Improvement of the Performance Using Received Messages on Learning of Communication Codes

Tatsuya Kasai Hayato Kobayashi Ayumi Shinohara
Graduate School of Information Sciences, Tohoku University, Japan

Introduction

In Multi-Agent Reinforcement Learning (MARL), each agent learns a cooperative policy \( \pi : S \rightarrow A \), where \( S \) and \( A \) are a set of states and a set of actions, respectively. If we utilize communication to facilitate multi-agent coordination, we must construct communication codes so that agents can communicate with each other. However, it is a hard task since we usually do not know workable communication codes and/or information on unknown problems. **We focus on a method that allows agents to learn communication codes autonomously.**

![Diagram](https://via.placeholder.com/150)

**Extension of learning method**

*Previous work*

**Signal Learning** (SL) [kasai08] allows agents to learn communication codes autonomously in MARL framework, where \( M \) is a set of messages whose meanings are not predetermined explicitly. In SL, agents can learn **two policies** as follows, concurrently.

Communication policy

\[ \pi_c : S \rightarrow M \]

Action policy

\[ \pi_a : S \times M \rightarrow A \]

**This work**

Our extension is just the change of \( \pi_c \) from \( \pi : S \rightarrow M \) to \( \pi : S \times M \rightarrow M \). We call this method **SL with Messages** (SLM).

**Comparisons of SL and SLM**

**Example problem**

The goal of the problem is that both agents, starting from their own SG states, go back to the SG states after activation (Figure 2, 3).

**Experiments**

We carried out experiments for comparing SL with SLM, where \(|M|\) is varied from 2 to 10.

**Discussion**

By comparing NC with SL, SL is clearly better than NC (Figure 4). This shows that some beneficial meaning emerges in messages in M through the learning processes in SL. In SL, \( \pi_c : S \rightarrow M \) probably allows each agent to include its own state in a message.

By comparing SL with SLM, SLM is clearly **better and more robust** than SL (Figure 4, 5). This means that SLM can allow each agent to include much more information in a message than SL.

By using an (deterministic) optimal policy, both agents reach the goal with the minimum number of steps. However, agents must remember the status of button for acquiring the deterministic optimal policies. In SLM, the messages should contain the information of status of button.

Table 1 shows the percentage of the successful trials in all 100 trials. As shown in the Table 1, SLM has the ability to acquire an deterministic optimal policy. Actually, SLM can allow the agent to get a deterministic optimal policy. Table 2 shows a simplest example of the acquired optimal policies (\(|M|=2\) in SLM).

**Conclusion**

We proposed SLM, and empirically showed that the performance was improved dominantly by using SLM, which is an extension of SL. In addition, we confirmed that SLM has the ability to acquire a deterministic optimal policy, which cannot be achieved by SL.