Decision Transformer Performance Analysis on CartPole-v1

Author: Jaeha Lee (The Ohio State University)

Abstract

This experiment investigates how the performance of a Decision Transformer varies with three primary factors: the amount of expert data, the number of training epochs, and the transformer model size. Using the offline CartPole-v1 dataset collected from a PPO expert policy, we trained multiple transformer configurations to predict the next action given the return-to-go (RTG), 4-dimensional state, and previous action. Results demonstrate that the balance between model capacity and dataset size critically determines generalization performance.

Method

We implemented a Decision Transformer using PyTorch, following the structure described in Chen et al. (2021). The input consisted of the return-to-go, state, and previous action, and the model was trained to predict the next action in an offline imitation learning setting. The dataset was generated from a PPO agent trained using the Stable-Baselines3 library for 20,000 timesteps. We then trained the transformer using varying Data Fractions, epoch counts, and model sizes as shown below:

Parameter	Values
Data fraction	25%, 50%, 100%
Epochs	3, 5, 10
Transformer size	Small: 64 dimensions, 2 layers, 2 heads Medium: 128 dimensions, 3 layers, 2 heads Large: 256 dimensions, 4 layers, 4 heads

Results

The following graphs show how each experimental variable affected the average return:

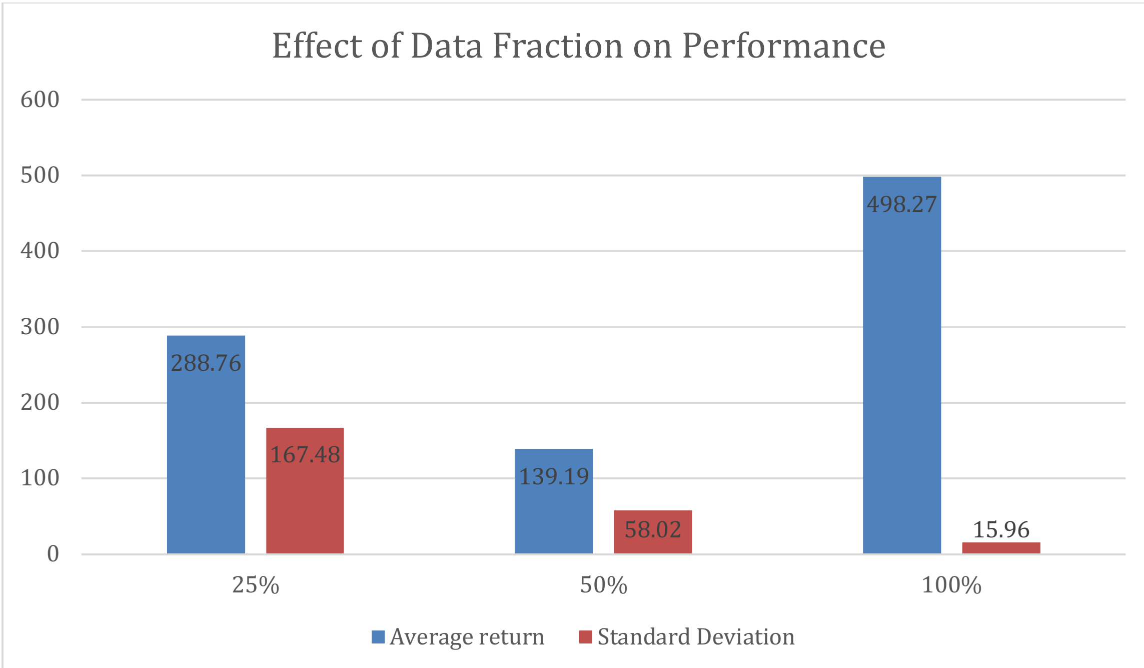

Figure 1. Effect of Data Fraction on Performance

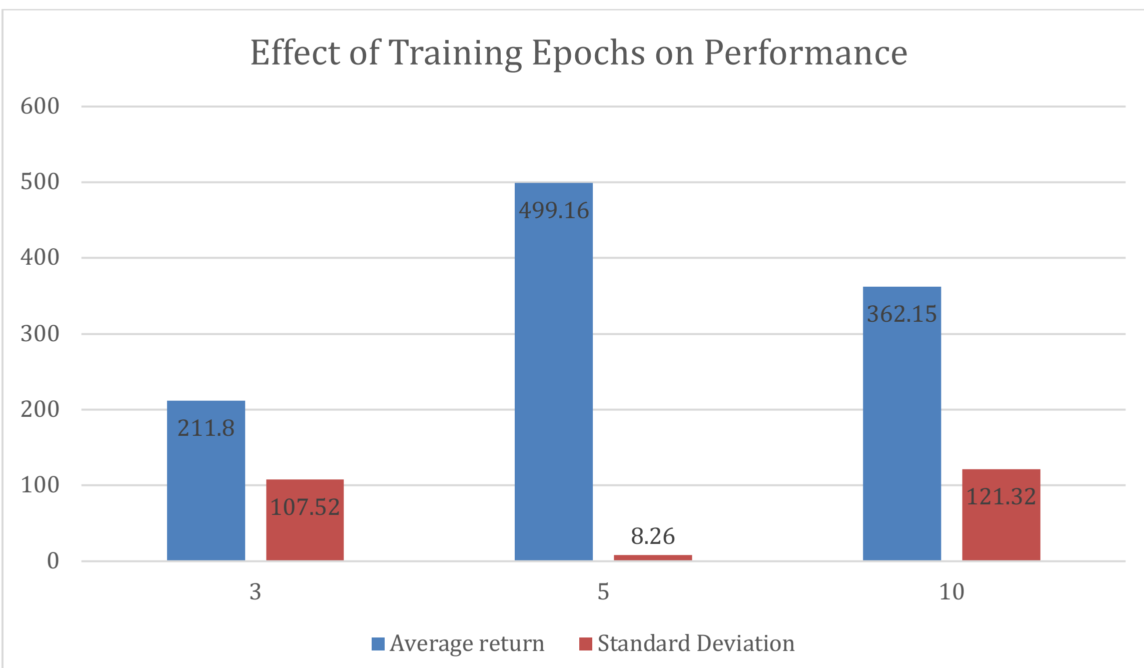

Figure 2. Effect of Training Epochs on performance.

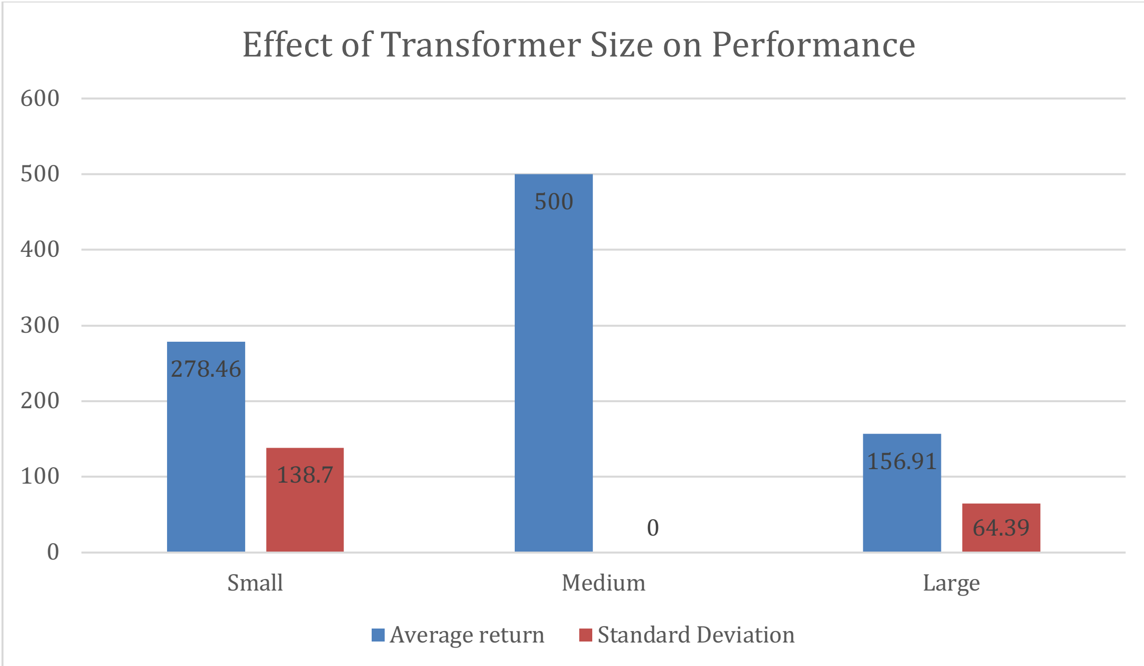

Figure 3. Effect of Transformer Size on Performance.

Discussion

The results indicate that the Decision Transformer's performance depends strongly on data size, model capacity, and the number of training epochs. The model achieved the best performance when trained with the full dataset, a medium-sized transformer, and five training epochs. Larger models exhibited overparameterization, meaning they had too many parameters relative to the dataset, which led to overfitting and unstable optimization. Conversely, smaller models lacked sufficient representational power, resulting in underfitting. These findings suggest that balanced model complexity is essential for stable offline reinforcement learning.

Conclusion

For the CartPole-v1 environment, the best performing configuration was: Data Fraction: 100%, Epochs: 5, Transformer Size: Medium (128 hidden, 3 layers, 2 heads), achieving an average return of $\approx 500 \pm 0$ over 100 episodes. This demonstrates that the Decision Transformer can reproduce near-optimal expert performance when properly tuned.

References

Chen, L., Lu, K., Rajeswaran, A., Lee, K., Grover, A., & Mordatch, I. (2021). Decision Transformer: Reinforcement Learning via Sequence Modeling. NeurIPS 2021.

Stable-Baselines3: https://github.com/DLR-RM/stable-baselines3

GitHub Repository: https://github.com/dlwogk7939/decision-transformer-cartpole

Original Decision Transformer Implementation: https://github.com/kzl/decision-transformer