучше понимать долгосрочные зависимости, чем классические альтернативы RNN. Кроме того, поскольку они не содержат повторяющихся единиц, они могут пропустить классические проблемы RNN, такие как исчезновение градиента и т. Д. Еще один важный факт о моделях на основе трансформаторов заключается в том, что они генерируют токены один за другим и ссылаются на сгенерированное предложение как новый вход для следующего поколения. который кажется очень гибким подходом.

Поскольку требования к архитектуре заключаются только в том, что они обращают внимание на прошлое и генерируют новое действие для агента, множественные

архитектуры на основе преобразователей могут рассматриваться как кандидаты, исключая только архитектуры на основе кодировщика, такие как BERT [5]. Поэтому оценка потенциала различных моделей на основе трансформаторов будет основной характеристикой этого исследования.

Пока исследование только в начале, были предприняты определенные начальные шаги для подготовки дальнейшего развития. Во-первых, были проведены исследования по актуальным темам, связанным с целью проекта. К ним относится трансформаторная архитектура [4]. GPT2 [1] и многочисленные разработки в области современных универсальных моделей генерации текста [1,5,6,7], а также основ обучения с подкреплением [8] и современного уровня техники модельные архитектуры [10]. Во-вторых, наиболее популярные модели на основе трансформаторов (Transformer, GPT2) были реализованы и экспериментированы с использованием нескольких библиотек. Из-за сложности моделей трансформаторов существуют различные возможности для ошибок. Кроме того, обучение этим моделям нетривиально. Для достижения хороших результатов требуются специальные графики скорости обучения и настройки гиперпараметра. Это одна из причин, почему трансформаторы в настоящее время не популярны для обучения с подкреплением. Был накоплен опыт обучения моделям, что окажется необходимым для дальнейших шагов.

Исходя из этого, следующим шагом будет развертывание моделей на основе трансформаторов в настройке обучения подкрепления для проверки их способностей в качестве генераторов действий агентов. Эти шаги в настоящее время проводятся.

Выводы

Пока проект продолжается, подготовка к созданию и тестированию новой модели была успешно завершена. С этим базовым уровнем были определены будущие цели для текущего проекта, включая создание и тестирование нескольких моделей на основе трансформаторов в различных средах. Что касается завершенного проекта, ожидается, что он достигнет соответствующих результатов в области обучения с подкреплением и, в частности, в области генерирования действий агента обучения с глубоким подкреплением. Эта область в настоящее время является популярной областью исследований, которая может многое открыть. Вклад в это предприятие соответствующим образом будет достигнут. Все эти выводы будут обобщены в документе, который станет конечным продуктом этого исследования.

List of abbreviations and conventions

ML - Machine Learning
NLP – Natural Language Processing
GPT2 - General Purpose Transformer 2
BERT- Bidirectional Encoder Representations from Transformers
ITMO - Information Technologies, Mechanics and Optics
RNN – Recurrent Neural Networks

References

1. Radford A. et al. Language models are unsupervised multitask learners //OpenAI Blog. – 2019. – Т. 1. – №. 8.

2. Upadhyay, Uddeshya, et al. Transformer Based Reinforcement Learning For Games.// arXiv preprint arXiv:1912.03918. – 2019.

3. Parisotto, Emilio, et al. Stabilizing Transformers for Reinforcement Learning. // *arXiv preprint arXiv:1910.06764*. – 2019.

4. Vaswani A. et al. Attention is all you need //Advances in neural information processing systems. – 2017. – С. 5998-6008.

5. Devlin J. et al. Bert: Pre-training of deep bidirectional transformers for language understanding //arXiv preprint arXiv:1810.04805. – 2018.

6. Trivedi (2019). OpenAI GPT-2 writes alternate endings for Game of Thrones. [online] Available at: https://towardsdatascience.com/openai-gpt-2-writes-alternate-endings-for-game-of-thrones-c9be75cd2425 [Accessed 27 Jun. 2019]

7. Kaggle (2019). Generating Long Sequences with Sparse Transformers. [online] Available at: https://www.youtube.com/watch?time_continue=197&v=se4ZM0es924 [Accessed 27 Jun. 2019]

8. Sutton, Richard S., and Andrew G. Barto. Introduction to reinforcement learning. Vol. 2. No. 4. // Cambridge: MIT press. – 2019.

9. Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q. V., & Salakhutdinov, R. Transformer-xl: Attentive language models beyond a fixed-length context. // *arXiv preprint arXiv:1901.02860*. – 2019

10. Kapturowski, S., Ostrovski, G., Quan, J., Munos, R., & Dabney, W. Recurrent experience replay in distributed reinforcement learning. ICLR. – 2019