Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)

Improving Deep Reinforcement Learning with Knowledge Transfer*

Ruben Glatt, Anna Helena Reali Costa
Escola Politécnica da Universidade de São Paulo, Brazil
{ruben.glatt,anna.reali}@usp.br

Abstract

Recent successes in applying Deep Learning techniques on Reinforcement Learning algorithms have led to a wave of breakthrough developments in agent theory and established the field of Deep Reinforcement Learning (DRL). While DRL has shown great results for single task learning, the multi-task case is still underrepresented in the available literature. This D.Sc. research proposal aims at extending DRL to the multi-task case by leveraging the power of Transfer Learning algorithms to improve the training time and results for multi-task learning. Our focus lies on defining a novel framework for scalable DRL agents that detects similarities between tasks and balances various TL techniques, like parameter initialization, policy or skill transfer.

Introduction and Context

Deep Learning (DL) architectures became one of the most powerful tools in Artificial Intelligence (AI) research in recent years, beating long standing records in many domains of Machine Learning like computer vision or language understanding (LeCun, Bengio, and Hinton 2015). They allow to learn abstract representations of high dimensional input data and can be used to improve many existing algorithms.

One of the techniques that can benefit from DL is the well-researched area of Reinforcement Learning (RL) (Sutton and Barto 1998). In RL, an agent explores the space of possible strategies to solve a task in a given environment, receives a feedback (reward or cost) on the outcome of the actions she takes and deduces a behavior policy from her observations over time. The goal of the agent is to determine a policy π that maps each state to an action, which maximizes the accumulated reward over a given horizon.

This form of sequential decision-making can be considered a Markov Decision Process (MDP). The core of the MDP is the Markov Property, which is given, if future states of the process depend only upon the present state and the action taken in this state. RL algorithms can be used to find a solution for this kind of problem and already achieve excellent results in a variety of domains like the board-game domain, autonomous helicopter flight or robot soccer.

Although it is not a new idea to use Neural Networks (NN) on RL problems (Riedmiller 2005), advances in algorithms for DL have brought up a new wave of successful applications as for example Atari game playing (Mnih et al. 2015) and established the field of Deep Reinforcement Learning (DRL). In this context a trained Deep Neural Network (DNN) can be seen as a kind of end-to-end RL approach, where the agent learns a state abstraction and a policy approximation in a single network directly from her input data. The learned policy follows the optimal action-value function $Q^*(s, a)$, which is approximated by the DNN and therefore termed Deep Q-Network (DQN).

However, RL algorithms have three major weaknesses: First, they need a long time and many examples to converge to an acceptable result during training; Second, a model that has been trained on one task can only be used on that same task; and third, the possible state and action spaces are not always fully observable.

One way to deal with those shortcomings, when learning multiple tasks, is to reuse already gathered knowledge, as is being researched in the field of Transfer Learning (TL) (Taylor and Stone 2009). The challenges in TL are that the agent has to decide, which knowledge she should transfer from which source and in which way to speed up and improve the training of a new task.

Motivation and Proposal

We have seen impressive results in many domains utilizing DRL over the last years but the main focus of the published work seems to be heavily on single task learning while the the multi-task case is still underrepresented, although it is finally improving. It is our firm believe, that the combination of RL and DL together with TL will play a great role in AI research towards building general AI agents. Since each of the introduced techniques provides successful methods to deal with different aspects of learning, but also has shortcomings in others, there is a great potential in combining the various advantages to improve current methods.

Many challenges remain the same for transfer in RL and DRL, as for example *negative transfer*, where the transferred

*We are grateful for the support from CAPES and CNPq (grant 311608/2014-0). The HPC resources for the computation are provided by the Superintendency of Information at the University of São Paulo. We also thank Google for the support due to the Google Research Awards 2016 for Latin-America and the Nvidia corporation for their hardware donation.

knowledge hurts the learning, or the determination of similarity between tasks, which helps to select the right source task or policy for the transfer. But the knowledge transfer for DRL comes with an additional level of difficulty compared to the classic domain because the state abstraction and the policy approximation are embedded within a single network and cannot easily be separated.

Consequently, the goal of our research is to overcome the restriction to single task learning for DRL by exploiting prior knowledge, acquired in previously or in parallel learned tasks, and help to fill the gap in the available literature for multi-task learning with DRL. More precisely, we propose to investigate autonomous RL agents to improve and speed up the learning for the multi-task case to build agents that can handle more tasks in a wider field of domains and thereby enhance the relevance for and the robustness in real-world applications.

During the course of this D.Sc. project, we intend to contribute a novel framework for scalable DRL agents that autonomously detects similarities between tasks and balances various TL techniques, like parameter initialization, policy or skill transfer, to improve the training time and results for multi-task learning, where skills are a combination of actions for common sub-tasks.

A focus will be on detecting underlying hierarchical structures in tasks and exploiting them while learning a new task. For this purpose, we will need an architecture that not only learns a state representation in regard to a single task, but one that can be generalized to various tasks without loosing specificity and at the same time learns to coordinate between different sources of knowledge to select only the relevant knowledge for the transfer. This could be achieved by introducing a kind of supervisory network, that learns to infer the best source or a set of weighted sources for knowledge transfer, which would also help to prevent *negative transfer* during learning.

Towards that goal, our next steps will be to adapt an approach from the classic domain, *Policy Reuse* (which organizes the transfer of whole policies) (Fernández and Veloso 2006), and compare it to two recently published approaches, *ADAAPT - A Deep Architecture for Adaptive Policy Transfer* (which uses an attention network to blend policies) (Rajendran et al. 2015) and *AMN - Actor-Mimic-Network* (which uses parameter transfer for initialization only) (Parisotto, Ba, and Salakhutdinov 2015) to get more experience with relevant approaches that use transfer in DRL.

Preliminary Results

In the course of determining the suitability of knowledge transfer for DRL, we conducted a number of experiments to encourage a broader discussion of the topic (Glatt, Silva, and Costa 2016) and developed a software framework to facilitate research in the area[1].

These initial experiments where conducted to evaluate possible effects on training time and overall cumulative reward when applying simple parameter transfer on DQNs. First, we trained networks on individual tasks with different configurations in regard to optimization algorithm of the gradient decent method and output nodes of the network. Then, we kept the structure of the network identical for all tasks and only transferred the network weights from an already trained network to initialize the network for a new task. The Atari game playing domain served as the experiment environment, but we evaluated only a very limited number of games mainly because of computational restraints we had at the time.

However, our results imply that the initialization of the DQN plays a far more important role than the choice of optimization algorithm or the number of output nodes. They further confirm the findings of TL in the classic RL domain, where transfer from a similar task is beneficial to learning, and transfer from less similar tasks has either a very small effect or even leads to *negative transfer*, if the tasks are very different.

[1] https://github.com/cowhi/deepatari

Conclusions

In conclusion, TL has shown great potential to accelerate learning in DRL tasks, but there are still many aspects to be understood before we can formulate a comprehensive framework for knowledge transfer for DRL. After conducting those initial experiments we will now focus on implementing more transfer learning approaches to get a better understanding of the impact of different aspects and factors for knowledge transfer for DRL.

References

Fernández, F., and Veloso, M. 2006. Probabilistic policy reuse in a reinforcement learning agent. In *AAMAS*, 720–727.

Glatt, R.; Silva, F. L. d.; and Costa, A. H. R. 2016. Towards knowledge transfer in deep reinforcement learning. In *Brazilian Conference on Intelligent Systems (BRACIS)*. IEEE.

LeCun, Y.; Bengio, Y.; and Hinton, G. 2015. Deep learning. *Nature* 521(7553):436–444.

Mnih, V.; Silver, D.; Rusu, A. A.; Riedmiller, M.; et al. 2015. Human-level control through deep reinforcement learning. *Nature* 518(7540):529–533.

Parisotto, E.; Ba, L. J.; and Salakhutdinov, R. 2015. Actor-mimic: Deep multitask and transfer reinforcement learning. *CoRR* abs/1511.06342.

Rajendran, J.; Prasanna, P.; Ravindran, B.; and Khapra, M. M. 2015. Adaapt: A deep architecture for adaptive policy transfer from multiple sources. *arXiv preprint arXiv:1510.02879*.

Riedmiller, M. 2005. Neural fitted q iteration–first experiences with a data efficient neural reinforcement learning method. In *ECML*. Springer. 317–328.

Sutton, R. S., and Barto, A. G. 1998. *Introduction to Reinforcement Learning*. Cambridge, MA, USA: MIT Press.

Taylor, M. E., and Stone, P. 2009. Transfer learning for reinforcement learning domains: A survey. *JMLR* 10:1633–1685.