Transfer Learning applied to Reinforcement Learning problem with continuous state space using Human-like recall/association

1 Abstract

Transfer knowledge in humans is often mimicked by machine learning algorithms. The objective in such cases is to improve learning performance in a given task by using information learned in a set of related tasks. In reinforcement learning, transfer techniques have achieved limited success, particularly in tasks with large state sets or continuous states where a function approximation is needed to estimate the value function. In this paper, we present a novel strategy to facilitate knowledge transfer when an agent is learning to solve a sequence of increasing difficulty tasks [1, 2]. We show how a sequence of tasks is an appropriate scenario to replicate the human transfer process. In particular, gradually increasing the difficulty of the tasks in a sequence allows us to quantify a notion of distance between tasks using a similarity function. By using This similarity function the agent is able to determine when to transfer knowledge from a previously learned task autonomously.

We first illustrate our strategy in a sequence of attack avoidance games where we employ a similarity function designed by hand. Our experiments show that learning time is improved by learning a sequence of increasing difficulty that leads to a target task compared to learning the target task directly. Naturally, for most task a similarity function can not be hand crafted. We propose to train a parameterized model that helps the agent identify states or sequences of states in the current tasks that are similar or dissimilar to those encountered in previously learned tasks and decide whether to transfer information accordingly. We are currently carrying out experiments using this idea in the task of an agent that learns to play the game Othello. We implemented several models, namely an One Class-SVM, an Auto Encoder, and an LSTM autoencoder.

2 Transfer learning in attack avoidance game

2.1 Game rules

The sequence of tasks for these experiments is shown in figure 1. The objective of the agent (gold square) is to arrive to the green area without crashing with any attacker or red zone. As the difficulty of the task increases by the presence of more adversaries, the state space increases, and thus the size of the neural network that approximates the value functions changes.

2.2 Similarity and Policy advice

The idea behind the similarity function is intuitive from the human learning perspective: When the agent is learning a new task T_k and faces a new situation the obvious first reaction is to relate the actual state to a previous experience from T_{k-1} to decide which could be the best action to take. If when learning task T_k the agent sees $k-1$ adversaries in its close neighborhood, it considers this situation familiar, and with a given probability uses the policy from task T_{k-1} to select an action.

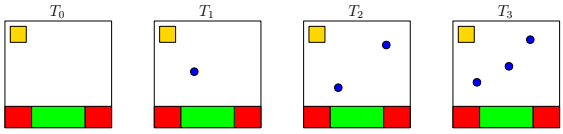

Figure 1: Sequences of attack avoidance tasks

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

Fifteen experiments where carried out, each one with a different combination of tasks and a given task as the target task. In experiments 8 through 15 the target task is T_3 Each transfer experiment $\mathrm{T}_{k-1} \rightarrow \mathrm{T}_k$ was run 15 times. The Q value function transferred corresponds to an agent whose performance was the average of the 15 runs. From this agent we obtain also the transfer rate ϕ. In table 1 we show the winning probability of the agent for each experiment type. The empty spaces in the table mean that that particular task was not seen by the agent.

Table 1: Experiments results - Target task T_3

				Winning probability			
Experiment	Transfer rate			T_0	T_1	T_2	T_3
	0-1	1-2	2-3	ϵ =0.1	ϵ =0.1	ϵ =0.1	ϵ =0.1
E_8	-	-	-	-	-	-	0.4008
							0.5336
E_9	-	-	-	0.9996	-	-	0.6130
E_{10}	-	-	0.2	-	0.4671	-	0.5246
E_{11}	0.4	0.5	-	0.9996	0.6574	-	0.5322
E_{12}	-	-	0.3	-	-	0.4593	0.5025
E_{13}	-	0.2	0.1	0.9996	-	0.5095	0.4931
E_{14}	-	0.1	0.9	-	0.4671	0.4542	0.4817
E_{15}	0.4	0.4	0.5	0.9996	0.6574	0.6002	0.6400

As it can be seen, when the agent has experienced all the previous task the probability of winning in task 3 is larger than when the agent experiences none or even only two preview task before the third task.

3 Work in progress: paremeterized similarity function

In our second set of experiments, we consider the task of learning to play Othello. An agent learns by facing a sequence of adversaries with increasing ability. In this scenario is more difficult for the agent to identify when the state it is facing is similar to a state observed in a previous task. Instead of using a hand crafted similarity function, we train a parameterized model that helps the agent identify states or sequences of states in the current tasks that are similar or dissimilar to those encountered in previously learned tasks and decide whether to transfer information accordingly. This approach is illustrated in figure 2

Currently, wea are running experiments using several models to represent our similarity function, namely, an One Class-SVM, an Auto Encoder and an LSTM autoencoder.

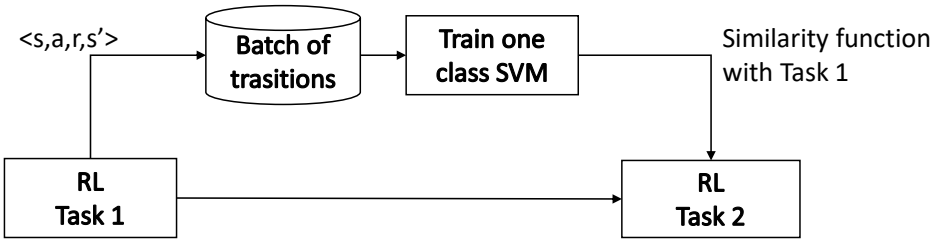

Figure 2: Construction of parameterized similarity function

References

[1] Michael G. Madden and Tom Howley. Transfer of experience between reinforcement learning environments with progressive difficulty. *Artif. Intell. Rev*, 21:375–398, 2004.

[2] Matthew E. Taylor, Gregory Kuhlmann, and Peter Stone. Accelerating search with transferred heuristics. In *ICAPS-07 workshop on AI Planning and Learning*, September 2007.