## Learning to Communicate with Deep Multi-Agent Reinforcement Learning

Multiple agents interacting in partially-observable settings need to collaoborate in order to maximize joint payoffs. Collaboration requires an effective communication protocol which provisions flow of information among agents. To this end, the work presents two novel schemes which facilitate learning of communication protocols using gradient-based information. Reinforced Inter-Agent Learning (RIAL) and Differentiable Inter-Agent Learning (DIAL) backpropogata error derivatives of communication messages within and among agents respectively. Adoption of RIAL and DIAL on challenging partially-observable tasks demonsrtates the need for communication.

The communication framework leverages the setup of centralized training with decnetralized control wherein multiple agents collaborate with each other to jointly optimize payoffs. RIAL and DIAL are incorporated as multi-agent schemes in this setup which make use of Q-Learning and backpropagation in noisy communicative channels in agents respectively. Firstly, RIAL consists of agents as deep rcurrent Q-networks which observe environment observations and messages from other agents. The action selector evaluates Q-values and selects environment and communicative actions which are passed on other agents at the next timestep. Flow of gradient in RIAL takes place within agent which is trained end-to-end. Contrary to RIAL, DIAL makes use of a C-Net which outputs Q-values for environment actions and the communicative message directly. The message bypasses the action selector and proceeds to a Discrete Regularised Unit (DRU) which discretises it during decentralized execution. Discretised messages are then fed to the receiver agent at the next timestep. DIAL provisions the flow of gradient from receiver to sender via the DRU. Practical implementation of DIAL architecture consists of message embeddings which are passed through a task-specific network.

Adoption of RIAL and DIAL on challenging partially-observable scenarios such as riddle solving and MNIST games demonstrates the suitability and necessity of a communication framework between agents. Furthermore, the effectiveness of DIAL is demonstrated as a result of higher normalized returns on tasks and activations observed due to the presence of noise in DRU. While RIAL and DIAL demonstate iproved performance, they certainly leave room for improvement on various facets. Effectiveness of communicative protocols is demonstrated for small number of agents having prior memory of their actions. A more challenging experimental setup would consist of larger number of agents without the need for memory-based modules. This would help demonstate the effectiveness of messages among agents. Furthermore, the framework provides an analysis of complexity of communication protocols which scales linearly in the number of agents and quadratically in the total number of actions (environmental and commuicative). This indicates that the approach may not be scalable to larger number of agents or complex tasks requiring significant messages between different types of agents. Improving the scalability of protocols could be an interesting direction for future work.

Usage of communication protocols amongmultiple agents highlights the necessity for inter-agent communication in partially-observable settings. This provides two new directions for future work. Firstly, protocols could be improved for larger number of agents and message spaces. And secondly, inter-agent communication could be studied among novel partners at execution time.