A Trajectory is Worth Three Sentences: Multimodal Transformer for Offline Reinforcement Learning (Supplementary Material)

Yiqi Wang[1] **Mengdi Xu**[2] **Laixi Shi**[1] **Yuejie Chi**[1]

[1]Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
[2]Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

The following information is provided to support discussion in the paper: (A) hyperparameter and training comparisons between the Decision Transformer [Chen et al., 2021] (DT), DT-large, and our Decision Transducer (DTd); (B) evaluations curves of DT-large and DTd on medium-expert, medium, medium-replay of the D4RL benchmark [Fu et al., 2020].

A HYPER-PARAMETERS FOR DT-LARGE AND DTD

In Table 1, we have listed the architecture difference between DT-large and DTd.

Table 1: **Hyperparameters**. While DT, DT-large, and DTd's Join net are all temporal transformers [Radford et al., 2019] taking multimodal input, their input modalities varies according to the design.

Models	size	dimension	Modality Encoders			Biasing			Combiner		Joint Encoder		
			layers	heads	modality	layers	heads	modality	layers	modality	layers	heads	modality
DTd	2.52M	128	3	1	uni-modal	1	1	bi-modal	1	bi-modal	1	2	bi-modal
DT	0.7M	128	\	\	\	\	\	\	\	\	3	1	tri-modal
DT-large	2.41M	213	\	\	\	\	\	\	\	\	4	3	tri-modal

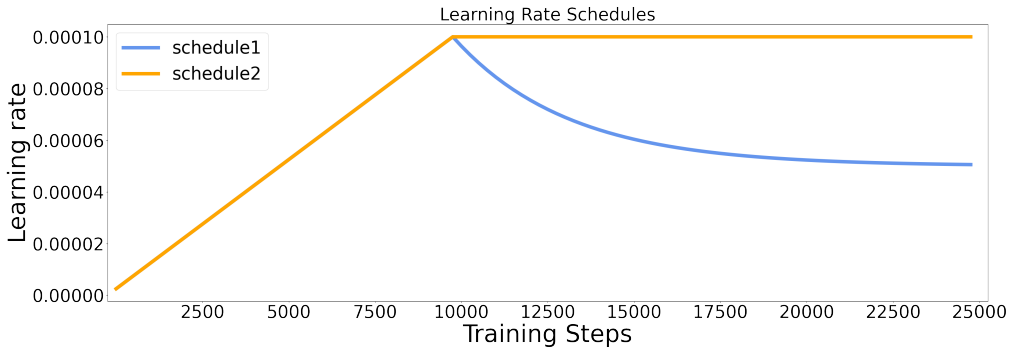

Figure 1: **Learning rate (lr) schedule**. DTd and DT-large are trained with the same number of gradient steps. However, on small dataset, DTd's lr (schedule 1) will be exponentially decayed after linear warm-up while DT-large don't (schedule 2). Since DTd is much more sample efficient than DT, a decaying lr will present it to overfit the data.

DT-large and DTd were both trained with 25k gradients steps of 256 batch size before evaluations. We found DTd learns faster than DT and may overfit the smaller dataset from D4RL benchmark [Fu et al., 2020] such as medium and medium-replay.

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

Therefore, we provide an exponentially decay learning rate schedule for DTd on smaller dataset other than medium-expert. Both DT-large and DTd has a peak learning rate 0.0001. We provide the learning rate curve throughout the training process as in Figure 1. DT-large always use schedule 2 provided by Chen et al. [2021] while DTd use schedule 2 when dataset is large (medium-expert) and choose exponentially decaying schedule 1 when dataset is small to avoid potential overfitting.

B EVALUATIONS CURVES OF DT-LARGE AND DTD

The normalized scores on D4RL benchmark [Fu et al., 2020] across 4 runs with independent training seed per run and 3 different evaluation seeds was plotted across all 3 environments (hopper, walker2d, halfcheetah) and across all dataset (medium-expert, medium, medium-replay). Both DT-large and DTd was trained 25k gradient steps with a batch size of 256 and was evaluated every 250 gradients steps following the protocol we discussed above. Curves are the average result across 4 runs and the shaded area corresponds to the standard deviation. As shown in the Figure 2, DTd is more sample efficient than DT on medium-expert data across all environment but such an advantage is not consistently observed on medium and medium-replay dataset across 3 environments as shown in Figure 3 and Figure 4.

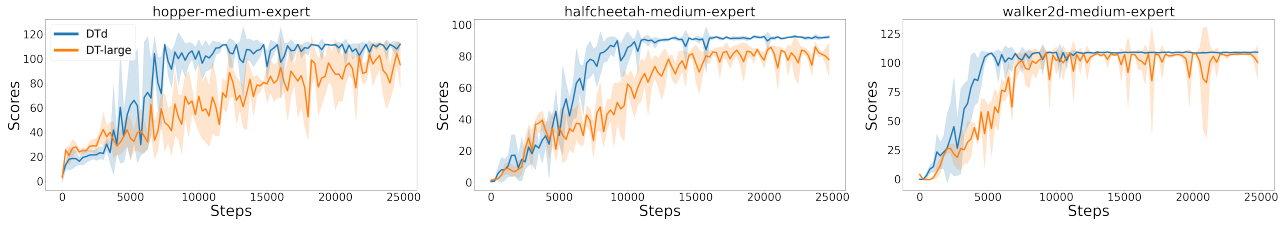

Figure 2: Evaluation Curve on Medium-Expert Dataset

.

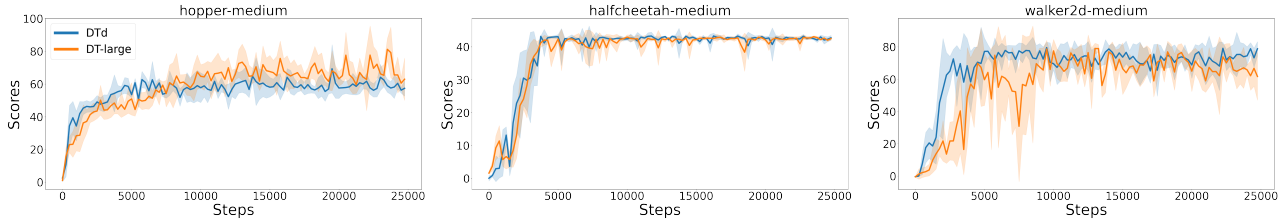

Figure 3: Evaluation Curve on Medium Dataset

.

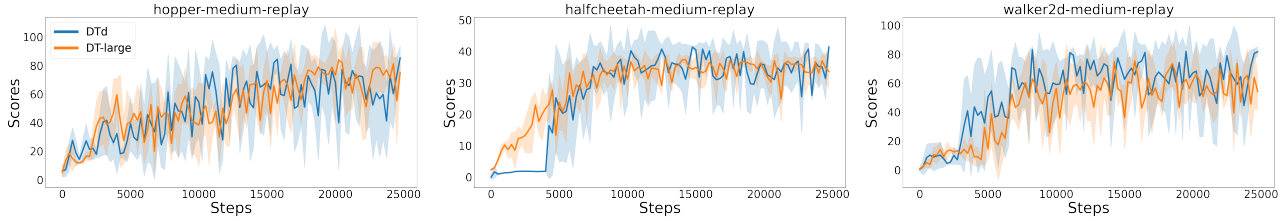

Figure 4: Evaluation Curve on Medium-Replay Dataset

.

References

Lili Chen, Kevin Lu, Aravind Rajeswaran, Kimin Lee, Aditya Grover, Misha Laskin, Pieter Abbeel, Aravind Srinivas, and Igor Mordatch. Decision transformer: Reinforcement learning via sequence modeling. *Advances in neural information processing systems*, 34:15084–15097, 2021.

Justin Fu, Aviral Kumar, Ofir Nachum, George Tucker, and Sergey Levine. D4rl: Datasets for deep data-driven reinforcement learning. *arXiv preprint arXiv:2004.07219*, 2020.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.