SENIOR THESIS 2021-22

Mid-Semester Report

Samar Rahmouni
srahmoun@andrew.cmu.edu

Advisor: Prof. Giselle Reis
giselle@cmu.edu

1 Abstract

Reinforcement learning (RL), though a powerful and simple trial-and-error procedure that found a lot of success in games like Go, cannot be deployed in the real world because of the lack of security guarantees. For instance, though an autonomous car trained with reinforcement learning is bound to learn how to drive, the AI needs to crash to learn that crashing is not desirable.

In the context of the proposed thesis, 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. We do so by experimenting and comparing both the optimization and security of three different RL implementations.

The first is a basic tabular Q-learning where we expect no security guarantees. The second is a hybrid architecture that incorporates the safe controller in [1] in a RL architecture. The safe controller (SC) computes a range of safe velocities given the current state of the environment in the car platooning scenario. Precisely, for multiple vehicles following each other, platooning aims to reduce the distance between them, hence taking less space on the road and allowing more vehicles to occupy highways, for instance. In the RL architecture, the SC then constricts the possible actions of the RL continuously at every given time step and the role of the RL is purely to find the optimal velocity in the safe range to minimize the distance between the cars. The third is what we will define as a logic-based inference RL. We investigate learning inferences rules in a deterministic environment where the result of an action given a state is not dependent on any probabilistic event. We approach the problem using inductive reasoning where the goal is to incorporate the learned rule knowledge into the decision making of a tabular Q-learning RL agent. In the given scenario, this will make use of the SC to develop a mapping from the state representation to a reward scheme, i.e. punish before a crash is bound to happen and adapt the Q-value and the epsilon-greedy approach.

2 Introduction

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], respectively for Go, StarCraft II and Dota 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. The survey in [3] lays the foundations of verifications goals for Artificial Neural Networks (ANNs) to be ensured for safety-critical tasks. The need for this behavioral constraints stems from the inability to analyze and describe the behavior of NNs. The same argument translates to RL. In general, reinforcement learning does not ensure any of the verification conditions in [3]. The observable behavior of a RL agent is unpredictable (e.g. AlphaGo gamestyle) and will more often than not take hazardous actions since it is a trial-and-error process. Hence, it is important to investigate the use of verified safety controller and the interpretability of RL in order to achieve these verifications goals. Add graph of proposed infrastructure.

3 Research Contribution

We (1) provide a framework to incorporate a verified safety controller into a reinforcement learning architecture, (2) investigate a compositional state representation for reinforcement learning in the lines of classical planning, and (3) incorporate rule knowledge into the decision-making of tabular reinforcement learning in a deterministic environment. The above is done as to test and experiment on the tradeoff of optimization and safety in basic RL, RL with a safe controller, and finally a logic-based inference RL.

4 Vehicle Platooning : Case Study

todo

5 Safe Controller as a Shield

todo

5.1 Architecture

Add graphics and description

5.2 Preliminary Results

todo

5.3 Discussion

todo

6 Related Work

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 (DL) and symbolic reasoning. The work has been a response to criticism 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 [4], 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, meta-learning involves adapting the learning parameters, balancing exploration and exploitation to direct the agent interaction [5, 6]. 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 [7]. As a result, neurosymbolic AI has various applications including vision-based tasks such as semantic labeling [8, 9], vision analogy-making [10], or learning communication protocols [11]. However, such representation has not yet been investigated in tabular reinforcement learning algorithms. Precisely, the problem of knowledge extraction and interpretability of RL architecture is yet to be tackled.

Artificial intelligence has moved away in the past years from symbolic to data-driven. Data-driven AI have proven to be both efficient and their algorithms easy to reuse. However, we are facing the limitations of such AI today whether in its incapability of developing abstract relations between components, or in the consistent problem of meta-learning in extrapolating learned knowledge in one environment to another, or finally, the safety properties that are harder to reason about and furthermore ensure. In the light of such limitations, hybrid architectures are becoming increasingly popular [4].

However, most approaches so far only considered neural networks. In [12], they present multiple formal methods that have been investigated and applied to Machine Learning, these include SMT-based and MILP-based formal methods which use constraint satisfaction and mixed-integer linear problems. The main idea is to reduce safety properties into a set of constraints. If a solution exists then the property does not hold, else it is safe. These methods, though sound and complete, do not scale for bigger networks.

Approaches that were concerned with safe reinforcement learning rather than the above, have approached the problem in previous work [13] in two ways. First was changing the optimization criteria [14], 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 [15] or adopting a risk directed exploration [16]. 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 [17], or proposes a reactive system called a shield [18] 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.

Our approach is novel as it (1) focuses on tabular reinforcement learning rather than neural networks, (2) provides a framework independent of the specifications of either the reinforcement learning controller or the safe controller and (3) investigates a translation from symbolic compositional state representation in RL to logic-based inference.

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] Zeshan Kurd and Tim Kelly. Establishing safety criteria for artificial neural networks. volume 2773, pages 163–169, 09 2003.

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

[5] 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.

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

[7] 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.

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

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

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

[11] 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.

[12] Caterina Urban and Antoine Min'e. A review of formal methods applied to machine learning. *ArXiv*, abs/2104.02466, 2021.

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

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

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

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

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

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