Intrinsic Motivated Multi-Agent Communication

Extended Abstract

Chuxiong Sun
The Institute of Software, Chinese Academy of Sciences
China, Beijing
chuxiong2016@iscas.ac.cn

Bo Wu
The Institute of Software, Chinese Academy of Sciences
China, Beijing
wubo2018@iscas.ac.cn

Rui Wang
The Institute of Software, Chinese Academy of Sciences
China, Beijing
wangrui@iscas.ac.cn

Xiaohui Hu
The Institute of Software, Chinese Academy of Sciences
China, Beijing
hxh@iscas.ac.cn

Xiaoya Yang
The Institute of Software, Chinese Academy of Sciences
China, Beijing
yangxiaoya17@mails.ucas.ac.cn

Cong Cong
The Institute of Software, Chinese Academy of Sciences
China, Beijing
congcong2020@iscas.ac.cn

ABSTRACT

Efficient communication is a promising way to achieve cooperation among agents in many real-world scenarios. However, aimless and motiveless information sharing may not work or even degrade the cooperative performance. Typically, the multi-agent communication behaviors are motivated by extrinsic rewards from environment. We conclude the mechanism as *'Communicate what rewards you'*. In this work, we present a novel communication mechanism called Intrinsic Motivated Multi-Agent Communication (IMMAC). Our key insight can be summarized as *'Communicate what surprises you'*. Concretely, we use an observation-dependent intrinsic value to represent the importance of observed information. Then a gating mechanism and an attentional mechanism based on intrinsic values are designed to control communication. By encouraging agent to communicate and focus on the observations with uncertain and important information, our algorithm achieves superior communication efficiency and cooperative performance. We evaluate IMMAC on a variety of challenging tasks, and demonstrate that intrinsic values are sufficient to drive efficient communication behaviors. Moreover, we found that the combination of intrinsic values and extrinsic values can further improve the communication efficiency. Consequently, intrinsic motivation is a promising way to control communication and it is capable of being a good complement to the existing extrinsic motivated communication methods.

KEYWORDS

Multi-Agent Reinforcement Learning; Multi-Agent Cooperation; Attentional and Gated Multi-Agent Communication; Intrinsic Motivation

ACM Reference Format:
Chuxiong Sun, Bo Wu, Rui Wang, Xiaohui Hu, Xiaoya Yang, and Cong Cong. 2021. Intrinsic Motivated Multi-Agent Communication: Extended Abstract. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, May 3–7, 2021*, IFAAMAS, 3 pages.

Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, Online.

1 INTRODUCTION

Essentially speaking, the purpose of communication is to improve the accuracy of decision-making by sharing the observed information. Consequently, how to extract information from local observations is the first challenge toward achieving efficient communication. However, there may exist useless information which can not aid in decisions or even degrade the cooperative performance. To this end, how to evaluate the importance of observed information is the second challenge in the literature of multi-agent communication. Typically, the existing communication protocols [7, 8, 11, 12, 14, 17, 19, 25, 29, 32, 33, 35] are trained or motivated by the extrinsic rewards from environment. To this end, the mechanism of existing works can be concluded as *'Communicate what rewards you'*. Concretely, it means that the observations and information which can help agents get more extrinsic return are more valuable to communicate. In this work, we propose a novel mechanism for communication. We utilize the agent's intrinsic uncertainty and curiosity about local observations to model the significance of shared information. We hold the view that the information generated by uncertain observations is also promising for communication and the observations with higher curiosity are deserved more attention. Our key insight can be concluded as *'Communicate what surprises you'*. It is worth remarking that the proposed intrinsic motivated communication is straightforward to combine with the existing extrinsic motivated communication. Furthermore, IMMAC should be regarded as a complement rather than an alternative to the existing algorithms without considering intrinsic values for communication.

2 METHOD

The purpose of communication is to overcome the difficulty of partial observability by information sharing. We hold the view that the information generated by novel and uncertain observations are more promising to communicate and deserved more attention than information extracted from familiar observations. Hence, the message m_i^t in our framework consists of two elements:

$$m_i^t = [\overbrace{h_i^t}^{information}, \underbrace{v_i^t}_{importance}] \tag{1}$$

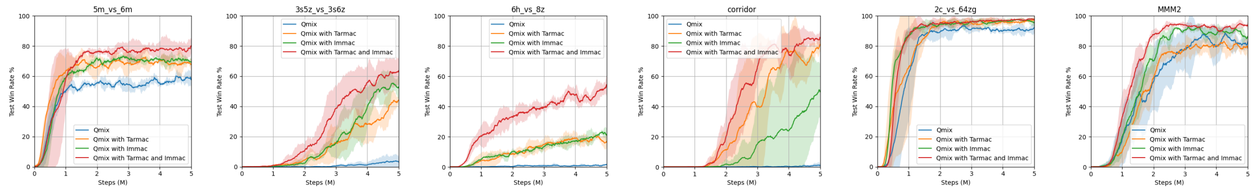

Figure 1: The learning curves of test win rates in SMAC scenarios. The shaded area represents 95% confidence intervals.

where h_i^t is the embeddings of local observations, we use it to represent the content of information; v_i^t is the output of intrinsic value network, it represents the intrinsic importance of the shared information. Concretely, the intrinsic importance are modeled based on Random Network Distillation [4].

During execution, $agent_i$ firstly encodes the observed information and measures the communicated values from local observations. Then the message m_i^t is passed through the gating mechanism which is designed to cut off unnecessary communication. In other words, the gating mechanism is required to decide whether to communicate based on current observations. Concretely, our framework can combine with any value-based gating mechanism, such as setting a threshold or designing more sophisticated rules. For convenience, we apply a simple heuristic based on v_i^t in this work. Each agent will share the observed information to others when the intrinsic importance is larger than a threshold δ. The gating mechanism endows agents with ability to decide when to communicate. The ability can help agents avoid unnecessary communication, reduce communication overhead and improve communication efficiency. It is especially promising in some real-world scenarios where the communicated resources (e.g. communication bandwidth and medium) are limited.

Then agents would send the messages to an attentional communication channel. The channel can be regarded as a shared communication medium which is responsible for integrating incoming messages then returning aggregated message to all agents. Concretely, the communication channel would leverage the intrinsic importance to compute an attention vectors for incoming messages.

$$(\alpha_1^t, ..., \alpha_n^t) = softmax(v_1^t, ..., v_n^t) \tag{2}$$

The attention weights would be high when the information is uncertain and important. Then the contents of shared information are aggregated using the intrinsic attention vectors:

$$c_i^t = \sum_{i=1}^{k} \alpha_i^t h_i^t \tag{3}$$

Obviously, the attentional information integration which allows agents to differentiate various messages is more sophisticated than the averaging combination. It endows agents with the ability to focus on information which can aid in their decisions. In addition, we adopt the paradigm of broadcast in this work (i.e. $c_1^t = c_2^t = ... = c_n^t$). At last, the integrated message c_i^t is combined with $agent_i$'s local observation o_j^t then fed into policy network.

$$a_i^t = \pi_j(o_i^t, c_i^t) \tag{4}$$

3 EXPERIMENT

In thiw work, we use Qmix [23] without communication and Qmix with Tarmac[7] (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 [24]. The detailed results are illustrated in Figure 1. Furthermore, we leave the more comprehensive evaluation of IMMAC including the performance of intrinsic motivated gating mechanism in the future work.

At first, we find that Qmix without considering communication presents a struggling performance in these scenarios. Especially in the four super hard tasks, Qmix almost fails to learn in 3 of them. On the other hand, the algorithms which take communication into account outperform Qmix by a large margin in almost all scenarios, except for $2c\_vs\_64zg$ which ally consists of only two units. The allied component may weaken the requirements of communication and result in the relatively smaller improvement. But the overall improvements in the other five scenarios are sufficient to demonstrate the effectiveness of communication. The shared information can significantly improve the quality of decision-making. Further, we surprisingly find that the intrinsic motivated communication can achieve comparable performance with extrinsic motivated communication. Concretely, IMMAC outperforms Tarmac by a considerable margin in $3s5z\_vs\_3s6z$ and MMM2, fails to match the performance in corridor and performs comparably in the other scenarios (i.e. $5m\_vs\_6m$, $6h\_vs\_8z$, $2c\_vs\_64zg$). In addition, the performances of Immac is better than Qmix without communication in all 6 scenarios. Overall, the results indicate that the intrinsic values can motivate efficient communication behaviors without considering any task-specific extrinsic signals. At last, we find that although there is a obvious difference in the performance of Tarmac and Immac, the combination of them almost achieve the best performance in all scenarios. It further demonstrates that intrinsic motivated communication is a good complement to extrinsic motivated communication. In other words, the intrinsic motivation and extrinsic motivation are different angles for evaluating the values of observations, but they are complementary. It is similar to the different senses of human beings which are corporate and complementary. The effective combination of them can largely aid in understanding the dynamic environment.

ACKNOWLEDGMENTS

This work is supported by the National Key Research and Development Program of China (No.2019YFB1405102) and the National Natural Science Foundation of China (Grant No.61802016).

REFERENCES

[1] Joshua Achiam and Shankar Sastry. 2017. Surprise-based intrinsic motivation for deep reinforcement learning. *arXiv preprint arXiv:1703.01732* (2017).

[2] Marc Bellemare, Sriram Srinivasan, Georg Ostrovski, Tom Schaul, David Saxton, and Remi Munos. 2016. Unifying count-based exploration and intrinsic motivation. In *Advances in Neural Information Processing Systems*. 1471–1479.

[3] Yuri Burda, Harri Edwards, Deepak Pathak, Amos Storkey, Trevor Darrell, and Alexei A Efros. 2018. Large-scale study of curiosity-driven learning. *arXiv preprint arXiv:1808.04355* (2018).

[4] Yuri Burda, Harrison Edwards, Amos Storkey, and Oleg Klimov. 2018. Exploration by random network distillation. *arXiv preprint arXiv:1810.12894* (2018).

[5] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1724–1734.

[6] Dennis Coon and John O Mitterer. 2012. *Introduction to psychology: Gateways to mind and behavior with concept maps and reviews.* Cengage Learning.

[7] Abhishek Das, Théophile Gervet, Joshua Romoff, Dhruv Batra, Devi Parikh, Mike Rabbat, and Joelle Pineau. 2019. Tarmac: Targeted multi-agent communication. In *International Conference on Machine Learning*. 1538–1546.

[8] Jakob Foerster, Ioannis Alexandros Assael, Nando De Freitas, and Shimon Whiteson. 2016. Learning to communicate with deep multi-agent reinforcement learning. In *Advances in neural information processing systems*. 2137–2145.

[9] Jakob N Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. 2018. Counterfactual multi-agent policy gradients. In *Thirty-second AAAI conference on artificial intelligence*.

[10] Rein Houthooft, Xi Chen, Yan Duan, John Schulman, Filip De Turck, and Pieter Abbeel. 2016. Vime: Variational information maximizing exploration. In *Advances in Neural Information Processing Systems*. 1109–1117.

[11] Jiechuan Jiang and Zongqing Lu. 2018. Learning attentional communication for multi-agent cooperation. In *Advances in neural information processing systems*. 7254–7264.

[12] Daewoo Kim, Sangwoo Moon, David Hostallero, Wan Ju Kang, Taeyoung Lee, Kyunghwan Son, and Yung Yi. 2019. Learning to schedule communication in multi-agent reinforcement learning. *arXiv preprint arXiv:1902.01554* (2019).

[13] J Zico Kolter and Andrew Y Ng. 2009. Near-Bayesian exploration in polynomial time. In *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 513–520.

[14] Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. 2016. Multi-agent cooperation and the emergence of (natural) language. *arXiv preprint arXiv:1612.07182* (2016).

[15] Qian Li and Zhichao Wang. 2017. Riemannian submanifold tracking on low-rank algebraic variety. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. 2196–2202.

[16] Ryan Lowe, Yi I Wu, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. 2017. Multi-agent actor-critic for mixed cooperative-competitive environments. In *Advances in neural information processing systems*. 6379–6390.

[17] Hangyu Mao, Zhengchao Zhang, Zhen Xiao, Zhibo Gong, and Yan Ni. 2020. Learning Agent Communication under Limited Bandwidth by Message Pruning. *AAAI 2020 : The Thirty-Fourth AAAI Conference on Artificial Intelligence* 34, 4 (2020), 5142–5149.

[18] Jarryd Martin, Suraj Narayanan Sasikumar, Tom Everitt, and Marcus Hutter. 2017. Count-based exploration in feature space for reinforcement learning. *arXiv preprint arXiv:1706.08090* (2017).

[19] Igor Mordatch and Pieter Abbeel. 2018. Emergence of grounded compositional language in multi-agent populations. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[20] Frans A Oliehoek, Christopher Amato, et al. 2016. *A concise introduction to decentralized POMDPs*. Vol. 1. Springer.

[21] Georg Ostrovski, Marc G Bellemare, Aaron van den Oord, and Rémi Munos. 2017. Count-based exploration with neural density models. *arXiv preprint arXiv:1703.01310* (2017).

[22] Deepak Pathak, Pulkit Agrawal, Alexei A Efros, and Trevor Darrell. 2017. Curiosity-driven exploration by self-supervised prediction. In *International Conference on Machine Learning (ICML)*, Vol. 2017.

[23] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder De Witt, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. 2018. QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning. *arXiv preprint arXiv:1803.11485* (2018).

[24] Mikayel Samvelyan, Tabish Rashid, Christian Schroeder de Witt, Gregory Farquhar, Nantas Nardelli, Tim GJ Rudner, Chia-Man Hung, Philip HS Torr, Jakob Foerster, and Shimon Whiteson. 2019. The starcraft multi-agent challenge. *arXiv preprint arXiv:1902.04043* (2019).

[25] Amanpreet Singh, Tushar Jain, and Sainbayar Sukhbaatar. 2018. Learning when to communicate at scale in multiagent cooperative and competitive tasks. *arXiv preprint arXiv:1812.09755* (2018).

[26] Kyunghwan Son, Daewoo Kim, Wan Ju Kang, David Earl Hostallero, and Yung Yi. 2019. QTRAN: Learning to Factorize with Transformation for Cooperative Multi-Agent Reinforcement Learning. In *ICML 2019 : Thirty-sixth International Conference on Machine Learning*. 5887–5896.

[27] Bradly C Stadie, Sergey Levine, and Pieter Abbeel. 2015. Incentivizing exploration in reinforcement learning with deep predictive models. *arXiv preprint arXiv:1507.00814* (2015).

[28] Alexander L Strehl and Michael L Littman. 2008. An analysis of model-based interval estimation for Markov decision processes. *J. Comput. System Sci.* 74, 8 (2008), 1309–1331.

[29] Sainbayar Sukhbaatar, Rob Fergus, et al. 2016. Learning multiagent communication with backpropagation. In *Advances in neural information processing systems*. 2244–2252.

[30] Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vinicius Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z Leibo, Karl Tuyls, et al. 2017. Value-decomposition networks for cooperative multi-agent learning. *arXiv preprint arXiv:1706.05296* (2017).

[31] Haoran Tang, Rein Houthooft, Davis Foote, Adam Stooke, OpenAI Xi Chen, Yan Duan, John Schulman, Filip DeTurck, and Pieter Abbeel. 2017. # Exploration: A study of count-based exploration for deep reinforcement learning. In *Advances in Neural Information Processing Systems*. 2753–2762.

[32] Rundong Wang, Xu He, Runsheng Yu, Wei Qiu, Bo An, and Zinovi Rabinovich. 2020. Learning Efficient Multi-agent Communication: An Information Bottleneck Approach. In *ICML 2020: 37th International Conference on Machine Learning*.

[33] Tonghan Wang, Jianhao Wang, Chongyi Zheng, and Chongjie Zhang. 2020. Learning Nearly posable Value Functions Via Communication Minimization. In *ICLR 2020 : Eighth International Conference on Learning Representations*.

[34] Yaodong Yang, Ying Wen, Jun Wang, Liheng Chen, Kun Shao, David Mguni, and Weinan Zhang. 2020. Multi-Agent Determinantal Q-Learning. In *ICML 2020: 37th International Conference on Machine Learning*, Vol. 1.

[35] Sai Qian Zhang, Qi Zhang, and Jieyu Lin. 2019. Efficient communication in multi-agent reinforcement learning via variance based control. In *Advances in Neural Information Processing Systems*. 3235–3244.