Efficient Communication in Multi-Agent Reinforcement Learning via Variance Based Control

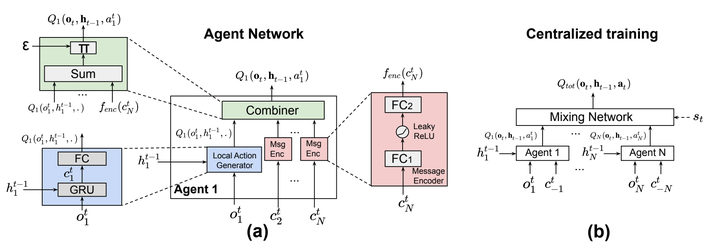
Sai Qian Zhang[1], **Qi Zhang**[2], **Jieyu Lin**[3]

[1]**Harvard University** [2]**Amazon** [3]**University of Toronto**

Abstract

- Multi-agent reinforcement learning (MARL) has recently received considerable attention due to its applicability to a wide range of real-world applications.
- Full communication among the agents leads to a large communication overhead and latency, which is impractical for the real system implementation with strict latency requirement and bandwidth limit (e.g., real-time traffic signal control, autonomous driving, etc).
- In this work, we propose Variance Based Control (VBC), a simple yet efficient technique to improve communication efficiency in MARL.

Design of the agent network



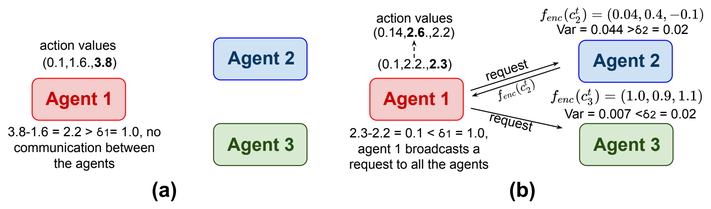
- The agent network consists of the following three networks: local action generator, message encoder and combiner.
- We employ a mixing network (shown in Figure 1(b)) to aggregate the global action value functions $Q_i(\mathbf{o}_t, \mathbf{h}_{t-1}, a_i^t)$ from each agents i, and yields the joint action value function, $Q_{tot}(\mathbf{o}_t, \mathbf{h}_{t-1}, \mathbf{a}_t)$.
- To limit the variance of the messages from the other agents, we introduce an extra loss term on the variance of the outputs of the message encoders $f_{enc}(c_j^t)$.
- The loss function during the training phase is defined as:

$$L(\theta_{local}, \theta_{enc}) = \sum_{b=1}^{B}\sum_{t=1}^{T}\left[(y_{tot}^b - Q_{tot}(\mathbf{o}_t^b, \mathbf{h}_{t-1}^b, \mathbf{a}_t^b; \boldsymbol{\theta}))^2 + \lambda \sum_{i=1}^{N} Var(f_{enc}(c_i^{t,b}))\right] \quad (1)$$

where $y_{tot}^b = r_t^b + \gamma max_{\mathbf{a}_{t+1}} Q_{tot}(\mathbf{o}_{t+1}^b, \mathbf{h}_t^b, \mathbf{a}_{t+1}; \boldsymbol{\theta}^-)$.

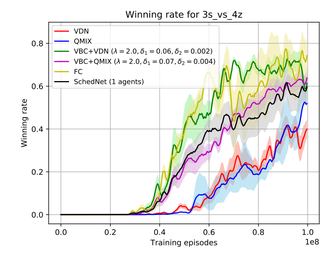
Communication Protocol

- Each agent only consults with the other agents when its confidence level on the local decision is low, and the other agents only reply when their messages can potentially change the final decision.
- The confidence level on the local decision is measured by computing the difference between the largest and the second largest element within the action values.
- When receiving the communication request, the agent replies to the request only when its message is informative, namely the variance of the message is high.

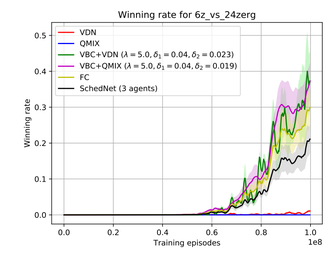


Evaluation

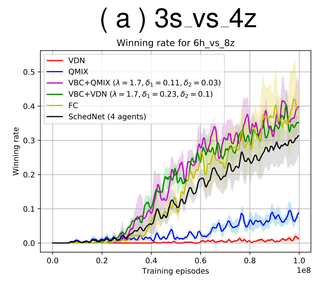
- We compare VBC and other banchmark algorithms, including VDN [3], QMIX [2] and SchedNet [1], for controlling allied units.
- We consider two types of VBCs by adopting the mixing networks of VDN and QMIX, denoted as VBC+VDN and VBC+QMIX.
- We notice that the algorithms that involve communication outperform the algorithms without communication in all the six tasks.
- VBC achieves the best performance with $2-10\times$ lower in communication overhead than the other algorithms.



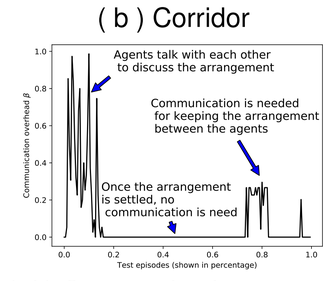
(a) 3s_vs_4z



(b) Corridor



(c) 6h_vs_8z



(d) Communication pattern

Conclusion

- By constraining the variance of the exchanged messages during the training phase, VBC improves communication efficiency while enables better cooperation among the agents.
- The test results of StarCraft Multi-Agent Challenge indicate that VBC outperforms the other state-of-the-art methods significantly in terms of both winning rate and communication overhead.

References

[1] D. Kim, S. Moon, D. Hostallero, W. J. Kang, T. Lee, K. Son, and Y. Yi. Learning to schedule communication in multi-agent reinforcement learning. *arXiv preprint arXiv:1902.01554*, 2019.

[2] T. Rashid, M. Samvelyan, C. S. de Witt, G. Farquhar, J. Foerster, and S. Whiteson. Qmix: Monotonic value function factorisation for deep multi-agent reinforcement learning. *arXiv preprint arXiv:1803.11485*, 2018.

[3] P. Sunehag, G. Lever, A. Gruslys, W. M. Czarnecki, V. Zambaldi, M. Jaderberg, M. Lanctot, N. Sonnerat, J. Z. Leibo, K. Tuyls, et al. Value-decomposition networks for cooperative multi-agent learning. *arXiv preprint arXiv:1706.05296*, 2017.

NeurIPS 2019, Conference on Neural Information Processing Systems, 8-14 December 2019, Vancouver, Canada