Decentralized Reinforcement Learning Inspired by Multiagent Systems

Doctoral Consortium

Dhaval Adjodah
MIT Media Lab
dval@mit.edu

KEYWORDS

Decentralized optimization, Reinforcement learning, Collective intelligence, Multiagent Systems

ACM Reference Format:
Dhaval Adjodah. 2018. Decentralized Reinforcement Learning Inspired by Multiagent Systems (Doctoral Consortium). In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 2 pages.

1 MOTIVATION

Existence can perhaps be viewed an exercise of searching high-dimensional, rugged, and approximated (using training data) landscapes for (often time-delayed) rewards. Bounded rationality imposes limits on the success of solutions that can be found by agents acting alone, causing them to potentially get stuck in 'effective local minima' [14]. To overcome these limits, agents can communicate and work together.

Historically, machine learning problems and algorithms were far enough from these limits that all problems could be abstracted as a single agent/model which was optimizing a loss using signals from the environment. However, we are now entering an era where the theoretical and engineering insights from multiagent systems and collective intelligence are becoming, again, critical for the continued growth and usefulness of large-scale real-world machine learning.

Theoretically, for example, modern reinforcement learning algorithms and problems are now high-dimensional and rugged enough that a collection of agents are often run in parallel (sometimes asynchronously) to speed up training because learning is fundamentally experience-based - the diversity and uniqueness of search trajectories of agents is of prime importance. Beyond reinforcement learning, machine learning models have gained from being trained as a collective through approaches ranging from student-teacher mechanisms (to transfer learning more effectively between agents), to population/evolutionary methods (to search more broadly the landscapes).

Engineering-wise, because we are still far from replicating human intelligence, we must build better interfaces for how humans and algorithms could collaborate for increased performance. But because humans are known to have very low communication bandwidth and are prone to biases, work must be done to understand not only how humans learn individually (for which there is extensive neuroscience, cognitive science, etc) but also how they learn from each other - what kind of data and model approximations do they make (and the biases and ensue), what network topologies are best for collaboration or exploration, what cognitive models do they use to sample and update their beliefs, etc. For example, personal AI assistants need to be able to learn more seamlessly from humans interaction.

2 COMPLETED WORK

2.1 Networked Evolution Strategies

Deep reinforcement learning algorithms run many learning agents in parallel to speed up learning and to minimize use of correlated data. There is evidence that the network structure of communication between nodes significantly affects the convergence rate and accuracy in decentralised optimization [5,6]. To our knowledge, no work has explored theoretically and experimentally how the topology of communication between learning agents affects deep reinforcement learning. In this work, we introduce the notions of ensembles, network topology and independent node-level agent updates to the Evolution Strategies paradigm. We prove that to sample the search space efficiently (parametrized by the variance of parameter updates), agents need to communicate within certain families of sparse network topologies. Our key findings [1] and contributions are as follows:

1. We derive Monte-Carlo estimates for update rules for fully-connected and sparsely-connected inter-agent learning based on biased inter-agent sampling, and additionally provide an upper bound for the variance of the Monte-Carlo estimate over a population of agents, which suggest sparser networks for higher variance.
2. Using this sparser family of communication networks, we observe faster and higher learning than when using fully-connected networks. Because the networks are sparser, learning incurs a lower communication cost.
3. We observe that this family of networks result in a multiplicative effect in total reward: networks with only 1,000 agents produce results competitive to fully-connected networks with 4,000 agents.
4. We find that sparser graphs can achieve up to 33.5% higher reward than a corresponding fully-connected network, and that they can reach the fully-connected maximum up to 32% earlier.

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.

2.2 Bayesian Optimality in the Wisdom of the Crowd

Fundamental to improving how human groups function is to understand how humans build and sample from their internal belief distributions, how they update their belief distribution after observing those of their peers, the conditions under which these sampling and update strategies fail, and how to aggregate each individual's belief into a collective belief. Although there has been extensive earlier work on how secondary factors (such as the effect of confidence [10,11]) affect the accuracy of groups estimates, there has been limited research on modeling the update and sampling procedure themselves (comparing model prediction to individual prediction), and on investigating how these procedures affect individual and group accuracy (comparing to the ground truth). In this work, we build upon the literatures of both cognitive science and the 'Wisdom of the Crowd' with the goal of modeling how humans learn from and influence [2,3,4,7,8,9] each other in order to understand how group accuracy emerges. We collect a large novel dataset (17K predictions from 2K people) and investigate a large variety of update and sampling models and find that, surprisingly, simple conjugate normal models do best at fitting the belief update and sampling behavior of humans. Because we also have data of the same individuals over different rounds, we also find that there is a collective tuning process where the more inaccurate individuals were in the past, the more they learn to trust the group's belief. We then reproduced previous subsampling strategies for improving the WoC strategies: we selected individuals that have been shown to be historically accurate (known as 'superpredictors'), and individuals that are resistant to social influence [12] and find that - although improvements can indeed be reproduced - they are dwarfed in comparison to when individuals are selected as per our novel metric of how far they are from the optimally Bayesian prediction. We then demonstrate that these results can also be observed in a separate dataset collected from a large prediction market, where we also show how to estimate influence in the absence of an explicit social signal (as in our data collection) using both heuristic geometrical and HMM models. However, we also find that social learning and improvement breaks down when the belief distribution of others is ambiguous (not statistically unimodal): when this signal of collective agreement is absent, then humans are then better off not updating their belief based on those of their peers.

3 FUTURE WORK

Regarding how to coordinate and improve learning between multiple AI agents, I am interested in two directions. One significant contribution would be to mathematically formalize learning between AI nodes, perhaps as a distributed optimization problem (building upon the formalism of Simulated Annealing or Free Energy minimization), or as an information-theoretic multi-sensor problem. Another direction is to forgo the static network topology and rethink the problem as a dynamic rewiring problem, where the network topology is also learned.

REFERENCES

(1) Adjodah, D., Calacci, D., Krafft, P., Moro, E., Pentland, S. *Improved Learning in Evolution Strategies via Sparser Inter-Agent Network Topologies* NIPS Deep Reinforcement Learning Symposium 2017

(2) Adjodah, D., Leng, Y., Chong, S.K., Krafft, P., Pentland, S. *Social Bayesian Learning in the Wisdom of the Crowd* International Conference on Computational Social Science 2017

(3) Adjodah, D., Chong, S.K., Leng, Y., Krafft, P., Pentland, S. *Large-Scale Experiment on the Importance of Social Learning and Unimodality in the Wisdom of the Crowd* Collective Intelligence Conference 2017

(4) Adjodah, D., Krafft, P., Moro, E., Pentland, S. *Cognitive Limitations in Financial Networks* International School and Conference on Network Science 2017

(5) Adjodah, D., Pentland, S. *Long-Term Effects of Social Influence on Behavior Change. In Review*

(6) Krafft, P., Adjodah, D., Pentland, S., Tenenbaum, J. *Particle Sharing across Heterogeneous Models. In Review*

(7) Adjodah, D., Krafft, P., Noriega, A., Pentland, S. *Harnessing Social Learning to Improve Crowd-Sourced Prediction* International Conference on Computational Social Science 2016

(8) Della Penna, N., Adjodah, D., Pentland, S. *Efficiency in prediction markets, evidence from SciCast.org* The Cutting Edge: Applications of Collective Intelligence, Collective Intelligence 2015

(9) Adjodah, D., Pentland, S. *Understanding Social Influence Using Combined Network Analysis and Machine Learning Models* NetSci 2013

(10) Lorenz, Jan, et al. *How social influence can undermine the wisdom of crowd effect.* Proceedings of the National Academy of Sciences 108.22 (2011): 9020-9025.

(11) Moussad, Mehdi, et al. *Social influence and the collective dynamics of opinion formation.* PloS one 8.11 (2013): e78433.

(12) Madirolas, Gabriel, and Gonzalo G. de Polavieja. *Improving collective estimations using resistance to social influence.* PLoS computational biology 11.11 (2015): e1004594.

(13) Bengio, Yoshua. *Evolving culture versus local minima.* Growing Adaptive Machines. Springer, Berlin, Heidelberg, 2014. 109-138.