Learning Transfer Automatic through Data Mining in Reinforcement Learning

Zeinab Arabasadi
Department of Computer Science
University of Bojnord
Bojnord, Iran

Nafiseh Didkar
Department of Computer Science
Amirkabir University of Technology
Tehran, Iran

ABSTRACT
One of the problems in reinforcement learning is that as the environment becomes more complex, the number of parameters used in decision making increase which leads us to a slow decision making process. The main idea here is to come up with a new algorithm which is able to transfer the information, using data mining techniques in extracting the patterns. We introduce a new algorithm for state transitions and actions which happen during the transfer by the agent are saved as a data set for data mining techniques which is presented Learning With Action Transfer (LAT). The main idea is to use the repeated action in each state, as a pattern in similar states as a means to improve learning speed and performance. The results in our algorithm will be compared to the results in Q-learning algorithm..

General Terms
Reinforcement learning

Keywords
Reinforcement learning, Transfer learning, Data mining

1. INTRODUCTION
Reinforcement learning [1, 2] is a branch of machine learning which studies the behavior of the intelligent agent facing stochastic and unknown environment. In reinforcement learning the intelligent agent explores the area with trial and error and gradually improves his behavior with supportive signals from the area. Reinforcement learning methods have achieved a lot of success but they often are not efficient enough in case of complex and large state space. If RL algorithms become able to use their previous experiences in learning new tasks, they would be more efficient. Using Transfer learning[3] is one way to achieve this goal. Transfer learning uses the information of previous learning as a means to improve performance in a new task.

A good idea of research has been done in the field of Transfer Learning, but in most cases the relationship between the source and target task is decided by a human. Here, the attempt is to find this relationship automatically. Transfer learning in RL(reinforcement learning) is an important topic to address at this time for three reasons. First, in recent years RL techniques have achieved notable successes in difficult tasks which other machine learning techniques are either unable or ill-equipped to address. Second, classical machine learning techniques such as rule induction and

Classifications are sufficiently mature that they may now easily be leveraged to assist with TL. Third, promising initial results show that not only are such transfer methods possible, but they can be very effective at speeding up learning.

Selfridge et al. [4] demonstrated that it was faster to learn to balance a pole on a cart by changing the task's transition function, T, over time.

Similarly, the idea of learning from easy missions Asada et al.[5] also relies on a human constructing a set of tasks for the learner. In this work, the task (for example, a maze) is made incrementally harder not by changing the dynamics of the task, but by moving the agent's initial.

Selfridge et al.[4] and Asada et al. [5] provide useful methods for improving learning, which follow fromSkinner's animal training work. While they require a human to be in the loop, and to understand the task well enough to provide the appropriate guidance to the learner, these methods are relatively easy ways to leverage human knowledge.

In Atkeson and Santamaria [6] transfer between tasks in which only the reward function can differ are again considered. Their method successfully transfers a locally weighted regression model of the transition function.

Asadi and Huber [7] have the agent identify states that "locally form a significantly stronger 'attractor' for state space trajectories" as subgoals in the source task (i.e., a doorway, between rooms that is visited relatively often compared to other parts of the state space). The agent then learns options to termed the decision-level model. Ravindran and Barto [8] learn relativized options in a small, human selected source task. When learning in the target task, the agent is provided these options and a set of possible transformations it could apply to them so that they were relevant in the target task. Ferguson and Mahadevan [9] take a unique approach to transfer information about the source task's structure. Proto-value functions (PVFs).

This paper introduces Reinforcement Learning and Action Transfer (LAT). LAT which is able to learn relationship between source task and target task autonomously by using of data mining techniques in extracting the patterns and replace the content with your own material.

2. PROPOSED METHOD
As stated above, increasing the size of state space will cause a decline in learning speed. In this paper an attempt has been made to improve the learning speed.

Suggested algorithm, divides the state space into several contexts. What we mean by a context is a part of state space in which all the variables are constant except one. That is, two areas are said to have the same context, when all their state space variables are identical, but they differ in one variable. So, for instance in a building with the same rooms on each floor (considering the number of rooms, their location and their size) there are two variables, room number and floor number. Each room on a floor is called a context since the floor number is equal but the room number differs. Or in the taxi problem

[10], there are two contexts: before picking up a passenger and after that.

When the contexts are defined, one of the reinforcement learning algorithms should be performed so that the agent can learn about the context. We used Q-Learning algorithm for this purpose. One of the popular reinforcement learning is the Q-Learning [11]. In this algorithm the agent updates the Q-function at each step. In each episode as the agent is exploring the context, some paths are traversed. If during the exploration reach these subgoals via a learned action-value function.

The agent passes a state and then comes to the same state again, there has been a loop. The saved paths which contain no loops are used in learning in similar contexts. In other words, these paths are a series of states in each of which an action has taken place. In each context, a number of stored paths pass from a state, the action that frequently occurs in a state is used as a frequent pattern for this state. In similar contexts we use

This pattern. For example if there are 6 patterns for state i, three of which going left, two going down and one going up, the state pattern would be going left.

In order to find similar contexts we use time series clustering [12]. For learning in a new context, the agent begins with exploring the context. After some episodes in new context it saves the paths which contain no loops and then with the saved paths learned before, using time series [13] and clustering, it finds the context, similar to the new one. Now for each state in new context its corresponding state in the similar context is identified and the frequent pattern of that state is transferred to the new context. During the process of learning we use greedy algorithm.

The advantage shows itself in contexts which vary in size or contexts which are rotated, since the algorithm could be performed using scale and rotation in saved paths.

LAT Algorithm

Repeat

(1). Interact with context & learn context

(2). Delete all loops from the paths that agent explore

(3). Store Paths without loop

(4). Extract pattern from stored paths for each state

Until no new context was found in the source task

Repeat

(1). Explore context

(2). Find similar context in source task

(3). Transfer action from similar context to corresponding state from current context

Until no new context was found in the target task

3. EXPERIMENTAL AND RESULTS

The algorithm will be performed in an environment consisting of two floors and environment consisting tower Hanoi, environment consisting of two floors with three rooms grid world on each floor "Fig1". In each state the agent is allowed to choose one of the four actions, left, right, up or down. As the agent starts exploring the first floor, it saves the paths which contain no loops and finds the frequent patterns of states. Then it starts exploring the rooms of the second floor and by transferring the data in similar situations, in tower Hanoi, we want to action transfer from two disks to three disks, we use a probability of 0.8 to decide the next action in which the state is chosen through data transfer. And with a probability of 0.2 the action which has the most value in Q-value, would be taken as the next action. And as illustrate in "Fig 2", learning with transfer shown an early improvement in performance in comparison with Q-Learning.

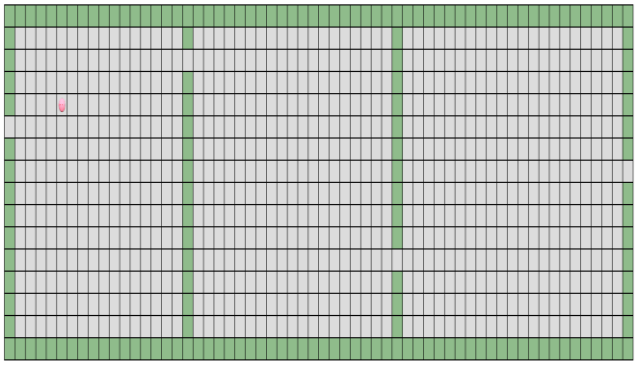

Fig 1. Experiment environment (consist of two floors, this picture is displayed one floor with three-room grid world)

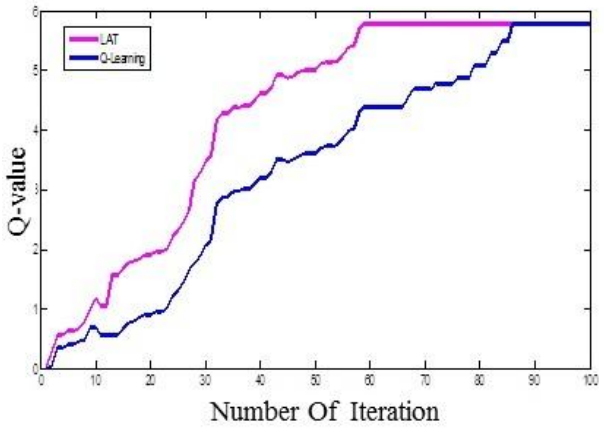

Fig 2. Comparison of the Q-learning and Q-learning with transfer learning for room a

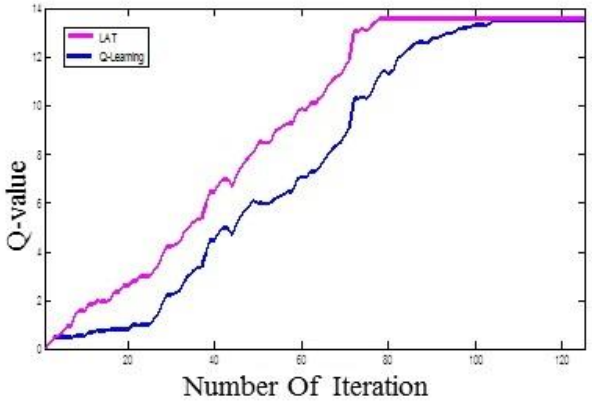

Fig 3.comparison of the Q-learning and Q-learning with transfer learning for room b

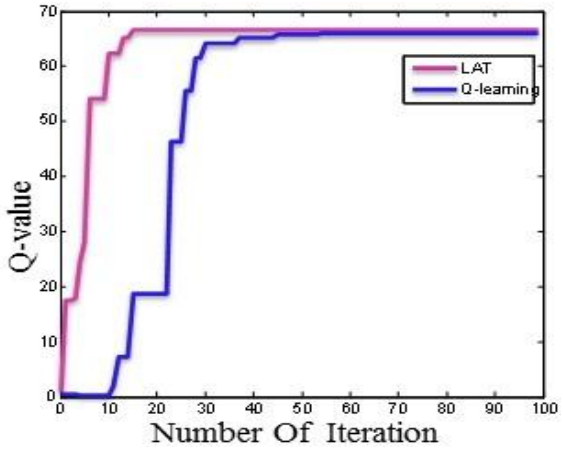

Fig 4. Comparison of the Q-learning and Q-learning with transfer learning for room c

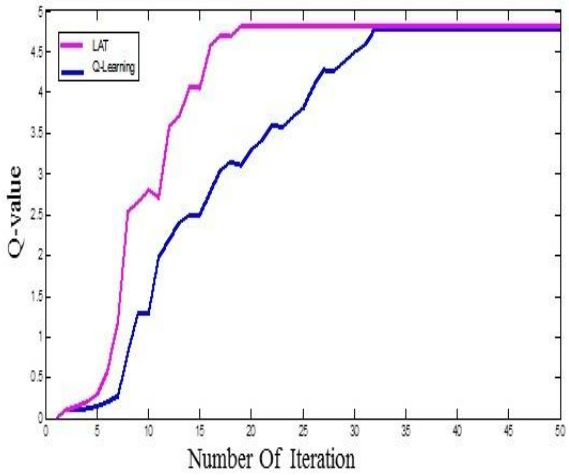

Fig 5. Comparison of the Q-learning and Q-learning with transfer learning for tower Hanoi

Three rooms are named a, b and c from the right two left, and each room reflects a context. "fig2" displays the comparison between Q-learning and Q-learning with transfer in room a,"fig3" refers to room b and "fig4" refers to room c, fig 5 refers to tower Hanoi.

4. CONCLUSION

In this paper represented algorithm which is able to transfer the information, using data mining techniques in extracting the patterns. The state transitions and actions which happen during the transfer by the agent are saved as a data set for data mining techniques. The main idea is to use the repeated action in each state, as a pattern in similar states as a means to improve learning speed and performance. In cases of larger areas the algorithm would be more efficient and a considerable amount of time would be saved. We compared the standard Q-Learning algorithm with our algorithm. As future work, Graph matching algorithms can be used instead of data mining algorithms.

5. REFERENCES

[1] R. Sutton and A. Barto, Introduction to Reinforcement Learning, MIT Press, Cambridge (1998).

[2] L.P. Kaelbling, M.L. Littman and A.W. Moore, Reinforcement learning: A survey, Journal of Artificial Intelligence Research 4 (1996), pp. 237–285.

[3] Matthew E. Taylor, Gregory Kuhlmann, Peter Stone: Autonomous transfer for reinforcement learning. AAMAS (1) 2008: 283-290

[4] Oliver G. Selfridge, Richard S. Sutton, and Andrew G. Barto. Training and tracking in robotics.

[5] Minoru Asada, Shoichi Noda, Sukoya Tawaratsumida, and Koh Hosoda. Vision-based behavior acquisition for a shooting robot by using a reinforcement learning. In Proceedings of IAPR/IEEE Workshop on Visual Behaviors-1994, pages 112–118, 1994.

[6] Christopher G. Atkeson and Juan C. Santamaria. A comparison of direct and model-based reinforcement learning. In Proceedings of the 1997 International Conference on Robotics and Automation.

[7] Mehran Asadi and Manfred Huber. Effective control knowledge transfer through learning skill and representation hierarchies. In Proceedings of the 20th International Joint Conference on Artificial Intelligence, 2007 , pages 2054–2059.

[8] Balaraman Ravindran and Andrew G. Barto. Model minimization in hierarchical reinforcement learning. In Proceedings of the Fifth Symposium on Abstraction, Reformulation and Approximation, 2002

[9] Kimberly Ferguson and Sridhar Mahadevan. Proto-transfer learning in Markov decision processes using spectral methods. In Proceedings of the ICML-06 Workshop on Structural Knowledge Transfer for Machine Learning, June 2006

[10] T. Dietterich, Hierarchical reinforcement learning with the MAXQ value function decomposition, Journal of Artificial Intelligence Research 3(2000) 227_303.

[11] P. Dayan, C. Watkins, Q-learning, Machine Learning 8 (1992) 279_292.

[12] Morgan Kaufmann Publishers is an imprint of Elsevier500 Sansome Street, Suite 400, San Francisco, CA94111Page468-489

[13] Cheng-Ping Lai, Pau-Choo Chung, Vincent S. Tseng: two-level clustering method for time series data analysis. Expert Syst. Appl (2010), 37