Intrinsic Motivated Multi-Agent Communication

Chuxiong Sun, Bo Wu, Rui Wang, Xiaohui Hu, Xiaoya Yang, Cong Cong
The Institute of Software, Chinese Academy of Sciences, China

Introduction

Recently, Multi-Agent Reinforcement Learning (MARL) has enjoyed great attentions in the literature.

The Challenges of MARL

- Scalability->CTDE
- Team Reward->Credit Assignment
- Local Observation->Communication

The Challenges of Communication

- How to extract information from local observations
- How to evaluate the importance of observed information

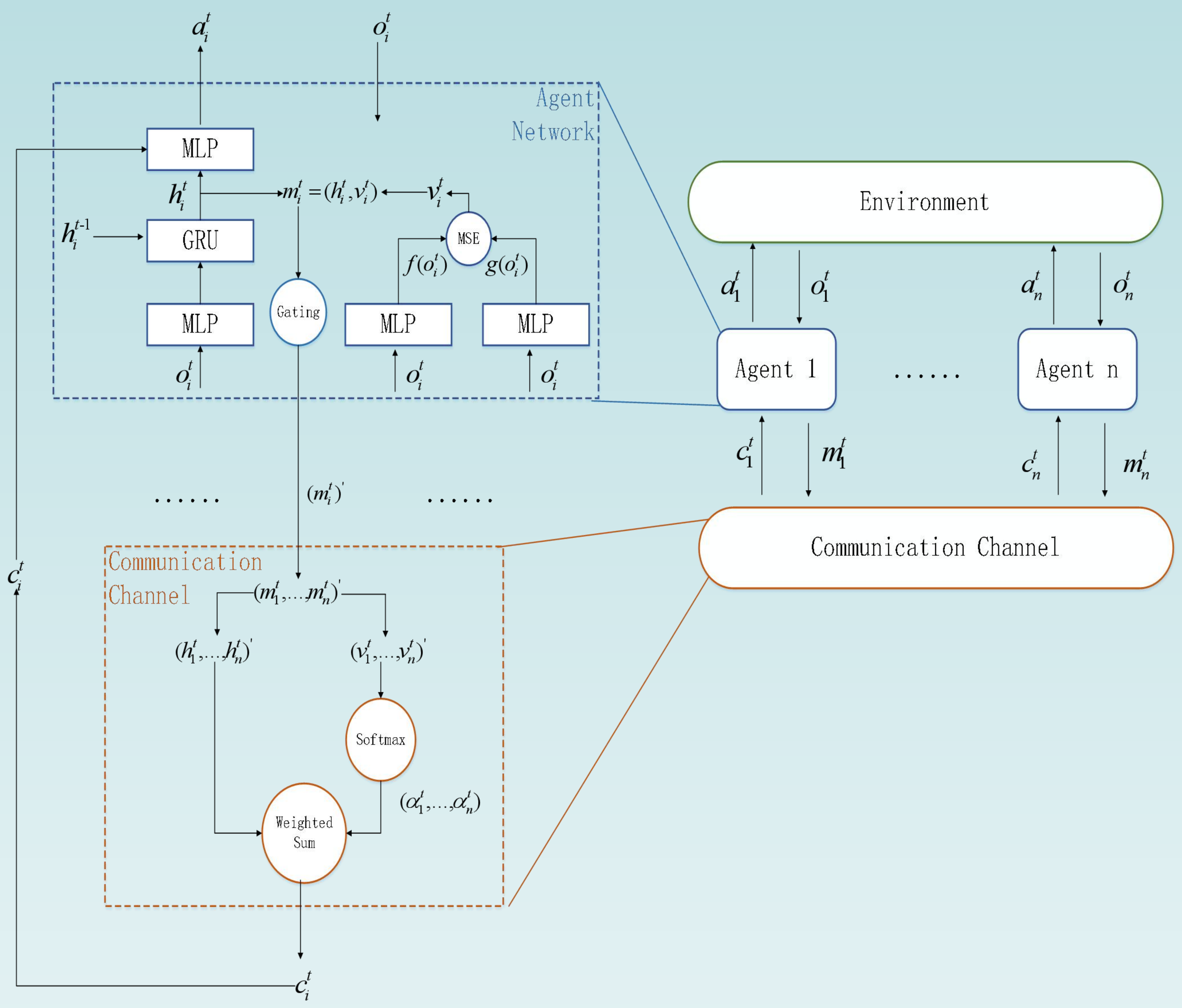
The Motivation of Communication

- The existing works can be summarized as 'Communicate what rewards you'.
- In this work, we propose a novel communication mechanism called 'Communicate what surprises you'.

Furthermore, we present a novel value-based communication framework /contribution

- The policy network is responsible for making decisions based on local observations and incoming messages.
- The intrinsic network is designed to measure the intrinsic importance of observed information.
- The gating mechanism is responsible for pruning useless messages.
- The attention communication channel is designed to integrate incoming messages.

Method



- At first, we use the mechanism proposed by [1] to measure the intrinsic importance of observed information.

$$v_i^t = f(o_i^t;\theta_f) - g(o_i^t;\theta_g)$$

- Furthermore, the message in our framework consists of two elements.

$$m_i^t = [h_i^t, v_i^t]$$

- Each agent will share the observed information to others when the intrinsic importance is larger than a threshold.
- Then the communication channel would leverage the intrinsic importance to compute an attention vectors for incoming messages.

$$(\alpha_1^t,\ldots,\alpha_n^t) = soft\max(v_1^t,\ldots,v_n^t)$$

- Then the contents of shared information are aggregated using the intrinsic attention vectors.

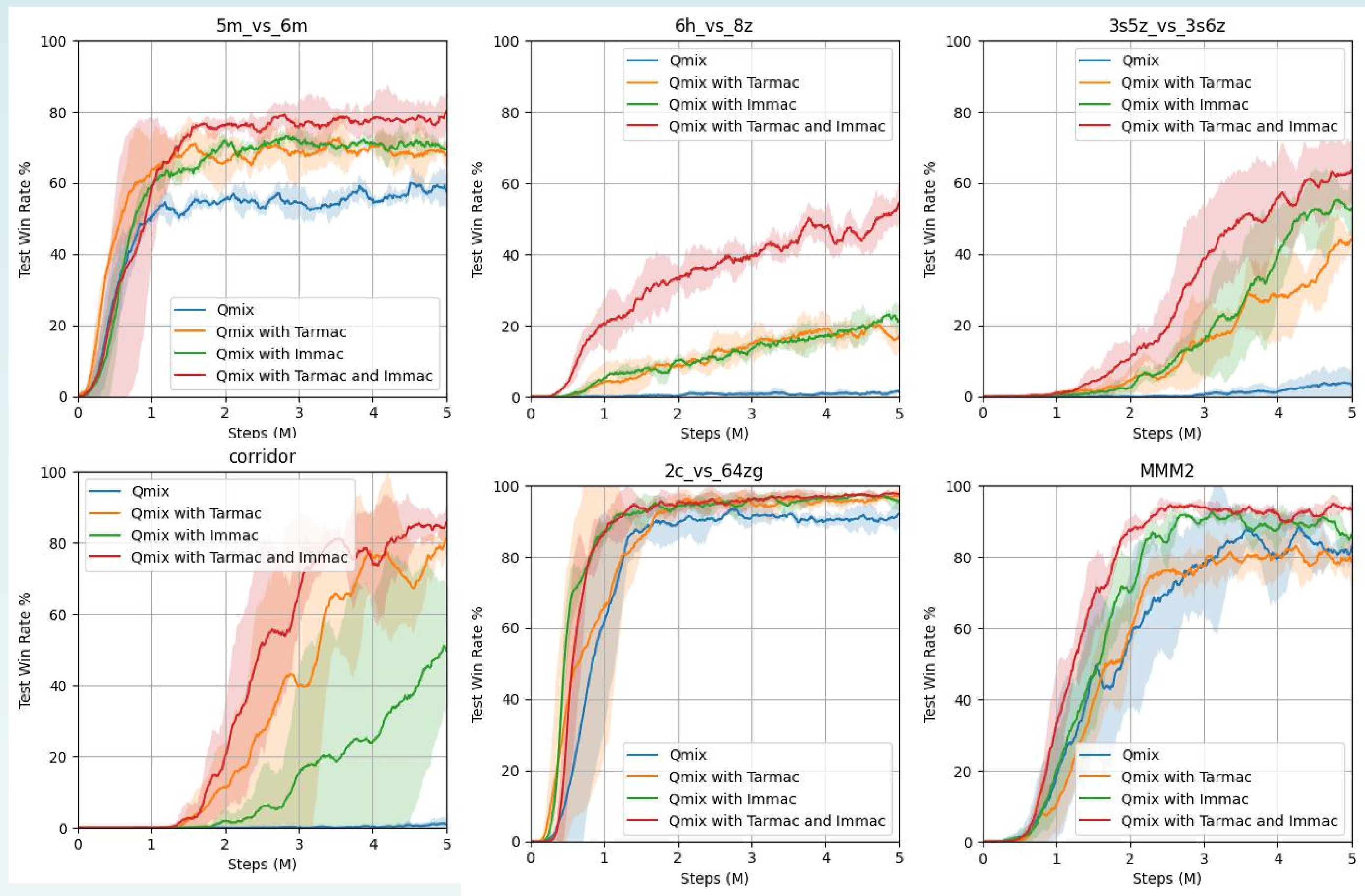
$$c_i^t = \sum_{i=1}^{k} \alpha_i^t h_i^t$$

- At last, the integrated message c_i^t is combined with agent's local observation o_i^t, then fed into policy network.

$$a_i^t = \pi_i(o_i^t, c_i^t)$$

Results

In this work, we use Qmix [2] without communication and Qmix with Tarmac[3] (i.e. Qmix improved by extrinsic motivated communication) as baselines. Then, we evaluate the proposed intrinsic value based attention mechanism on the six challenging scenarios from SMAC [4]. The detailed results are illustrated in the following figure. Furthermore, we leave the more comprehensive evaluation of IMMAC including the performance of intrinsic motivated gating mechanism in the future work.



References

[1]Exploration by random network distillation. arXiv preprint arXiv:1810.12894 (2018).

[2]QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning. arXiv preprint arXiv:1803.11485 (2018).

[3]Tarmac: Targeted multi-agent communication.

In International Conference on Machine Learning. 1538–1546.

[4]The starcraft multi-agent challenge. arXiv

preprint arXiv:1902.04043 (2019).