Statement of Purpose

Qiushi (Max) Lin

My research interests are broadly in theory and applications of reinforcement learning. Reinforcement learning handles the challenges of sequential decision-making through interaction with environments. State-of-the-art deep reinforcement learning methods have a wide range of real-world applications, yet most of them lack theoretical foundations. In my Ph.D., I plan to advance our theoretical understanding of reinforcement learning algorithms and, based on that, develop reinforcement learning algorithms that are both theoretically principled and practically relevant.

In fact, such motivation comes from my research experience in the empirical world of reinforcement learning. I have been working with Prof. Hang Ma for the past two years. We focused on designing practical and efficient algorithms for partially observable multi-agent path finding. Although recent works have proposed several reinforcement learning-based methods for this task, they cannot generalize to large-scale instances, and complex cooperation among agents is difficult to achieve. I designed a form of explicit cooperative guidance based on heuristics and extended the multi-agent actor-critic framework to enhance the generalizability of the learned model. This proposed algorithm arises from the marriage of the planning algorithm (heuristic search) and the reinforcement learning framework (actor-critic), and based on the extensive empirical evaluation, this algorithm achieves outperformance compared to other existing works in various benchmarks, which leads to the publication of my first research paper [LM23b].

For our second line of research, I extended learning-based methods to another variation of multi-agent path finding, moving agents in formation. This task requires agents to optimize over two different objectives of path planning and formation control. The challenges come from the curse of dimensionality and the complication of balancing two given objectives during decentralized execution. To address these issues, inspired by the literature on mean field reinforcement learning and multi-objective reinforcement learning, I proposed a novel framework that can handle the bi-objective optimization in large-scale multi-agent reinforcement learning. Using this method, the learned policy can lead to a greater solution quality than centralized planning algorithms in most cases and have the flexibility to handle dynamic formations. Based on that, we wrote another research paper [LM23a] which has been submitted and is currently under review.

Although I love the idea of connecting the field of reinforcement learning with multi-agent path finding, I am not satisfied with the underlying rationale of the proposed learning algorithms. Oftentimes, I would find the details in code-level implementation have a much stronger impact on the performance than the actual algorithm design, which could be somewhat demotivating. I would much rather work in a field in which every concept is rigorously defined and the performance of algorithms is measurable, predictable, and analyzable. Carrying the unsatisfactory from my past research, I took two graduate-level theory courses, *Theoretical Foundations of Reinforcement Learning* and *Optimization for Machine Learning*, taught by Prof. Sharan Vaswani at SFU. In these courses, I learned a wide range of knowledge for machine learning theory, from convex optimization to online learning, from multi-armed bandits to policy gradient methods, etc. Exploring all related literature opens another door for me to understand reinforcement learning from a whole new perspective, and I find myself deeply fascinated by the theory literature. I also realize there are so many unanswered open questions in this field, and without solving these questions, researchers would never get a hold of the essence of algorithm design for reinforcement learning.

Motivated by my appreciation for theory and my confusion about empiricism, I made my first attempt at doing research in reinforcement learning theory. Guided by Prof. Sharan Vaswani, my teammates and I focused on analyzing the convergence rates for policy gradient methods with linear function approximation. We developed a general framework to derive convergence rates of policy gradient methods for log-linear policies and reduced the problem to tabular softmax settings. With this recipe, we extended theoretical guarantees of softmax policy gradient methods to derive theoretically guaranteed algorithms for log-linear policies with both exact and inexact policy evaluation. To the best of our knowledge, we are the first to develop algorithms that have similar convergence rates for log-linear policies compared to softmax policy gradient methods. This experience is extremely rewarding for me, as it is the first time I could propose algorithms based on solid theoretical foundations. We summarized our results in this report [LAdAV23].

Even though the pivot of my research interests towards reinforcement learning theory was recent, I would argue that my motivation has always been a natural match for the subject. My favorite problems to think about are those that are abstract and well-defined, and my favorite solutions to problems are those that are simple and elegant, yet require deep and sophisticated thoughts. Such inclination provides more joy for me to do theory research than empirical reinforcement learning. Moreover, through hard studying in Prof. Sharan Vaswani's two theory courses, I have been equipped with basic knowledge to understand most cutting-edge theory literature and to potentially come up with novel ideas, and now through this application, I hope I can find a place to continue doing so.

Reinforcement learning algorithms can not only be impactful in practice but also beautiful and elegant in theory. In the future, I aim to develop theoretically principled learning algorithms that could potentially be relevant to real-world applications. Getting a Ph.D. would allow me to continue working in this field and to acquire the skills and maturity I need to become a Professor in Computer Science. Thank you for your consideration.

References

[LAdAV23] Qiushi Lin, Matin Aghaei, Anderson de Andrade, and Sharan Vaswani. On the Convergence Rates of Log-Linear Policy Gradient Methods. 2023. Preprint on Webpage at `https://qiushi-lin.github.io/documents/lin2023llpg.pdf`.

[LM23a] Qiushi Lin and Hang Ma. Mean Field Control with Envelope Q-learning for Moving Decentralized Agents in Formation. 2023. Preprint on Webpage at `https://qiushi-lin.github.io/documents/lin2023mfceq.pdf`.

[LM23b] Qiushi Lin and Hang Ma. SACHA: Soft Actor-Critic with Heuristic-Based Attention for Partially Observable Multi-Agent Path Finding. *IEEE Robotics and Automation Letters*, 8(8):2377–3766, 2023.