Research Statement

Arambam James SINGH

My primary research interest lie in the area of reinforcement learning (RL). Recent advancements in reinforcement learning research have lead to achievement of crucial milestones in the field of artificial intelligence. Especially, the major breakthroughs so far have been on single agent settings where one agent learn to solve a task. However, in many real world problems such as maritime traffic navigation, air-traffic control, road traffic control etc., many learning agents operate in the same environment. In a multiagent environment, each individual agent take independent decisions based on local observations to achieve a global objective or to maximize individual reward. Such systems are ubiquitous in real-world applications and can potentially impact many business sectors. Therefore, I emphasize the majority of my research to study the problem of large scale multiagent decision-making under uncertainty. Particularly, I focus on the fundamental problems arising in multiagent systems with a large number of agents and develop effective solutions based on multiagent reinforcement learning framework.

Current Research

The study of multiagent system can be broadly classified into — cooperative, competitive and mixed. Currently my research mainly focuses on the cooperative setting, where all agents jointly maximize a global reward. Such cooperative multiagent systems are challenging because of the infamous *multiagent credit assignment* problem. Since all agents receive the same global reward based on their joint action, each agent's individual contribution (and its specific actions) to the global reward is not clearly determined. Even in settings where agents receive local rewards, assigning appropriate credit to each agents is still difficult because of the interactions among agents. This make the multiagent learning challenging and sample inefficient. Generally, the credit assignment problem becomes much more challenging as the number of agents increases. Existing works either do not scale well to large numbers of agents [4, 5], or their credit assignment mechanisms are not efficient [6, 7]. One of the main objective in my PhD thesis is bridging the gap of existing approaches by developing scalable and effective algorithms to address the problem of credit assignment in large scale cooperative multiagent systems.

I use the maritime vessel traffic management problem as one of the motivating application domains for large scale multiagent system. The key research question is how to coordinate vessels in a heavily trafficked maritime traffic environment, such as the Singapore Strait, to increase the safety of navigation by reducing traffic hotspots. I formulate the problem as a multiagent reinforcement learning problem with cooperative setting, where all vessel agents collaborate to achieve a global objective of less traffic congestion and high traffic throughput. In addition to the underlying problem of multiagent credit assignment, maritime traffic control problem also exhibit other challenges such as — vessels unlike car can never stop, highly dynamic environment due to unknown factors (e.g., weather conditions, tidal waves), *variable action duration*: action that takes variable amount of time to complete. I use travel time to cross a

zone (a sea space) as an action, and due to highly dynamic nature of vessel movement, the travel time is stochastic in nature. In my solution approach, I exploit the anonymity structure present in multiagent system with large number of agents, i.e an agent's behavior is mainly influenced by the aggregate information about the neighboring agents rather than its identity. Such a structure can be observed in the maritime traffic problem also where a vessel agent's decision-making process is largely affected by the number of other vessels present in the vicinity. To this end, I first propose an aggregate based multiagent model for the maritime traffic control problem. Then, I develop an aggregate based realistic maritime traffic simulator which is used for training a policy gradient based method to provide an effective solution strategy. I also address the credit assignment problem by developing a *vessel-based* value function, which performs effective credit assignment by computing precisely the effectiveness of the agent's policy by filtering out the contributions from other agents. However, computing this value function is computationally expensive because it requires to keep track of other vessel's state-action pair. To address this challenge, I propose an alternative aggregate based method to compute the value function by exploiting the anonymity structure. Through extensive evaluation on both synthetic and real world problem instances, the results demonstrate the efficacy our proposed approach. This work was accepted at AAAI-2019 [3].

Although [3] achieved promising results, its ability to handle variable action duration is rather limited, which is a crucial feature of the problem domain. Thus, in the next chapter of my dissertation, I address this challenge using hierarchical reinforcement learning (HRL), a framework for control with variable action duration. In HRL, there is a notion of high level actions that can take variable amount of time to complete, unlike primitive actions which are executed at every time step. One key benefit of HRL is structured exploration, i.e exploration using high level actions rather than just primitive actions. By exploiting this property, I develop a novel hierarchical learning based approach for the maritime traffic control problem. Particularly, I introduce the notion of high level actions that intuitively corresponds to different traffic situations. Each high level action provides a mapping to a low level navigation action that provide vessels a recommended travel time to cross a zone. Using such high and low level policies, I showed both theoretical advantages (such as better exploration while learning using the high level action policy), and empirical benefits on both synthetic and real world datasets. This work was accepted at AAMAS-2020 [2].

In my dissertation, I also developed a general approach to address the credit assignment problem for any large scale cooperative multiagent system for both discrete and continuous actions settings. The proposed methodology is motivated by a popular technique of *difference rewards* (DR) [8]. DR is computed as the difference between the global reward and a counterfactual reward when the particular agent's impact is removed from the system, quantifying the agent's contribution to the global reward. However, computing DR is challenging because it requires access to the actual reward model (not available in a model free setting), or performing additional simulations, which are computationally challenging. First, I propose a novel approach to learn a differentiable reward model by exploiting the collective nature of interactions among agents. Using the learned reward model, I propose a scalable approach to estimate DR for effective multiagent credit assignment in large multiagent systems. Unlike previous techniques,

our method does not require domain expertise or extra simulations, and is highly scalable with the number of agents. I evaluate our proposed approach on air-traffic control problem, which is also a large scale multiagent systems. On a wide range of synthetic and real world instances, our proposed method significantly improve the performance against several other competing algorithms. This work is accepted at ICAPS-2021 [1]. A complete list of my research publications is provided in my resume.

Future Directions

In future, I am interested in continuing my studies on multiagent RL because of its nature in many real world problems. While there are many existing works on multiagent RL, they are not scalable to large agent settings. The progress on scalable multiagent RL research is relatively limited, which is one of my main research direction.

I am also interested in developing better theoretical understanding of the field. Despite recent successes, RL's real world applications are limited compared to other sub-field of machine learning such as supervised learning. There are several reasons behind it — poor sample efficiency of learning algorithms, having access to realistic simulator of the problem domain, safety issues during training a real world agent e.g., physical robot(s), etc. I am particularly interested in studying the sample complexity issues in RL algorithms by exploring the directions of hierarchical RL and distributional RL.

In summary, my longterm research goal is to develop theoretically sound and effective research methodologies to advance the field of artificial intelligence.

References

1. **Arambam James Singh**, Akshat Kumar, Hoong Chuin Lau, *Learning and Exploiting Shaped Reward Models for Large Scale Multiagent RL*. Accepted at International Conference on Automated Planning and Scheduling (ICAPS-2021).
2. **Arambam James Singh**, Akshat Kumar, Hoong Chuin Lau, *Hierarchical Multiagent Reinforcement Learning for Maritime Traffic Management*. In Proceedings of the International Conference on Autonomous Agents and MultiAgent Systems (AAMAS-2020).
3. **Arambam James Singh**, Duc Thien Nguyen, Akshat Kumar and Hoong Chuin Lau, *Multiagent Decision Making For Maritime Traffic Management*. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI-19).
4. Rashid, Tabish et al. *QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning*. In Proceedings of the International Conference on Machine Learning, (ICML-2018).
5. Shariq Iqbal and Fei Sha. *Actor-attention-critic for multi-agent reinforcement learning*. In Proceedings of the International Conference on Machine Learning, (ICML-2019).
6. Yaodong Yang et al. *Mean Field Multi-Agent Reinforcement Learning*. In Proceedings of the International Conference on Machine Learning (ICML-2018).
7. Duc Thien Nguyen et al. *Credit assignment for collective multiagent RL with global rewards*. In Advances in Neural Information Processing Systems (NeurIPS-2018)
8. Wolpert et al. *Optimal payoff functions for members of collectives*. In Modeling complexity in economic and social systems. World Scientific. 355-369, 2002.