In-Context Learning with Decision Transformers

Koo Jin Woo, Lee Jung Woo
Seoul National University

jinukoo@snu.ac.kr, junglee@snu.ac.kr

Decision Transformer 를 이용한 In-Context Learning

구진우, 이정우
서울대학교

Abstract

In this work, we tried to apply in-context learning which has brought huge success in the field of natural language processing to offline reinforcement learning. We pre-trained a Decision Transformer model with a non-expert dataset and by giving a segment of expert dataset as a context, we achieved similar or better performance in Hopper and HalfCheetah tasks in Gym-MuJoCo domain.

Ⅰ. Introduction

Reinforcement Learning aims to find the optimal policy to achieve maximum return in a certain environment and a task. There have been many methods to solve this problem. The most popular way is to estimate the value function to obtain the policy by choosing the action that has the maximum value. Offline reinforcement learning is a problem of finding the optimal policy and from a given, fixed dataset of interactions, without exploration.[1]

Decision Transformer viewed this offline reinforcement learning problem as a sequence modeling problem.[2] Instead of estimating the value function, it uses the powerful generalizing ability of transformer architecture to learn the policy. Decision Transformer shows performance on par with state-of-the-art offline reinforcement learning algorithms, but they lack stitching ability.[6, 7] It is because Decision Transformer does not have reason to output the optimal action; It outputs the most probable action according to the trained dataset.[8]

After GPT3 was introduced, in-context learning has been a popular method in the field of natural language processing.[3] In-context learning is a method of inferencing a pre-trained model with a few input and output pairs to obtain good performance on downstream tasks, which are not trained in the pre-training step. It has the advantage of achieving similar or better results compared to fine-tuning models, even without additional fine-tuning step.

This work tried to apply in-context learning in the field of offline reinforcement learning by using Decision Transformers. Since Decision Transformers and language models that were successful in-context learners have transformer architecture in-common, we hypothesized that in-context learning would also improve performance in reinforcement learning. Although in-context learning is a meta-learning algorithm, we thought that this method can be applied to single task and solve the optimality problem that Decision Transformers have.

By providing additional expert demonstrations to Decision Transformer models pre-trained with non-expert dataset, the model achieved similar or better performance on some control tasks.

Ⅱ. Method

To apply in-context learning, we first pre-trained a Decision Transformer model. We used the official code the authors of the Decision Transformer paper had released. We thought that the maximum context length of the original Decision Transformer model, which was 20, was too small for in-context learning. Therefore, we pre-trained the model with the context length of 50.

The model was pre-trained for 10 epochs, using D4RL dataset with the same parameters as the original paper suggested.[4] We experimented on 3 control tasks in Gym-MuJoCo domain: Hopper, HalfCheetah, and Walker2d. We used medium and medium-replay datasets to pre-train the model.

After pre-training, we selected the highest rewarding 20-timestep trajectory segment from the expert dataset. Then, we evaluated the pre-trained model with the segment prefixed, giving the model context of an expert policy. The results are shown in Table 1. The measurements are sum of total rewards until the episode ends or 1000 timesteps have passed.

Table 1. Experiment results

		HalfCheetah		Hopper		Walker2d	
		Medium	Medium-replay	Medium	Medium-replay	Medium	Medium-replay
Regular Inference	mean	4917.1	4243.5	2023.2	2054.8	3105.0	2757.0
	std	467.8	800.5	443.4	588.4	787.3	980.7
In-Context Learning	mean	4976.7	4200.7	2575.1	2082.7	3001.1	1695.0
	std	396.8	706.7	580.6	368.3	871.7	1107.5

Compared to regular inference, in-context learning showed similar or better performance on HalfCheetah and Hopper tasks. Especially, in-context learning brought about 25% increase of total return for Hopper pre-trained with medium dataset.

However, in Walker2d domain, in-context learning brought decrease in performance. This could be because of the model size. The pre-trained model has only 727695 parameters, while language models that showed emergent abilities by in-context learning have very large number of parameters, greater than 100 billion.[5]

Ⅲ. Conclusion

In this work, we tried to apply in-context learning in the field of offline reinforcement learning. We pre-trained Decision Transformer in 3 control tasks and applied in-context learning by additionally giving segment of an expert dataset when inferencing. Although it showed similar or better performance in 2 control tasks, further scaling the size of pre-trained model or training the model in a multi-task scenario can bring a better result.

ACKNOWLEDGMENT

This work is in part supported by National R&D Program through the National Research Foundation of Korea (NRF, 2021M3F3A2A02037893), Institute of Information & communications Technology Planning & Evaluation (IITP, 2021-0-00106) grant funded by the Ministry of Science and ICT (MSIT), INMAC, and BK21 FOUR program.

REFERENCES

[1] Levine, Sergey, "Offline Reinforcement Learning: Tutorial, Review, and Perspectives on Open Problems," 2020, (https://arxiv.org/pdf/2005.01643.pdf).

[2] Chen, Lili, "Decision Transformer: Reinforcement Learning via Sequence Modeling," 2021, (https://arxiv.org/pdf/2106.01345.pdf).

[3] Brown, B. Tom, "Language Models are Few-Shot Learners," Advances in neural information processing systems 33 (2020), pp. 1877-1901, 2020.

[4] Fu, Justin, "D4RL: Datasets for Deep Data-Driven Reinforcement Learning", 2020, (https://arxiv.org/pdf/2004.07219.pdf).

[5] Wei, Jason, "Emergent Abilities of Large Language Models", Transactions on Machine Learning Research, 2022.

[6] Yamagata, Taku, "Q-learning Decision Transformer: Leveraging Dynamic Programming for Conditional Sequence Modelling in Offline RL", 2022, (https://arxiv.org/pdf/2209.03993.pdf).

[7] Wu, Yeuh-Hua, "Elastic Decision Transformer", 2023, (https://arxiv.org/pdf/2307.02484.pdf)

[8] Lee, N. Jonathan, "Supervised Pretraining Can Learn In-Context Reinforcement Learning", 2023, (https://arxiv.org/pdf/2306.14892.pdf)