Learning Structured Communication for Multi-Agent Reinforcement Learning

Colin Acker

October 2023

1 Notes

This paper proposes LSC, a solution to ineffective communication in multi-agent systems. We first consider a group of n agents. The first component of the LSC is the Structured Communication Module. "The structured communication module is designed by three principles: 1) agents in the same group are more likely to understand and cooperate inner group; 2) high-level agents are more likely to capture the global perception through the exchanged messages; 3) high-level agents are distributed sparsely to lower the communication cost." The Structured Communication Module includes the Weight Generator (WG) and the CBRP. The Weight Generator is a neural network, $f_{wg} : o_i \rightarrow w_i$, which determines the communication importance for each agent. The point of the WG is to determine the confidence of an agent to become high-level. Next, the CBRP takes $\vec{w}$ and the local geometry, and constructs a hierarchical communication structure. There are high-level agents (HLAs) and low-level agents (LLAs). The CBRP links LLAs to HLA, and makes sure that in each group, there is a single HLA.

Since the CBRP is non-differentiable, the paper proposes an auxilary RL task, where each agent's action corresponds to a weight choice in the task, and $w_i \in \{0, 1, 2\}$. The discreteness enables a task-driven, closed-loop communication weight generation. The paper uses Q-networks for the weight generator, and uses loss function:

$$l(\theta^w) = \mathbb{E}_{\vec{o},\vec{w},r,\vec{o}}[\sum_{i=1}^{n}(Q_{\theta w}(o_i, w_i) - y_i)^2]$$
$$y_i = r_i + \gamma \max_{\vec{\theta^w}} Q_{\theta w}(\vec{o_i}, \vec{w_i})$$

Once the communication network topology is determined, the communication-based policy module learns a global collaboration policy. The communication-based policy module consists of a GNN-based sub-module and a Q-net sub-module. The GNN-based sub-module, $f_{\theta^{gnn}}$ learns a policy based on communication messages and updates overall state perceptions. It employs a three-phase

Table 1: The proposed GNN-based communication architecture with three steps.

Type	Edge $(i \to j) \in \mathcal{E}$	Edge Update Scheme	Node Update Scheme
Step 1: intra-group aggregation	$i \in \mathcal{V}_l$, $j \in \mathcal{V}_h$	$e_{ij} = \phi(v_i^l)$, $\bar{e}_j = \rho(\{e_{ij}\}_{(i \to j) \in \mathcal{E}})$	$v_j^h = \phi(\bar{e}_j, v_j^l)$
Step 2: inter-group sharing	$i \in \mathcal{V}_h$, $j \in \mathcal{V}_h$	$e_{ij} = \phi(v_i^h, v_i^l)$, $\bar{e}_j = \rho(\{e_{ij}\}_{(i \to j) \in \mathcal{E}})$	$v_j^g = \phi(\bar{e}_j, v_j^l)$
Step 3: intra-group sharing	$i \in \mathcal{V}_h$, $j \in \mathcal{V}_l \cup \mathcal{V}_h$	$e_{ij} = \phi(v_i^g, v_i^h, v_i^l)$, $\bar{e}_j = \rho(\{e_{ij}\}_{(i \to j) \in \mathcal{E}})$	$v_i^l = \phi(\bar{e}_i, v_i^l)$, $v_j^l = \phi(\bar{e}_j, v_j^l)$

Figure 1: GNN-Based Communication Architecture

communication strategy. During intra-group aggregation, each LLA embeds its local perception (v_i^l) and transmits it to its connected HLA. The HLAs aggregate the information they received and obtains the group perception (v_i^h). Next, during inter-group sharing the HLAs communicate with each other and obtain the global perception (v_i^g). Lastly, during intra-group sharing, each HLA embeds its local, group, and global perception as a message and sends it to each connected LLA. The LLAs then update their local perceptions.

After this, all that is left is to learn a policy based on the new state perceptions. The paper uses a Q-Net for each agent i, $Q^i_{\theta^Q}$. Note that each Q is parameterized by θ^Q. The gradient can be back-propagated from Q-net to the GNN, so the loss of the communication-based policy module is:

$$l(\theta^Q, \theta^{gnn}) = \mathbb{E}_{\vec{o},\vec{a},r,\vec{o}}[\sum_{i=1}^{n}(Q^i_{\theta Q}(f_{\theta^{gnn}}(\vec{o}), \vec{a}_i) - y_i)^2]$$

$$y_i = r_i + \gamma max_{\vec{a}_i} Q^i_{\theta Q}(f_{\theta^{gnn}}(\vec{o}), \vec{a}_i)$$

They tested their results on two different multi-agent environments. They showed that the LSC outperforms pre-trained agents. Also, they showed that the Weight Generator provided benefits, as it significantly outperformed a fixed weight generator. They compared LSC with multiple other communication-topology architectures, and found that LSC performed best, likely because of the 3-step process in the GNN-based sub-module.