Plan of Study for 2019-2020

Silviu Pitis, April 2019

My ultimate research interest lies in the *normative design of general purpose artificial agency.* That is, how *should* we design artificial agents that learn to solve general tasks and contribute positively to human society? This a broad question, and I am focusing my efforts on two specific areas that I view as essential pieces of the larger puzzle. The first asks: how should we construct agents that can reason about the world, do so at multiple levels of abstraction, and communicate their reasoning via multiple modalities (especially natural language)? The second asks: what objective functions should agents optimize, from a normative perspective? Below I describe the specific research topics within each that I plan to continue working in the coming academic year (2019-2020).

(Area #1) Integrating multiple goal spaces in multi-goal reinforcement learning. A human can accomplish arbitrarily defined goals, whether expressed by an image or via language, and can translate image goals into language goals and vice versa. In our 2019 extended abstract [7] we propose to enable multi-goal reinforcement learning (RL) agents [12, 1, 10] to do the same by treating goal spaces as topologies (families of sets) over the underlying state space and learning "Protoge" maps between them. This allows us to transfer knowledge between goal spaces and learn a universal goal representation. Significant additional work is needed to develop these ideas into a general-purpose framework for reinforcement learning agents.

(Area #1) Using destination-based general value functions for planning. General value functions (GVFs) [13] encode world knowledge in the form of a value function. Destination-based GVFs are the default GVF in multi-goal reinforcement learning, characterized by a sparse reward function that gives positive signal only at the goal. I am interested in using learned GVFs to solve navigation tasks by top-down hierarchical planning [2, 4]. This is also a promising application of our recent work on modeling metrics with neural networks [8].

(Area #1) Causal reasoning with RL. As humans naturally understand notions of causality, it is likely that reinforcement learning agents can as well—how can we enable this? I believe destination-based GVFs, together with the ideas in my 2018 paper [5] will be applicable here.

(Area #2) Individual intelligence. In [6] I showed that the standard MDP model may be insufficient to model all "rational" preferences. As this work was strictly theoretical, an interesting follow-up question I would like to explore is whether we can demonstrate an empirical benefit to my proposal to use a variable discount factor, e.g., in inverse reinforcement learning [3].

(Area #2) Pro-social and group intelligence. How should an agent integrate feedback signals from multiple principals in a normatively justified way? [11]. How do we design a justified social welfare objective, especially when signals are a mix of ordinal and cardinal values? My recent project [9] formalizes this problem mathematically, and provides a first step toward an answer.

(Area #2) Procedural safety. Many social and safety issues that might arise during an agent's lifetime are unforeseeable or otherwise impossible to hard or soft code (via inductive bias) into our models. We can, however, borrow an idea from the legal system: rather than construct substantive rules in difficult cases, provide the necessary procedure (e.g., due process) so that ultimate outcomes are deemed fair, even if they are considered "wrong" by some (or many!). Can we design a procedural framework to help us manage the undoubtedly many AI policy issues that will arise in the future?

References

[1] M. Andrychowicz, F. Wolski, A. Ray, J. Schneider, R. Fong, P. Welinder, B. McGrew, J. Tobin, O. P. Abbeel, and W. Zaremba. Hindsight experience replay. In *Advances in Neural Information Processing Systems*, pages 5048–5058, 2017.

[2] L. P. Kaelbling. Hierarchical learning in stochastic domains: Preliminary results. In *Proceedings of the tenth international conference on machine learning*, volume 951, pages 167–173, 1993.

[3] A. Y. Ng, S. J. Russell, et al. Algorithms for inverse reinforcement learning. In *Icml*, volume 1, page 2, 2000.

[4] S. Pitis. Reasoning for reinforcement learning. In *Hierarchical Reinforcement Learning Workshop at the 31st Conference on Neural Information Processing Systems (HRL@NIPS)*, 2017.

[5] S. Pitis. Source Traces for temporal difference learning. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. AAAI Press, 2018.

[6] S. Pitis. Rethinking the discount factor in reinforcement learning: A decision theoretic approach. In *Proceedings of the 33nd AAAI Conference on Artificial Intelligence*. AAAI Press, 2019.

[7] S. Pitis, H. Chan, and J. Ba. ProtoGE: Prototype goal encodings for multi-goal reinforcement learning. In *The 4th Multidisciplinary Conference on Reinforcement Learning and Decision Making (RLDM)*, 2019.

[8] S. Pitis, H. Chan, K. Jamali, and J. Ba. An inductive bias for distances: Neural nets that respect the triangle inequality. In *Proceedings of the Eighth International Conference on Learning Representations (ICLR)*, 2020.

[9] S. Pitis and M. R. Zhang. Objective social choice with non-i.i.d. votes. In *Proceedings of International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 2020.

[10] M. Plappert, M. Andrychowicz, A. Ray, B. McGrew, B. Baker, G. Powell, J. Schneider, J. Tobin, M. Chociej, P. Welinder, et al. Multi-goal reinforcement learning: Challenging robotics environments and request for research. *arXiv preprint arXiv:1802.09464*, 2018.

[11] A. D. Procaccia. How is voting theory really useful in multiagent systems? *available online, URL: http://www. cs. cmu. edu/arielpro/papers/vote4mas. pdf (DOA: 15.01. 2013)*.

[12] T. Schaul, D. Horgan, K. Gregor, and D. Silver. Universal value function approximators. In *International Conference on Machine Learning*, pages 1312–1320, 2015.

[13] R. S. Sutton, J. Modayil, M. Delp, T. Degris, P. M. Pilarski, A. White, and D. Precup. Horde: A scalable real-time architecture for learning knowledge from unsupervised sensorimotor interaction. In *The 10th International Conference on Autonomous Agents and Multiagent Systems*, 2011.