Logically-Constrained Reinforcement Learning

Mohammadhosein Hasanbeig, Alessandro Abate, and Daniel Kroening

University of Oxford

In this work we employ RL to synthesize a control policy for a Markov Decision Process (MDP) such that the generated traces satisfy a Linear Temporal Logic (LTL) property. In this problem we assume that the MDP transition probabilities are unknown. LTL, a modal logic, is able to express rich logical time-dependent properties such as safety and liveness. An LTL property can be represented as an automaton. However, LTL-to-automaton translation sometimes results in non-deterministic automata with which model checking cannot generally be performed. The standard solution is to use Safra's construction to determinize the automaton, which increases the size of automata dramatically [1, 2].

Another solution is to direcly convert a given LTL property into a Deterministic Rabin Automaton (DRA), since the resulting automaton will by definition have deterministic transitions. Nevertheless, it is proven that any such conversion, in the worst case, results in automata that are doubly exponential in size of formulae [3]. Although a fully non-deterministic automaton cannot be used for probabilistic model checking, it is known that automata do not need to be fully deterministic if restricted forms of non-determinism are allowed.

In this work we propose to convert the LTL property into a Limit Deterministic Büchi Automaton (LDBA) [4]. Intuitively, a Büchi automaton is called limit deterministic if every state reachable from an accepting state has deterministic transitions. It is shown that this construction results in an exponential-sized automaton for LTL\GU (a fragment of LTL), and results in nearly the same size as a DRA for the rest of LTL formulae. Furthermore, a Büchi automaton is simpler than a Rabin automaton in terms of its acceptance conditions, which makes the algorithm implementation much simpler, e.g. [5].

Once the LDBA is generated from the LTL property, we construct a synchronous product between the MDP and the resulting LDBA and then assign a reward function based on the acceptance condition of the automaton to the product MDP. By following this reward function, RL is able to generate a policy that satisfies the given LTL property. The proposed algorithm is completely model-free, which means that we are able to synthesize policies without knowing or approximating the transition probabilities. Note that classical Dynamic Programming (DP) is of limited utility in this problem, both because of its assumption of a perfect model and also because of its great computational expense [6].

In the standard DP value iteration method, the value estimation is simultaneously updated for all states. However, an alternative method is to update the value for one state at a time. This method is known as asynchronous value iteration. We show that the RL procedure sets up an online asynchronous value iteration method to calculate the maximum probability of satisfying the given property, at any given state of the MDP. Therefore, In addition to the control

synthesis problem, we employ a value iteration method to calculate the probability of satisfaction of the LTL property. Use of RL for policy generation allows the value iteration algorithm to focus on parts of state space that are relevant to the property. This results in a faster calculation of probability values when compared to DP, where these values are updated for the whole state space.

The performance of the algorithm is evaluated via a set of numerical examples in [7]. We observe an improvement of one order of magnitude in the number of iterations required for the synthesis compared to existing approaches.

References

1. Safra, S.: On the complexity of omega-automata. In: Foundations of Computer Science, 1988., 29th Annual Symposium on, IEEE (1988) 319–327
2. Piterman, N.: From nondeterministic Büchi and Streett automata to deterministic parity automata. In: Logic in Computer Science, 2006 21st Annual IEEE Symposium on, IEEE (2006) 255–264
3. Alur, R., La Torre, S.: Deterministic generators and games for LTL fragments. ACM Transactions on Computational Logic (TOCL) **5**(1) (2004) 1–25
4. Sickert, S., Esparza, J., Jaax, S., Křetínskỳ, J.: Limit-deterministic Büchi automata for linear temporal logic. In: International Conference on Computer Aided Verification, Springer (2016) 312–332
5. Tkachev, I., Mereacre, A., Katoen, J.P., Abate, A.: Quantitative model-checking of controlled discrete-time Markov processes. Information and Computation **253** (2017) 1–35
6. Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction. Volume 1. MIT press Cambridge (1998)
7. Hasanbeig, M., Abate, A., Kroening, D.: Logically-constrained reinforcement learning. arXiv preprint arXiv:1801.08099 (2018)