SENIOR THESIS 2021-22

Abstract and Problem Statement

Samar Rahmouni
srahmoun@andrew.cmu.edu

Advisor: Prof. Giselle Reis
giselle@cmu.edu

Implementing autonomous agent controllers that can robustly and efficiently adapt to different dynamic and complex scenarios is still an open challenge in robotics and AI. For instance, it is a challenge that self-driving vehicles and humanoid robots face because of a continuously changing real-world. Thus, autonomous agents have to face multiple sources of stochasticity once deployed. Precisely, in the case of Reinforcement Learning, an autonomous car trained by trial-and-error is bound to learn how to drive. In practice, this cannot be an option, considering that the AI needs to crash to learn that crashing is not desirable. In other words, actions that lead to crashing can only be learnt once the crash happens. This makes RL nearly impossible to deploy in the real world. We investigate both the security and the interpretability aspect of reinforcement learning in a cooperative adaptive cruise control inspired from [1], in the aim of finding how formal security frameworks can guide the representation, robustness and extrapolation of knowledge in Reinforcement Learning agents.

1 Problem Statement

Implementing a robust adaptive controller that is effective in terms of precision, time, and quality of decision when facing dynamic and uncertain scenarios, has always been a central challenge in AI and robotics. As autonomous cars are deployed, IoT is popularized, and human-robot interactions become more complex, we are more and more confronted with the need for robotic agents that can effectively and continually adapt to their surroundings, not only in simulation, but also in practice, when deployed as a cyber-physical system. Since we are unable to provide a repertoire of all possible scenarios and actions, our agents need to be able to autonomously predict and adapt to new changes. RL is an approach that supports developing these capabilities, it is also the solution that AlphaGo, Deepmind AlphaStar, and OpenAI Five have adopted [2] and found success in. However, as RL is a trial-and-error process, a car trained using RL is bound to crash to learn not to crash again. Safe Reinforcement Learning is then crucial to investigate in order to be able to deploy it in larger scales, but also out of simulation.

One of the many ways safety has been approached is by formalization and symbolic reasoning. In the case of artificial intelligence, recent work proposes Neurosymbolic integration. Neurosymbolic integration has been an ongoing work in the last years towards a combination of Deep Learning and Symbolic Reasoning. The work has been a response to critics of DL, precisely, the lack of formal semantics and intuitive explanation and the lack of expert knowledge towards guiding machine learning models. Key questions the field targets are identifying the necessary and sufficient building blocks of AI [3], namely, how can we provide the semantics of knowledge, and work towards meta-learning? Meta-learning in reinforcement learning is the problem of learning-to-learn, which is about efficiently adapting a learned policy to conditions and tasks that were not encountered in the past. In RL, metalearning involves adapting the learning parameters, balancing exploration and exploitation to direct the agent interaction [4, 5]. Meta-Learning is a central problem in AI, since an agent that can solve more and more problems it has not seen before, approaches the ideal of a general-purpose AI.

Current Neurosymbolic AI trends are concerned with knowledge representation and reasoning, namely, they investigate computational-logic systems and representation to precede learning in order to provide some form of incremental update, e.g. a meta-network to group two sub-neural networks. [6] This leads to neurosymbolic AI finding various applications including vision-based tasks such as semantic labeling [7, 8], vision analogy-making [9], or learning communication protocols [10]. However, such representation has not been yet investigated in tabular reinforcement learning algorithms. Precisely, the problem of knowledge extraction and interpretability of RL architecture is yet to be tackled. Approaches that can be considered include state representation in the form of classical planning and ways to build hybrid architecture to include a symbolic module.

In the following thesis, we investigate different ways formal methods can be used in an adaptive cruise control scenario. Precisely, we look into a combination of Reinforcement Learning and Symbolic Reasoning that allow us to ensure given safety properties, i.e. a Safe RL architecture.

The problem of safe reinforcement learning has been approached in previous work [11] in two ways.
First was changing the optimization criteria [12], precisely by incorporating risk into the performance of the policy. Namely, either considering the worst case scenario and constraining on it, reducing the variance to be more sensitive or applying constraints i.e. only update the policy if the action is in a safe set. This does not however guarantee safety, but tends to minimize the probability of risk, hence not allowing RL systems to be deployed in the physical world. Second was to consider the exploration process, either by incorporating external knowledge i.e. learning from demonstration [13] or adopting a risk directed exploration [14]. These approaches can be considered rigid, as requiring more data and expert knowledge that needs to be proven safe and sound, or in the latter, decreasing the efficiency and requiring more time for the learning process. In particular, in these previous two approaches, work that makes use of formal security frameworks is yet to be investigated.

Some of the more recent work that does investigate a hybrid architecture (i.e. RL and symbolic reasoning) makes use of set-theoretic techniques and constraint satisfaction problems to optimize from the constraints [15], or proposes a reactive system called a shield [16] to either constraint the actions given by the environment or adapt them once the RL module chooses one. In both cases, the added safety controller is described by its specifications, rather than being an independent existing controller.

We propose two ways to combine the controller specified in [1] and Reinforcement Learning in order to ensure safety properties. (1) Similar to the previous work, given the range of safe speeds the controller returns, we investigate the use of RL in finding the best value that minimizes the gap between the vehicles. (2) Given the safety sets given by the controller, we are able to adapt the rewards, i.e. an action that leads to crashing can be punished before the crashing happens, hence investigating whether the safety sets learnt by the RL agent converge to those given by the controller.
Precisely, the first part highlights the use of RL in optimizing a safe controller, while formally ensuring the safety properties. The second part aims to answer whether a formal reward system inspired from a safe controller can either minimizes risks or ensure the convergence of safety. However, towards answering both these questions, the symbolic reasoning of Reinforcement Learning i.e. in the way neurosymbolic AI is investigated, needs to be formalized.

References

[1] Yuri Gil Dantas, Vivek Nigam, and Carolyn Talcott. A formal security assessment framework for cooperative adaptive cruise control. In *2020 IEEE Vehicular Networking Conference (VNC)*, pages 1–8, 2020.

[2] Yuxi Li. Reinforcement Learning Applications. Technical Report arXiv:1908.06973, August 2019.

[3] Artur d'Avila Garcez and Luis C. Lamb. Neurosymbolic ai: The 3rd wave, 2020.

[4] Abhishek Gupta, Russell Mendonca, YuXuan Liu, Pieter Abbeel, and Sergey Levine. Meta-Reinforcement Learning of Structured Exploration Strategies. In *Conference and Workshop on Neural Information Processing Systems (NeurIPS)*, page 10, 2018.

[5] Nicolas Schweighofer and Kenji Doya. Meta-learning in Reinforcement Learning. *Neural Networks*, 16(1):5–9, January 2003.

[6] Tarek R. Besold, A. Garcez, Sebastian Bader, H. Bowman, Pedro M. Domingos, P. Hitzler, Kai-Uwe Kühnberger, L. Lamb, Daniel Lowd, P. Lima, L. Penning, Gadi Pinkas, Hoifung Poon, and Gerson Zaverucha. Neural-symbolic learning and reasoning: A survey and interpretation. *ArXiv*, abs/1711.03902, 2017.

[7] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. pages 3156–3164, 06 2015.

[8] Andrej Karpathy and Fei Li. Deep visual-semantic alignments for generating image descriptions. pages 3128–3137, 06 2015.

[9] Scott E. Reed, Yi Zhang, Y. Zhang, and Honglak Lee. Deep visual analogy-making. In *NIPS*, 2015.

[10] Jakob N. Foerster, Yannis M. Assael, N. D. Freitas, and S. Whiteson. Learning to communicate to solve riddles with deep distributed recurrent q-networks. *ArXiv*, abs/1602.02672, 2016.

[11] Javier García and F. Fernández. A comprehensive survey on safe reinforcement learning. *J. Mach. Learn. Res.*, 16:1437–1480, 2015.

[12] R Rockafellar and Stan Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 01 2000.

[13] Nils T. Siebel and G. Sommer. Evolutionary reinforcement learning of artificial neural networks. *Int. J. Hybrid Intell. Syst.*, 4:171–183, 2007.

[14] Edith Law. Risk-directed exploration in reinforcement learning. 01 2005.

[15] Yutong Li, N. Li, H. E. Tseng, A. Girard, Dimitar Filev, and I. Kolmanovsky. Safe reinforcement learning using robust action governor. In *L4DC*, 2021.

[16] Mohammed Alshiekh, Roderick Bloem, Rüdiger Ehlers, Bettina Könighofer, Scott Niekum, and Ufuk Topcu. Safe reinforcement learning via shielding. 08 2017.