Emergent Compositional Concept Communication through Mutual Information in Multi-Agent Teams

Extended Abstract

Seth Karten
Carnegie Mellon University
Pittsburgh, PA, USA
skarten@cs.cmu.edu

Siva Kailas
Carnegie Mellon University
Pittsburgh, PA, USA

Katia Sycara
Carnegie Mellon University
Pittsburgh, PA, USA

ABSTRACT

In multi-agent reinforcement learning (MARL) with communication, coordination information (ordinal) is often required in addition to referential info about one's observations. The information bottleneck defines a trade-off between complexity and utility, which loses structure of latent information when compressed solely for utility. Thus, in this work, we use information theory to introduce information-rich, variational compositional communication to adequately embed referential information and to provide a contrastive objective to ground communication in intent-specific features without relying on reward. Each message is composed of a set of emergent concepts, which we show span the observations and intents. Messages are naturally compressed to the least number of bits.

KEYWORDS

Emergent Communication; Multi-Agent Reinforcement Learning; Information Theory; Concept Whitening; Sparse Communication

ACM Reference Format:
Seth Karten, Siva Kailas, and Katia Sycara. 2023. Emergent Compositional Concept Communication through Mutual Information in Multi-Agent Teams: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

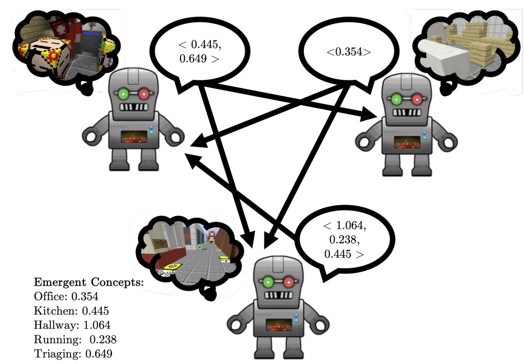

Figure 1: With emergent compositional concept communication, a multi-agent team compresses their observation and intent to communicate learned white-box messages. Here, agents communicate with compositional messages of at most three in length. Each token within the message represents a discrete emergent concept.

1 INTRODUCTION

Emergent communication studies the creation of artificial language. Often phrased as a Lewis game, speakers and listeners learn a set of tokens to communicate complex observations [13]. However, in multi-agent reinforcement learning (MARL), agents suffer from partial observability and non-stationarity (due to unaligned value functions) [17], which aims to be solved with decentralized learning through communication. In the MARL setup, agents, as speakers and listeners, learn a set of tokens to communicate observations, intentions, coordination or other experiences which help facilitate solving tasks [8, 9, 22]. Agents learn to communicate effectively through a backpropagation signal from their task performance [5, 6, 12, 15, 19, 20]. This has been found useful for applications in human-agent teaming [9–11, 16], multi-robot navigation [6], and coordination in complex games such as StarCraft II [18].

Traditionally, in MARL with communication, the communication system is learned in an unsupervised manner from a gradient signal based on the actions taken for the task. However, choosing the correct action requires a sufficient communication protocol, creating non-stationarity. This work aims to ground the communication to more accurately represent the intent through goal-grounded contrastive learning. Contrastive learning [4], which builds on the MaxEnt reinforcement learning objective [3], aims to build current representations which are closer to future states than random states. Information theory objectives have been used in conjunction with contrastive learning to invoke independently principled subspaces [1], or, in our context, concepts. We introduce compositional emergent communication grounded in task-specific information through contrastive learning.

This work enables a compositional emergent communication paradigm, which exhibits clustering and informativeness properties. We show theoretically and through empirical results that compositional language enables independence properties among tokens with respect to referential information. When combined with contrastive learning, our method outperforms competing methods that only ground communication on referential information. Finally, we show that contrastive learning acts as an optimal critic for communication, reducing sample complexity for the unsupervised emergent communication objective. In addition to the more human-like format, compositional communication is able to create variable-length messages, meaning that our method does not generate unnecessarily large messages with little information. We show the utility

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

of our method in multi-agent settings, with a focus on teams of agents and high-dimensional pixel data. Please refer to our evolved paper [7] for full derivations, methodology, and experiments.

2 COMPOSITIONAL COMMUNICATION

In our scenario, the information bottleneck is a trade-off between the complexity of information $I(H^i; M^i)$ (representing the encoded information exactly) and representing the relevant information $I(M^{j\neq i}; Y^i)$, which is signaled from our contrastive objective. In our setup, the relevant information flows from other agents through communication, signaling a combination of the information bottleneck and a Lewis game. We additionally promote complexity through our compositional independence objective, $I(M_1^i; \ldots; M_L^i | H^i)$. This is formulated by the following Lagrangian,

$$\mathcal{L}(\ p(m^i|h^i)\ ) = \beta \hat{I}(M^{j\neq i}; Y^i)\ - \beta \hat{I}(H^i; M^i)\\ - \beta \hat{I}(M_1^i; \ldots; M_L^i | H^i)$$

where the bounds on mutual information $\hat{I}$ are defined in equations 1, 2, and 3. Overall, our objective is,

$$J(\theta) = \max_{\pi} \mathbb{E}\left[\sum_{t\in T}\sum_{i\in N} \gamma_t \mathcal{R}(s_t, a_t) + \mathcal{L}(\ p(m_t|h_t)\ )\right]\\ \text{s.t.}(a_t, m_t, h_t) \sim \pi^i, s_t \sim \mathcal{T}(s_{t-1})$$

Since we want the mutual information to be minimized in our objective, we minimize,

$$\hat{I}(m_1; \ldots; m_L | h) =\\ \mathbb{E}_{h\sim p(h)}\left[D_{KL}\left(q(\hat{m}|h) || \pi_m^i(m_1|h) \otimes \cdots \otimes \pi_m^i(m_L|h)\right)\right] \quad (1)$$

To induce complexity in the compositional messages, we additionally want to minimize the mutual information $I(H; M)$ between the composed message $\hat{m}$ and the encoded information h. For the mutual information between the composed message and encoded information, the following upper bound holds,

$$I(H; M) \leq \hat{I}(H^i, M^i) = \sum_{l}^{L} \mathbb{E}_{h\sim p(h)}\left[D_{KL}\left(q(m_l|h) || z(m_l)\right)\right] \quad (2)$$

First, note that our Markov Network is as follows: $H^j \rightarrow M^j \rightarrow Y^i \leftarrow H^i$. Continue to denote i as the agent identification and j as the agent ID such that $j \neq i$. We aim to satisfy the utility objective of the information bottleneck, $I(M^j; Y^i)$, through a contrastive learning objective,

$$\hat{I}(M^j, Y^i) = \log\left(\sigma(f(s, m, s_f^+))\right) + \log\left(1 - \sigma(f(s, m, s_f^-))\right) \quad (3)$$

which lower bounds the mutual information, $I(M^j, Y^i) \geq \hat{I}(M^j, Y^i)$.

3 EXPERIMENTS AND RESULTS

Our method considers conditioning on inputs, especially rich information, such as pixel data, and task-specific information. When evaluating an artificial language in MARL, we only are interested in referential tasks, in which communication is *required* to complete the task. With regard to intent-grounded communication, we study ordinal tasks, which require coordination information between agents to successfully complete. Thus, we consider tasks with a team of agents to foster messaging with both coordination information and observations. The blind traffic junction environment [19] requires multiple agents to navigate a junction without observing other agents and must coordinate with communication to traverse through the lanes without colliding with agents. We further evaluate the complexity of compositional communication with a referential Pascal VOC [2] game. We evaluate each scenario over 10 seeds against baselines [14, 19, 21].

Our β ablation in table 1 yields a bijection between each token in the vocabulary and the possible emergent concepts, i.e., the enumerated observations and intents. Thus for $\beta = 0.1$, there is no redundancy. Despite a trivially small amount of mutual information between tokens, our compositional method is able to reduce the message size in bits by 2.3x using our derived regularization, for a total of an 8x reduction in message size over non-compositional methods such as `ae-comm`.

Overall, figure 2 shows that our compositional, contrastive method outperforms all methods focused on solely input-oriented communication grounding. In the blind traffic junction, our method yields a higher average task success rate and is able to achieve it with a lower sample complexity. Training with the contrastive update tends to spike to high success but not converge, often many episodes before convergence, which leaves area for training improvement. That is, the contrastive update begins to find aligned latent spaces early in training, but it cannot adapt the methodology quickly enough to converge. The exploratory randomness of most of the early online data prevents exploitation of the high utility f^+ examples. This leaves further room for improvement for an adaptive contrastive loss term.

Table 1: Beta ablation: Redundancy measures the capacity for a bijection between the size of the set of unique tokens and the enumerated observations and intents. Min redundancy is 1.0 (a bijection). Lower is better.

β	Success	Message Size in Bits	Redundancy
0.1	1.0	64	1.0
0.01	.996	69.52	1.06
0.001	.986	121.66	2.06
0	.976	147.96	2.31
non-compositional	.822	512	587

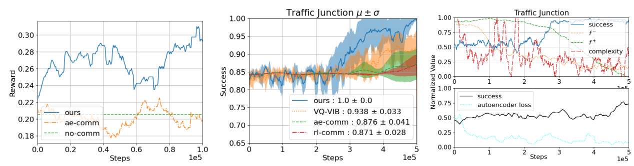

Figure 2: Left: Pascal VOC Game. Middle: Comparison with baselines in Traffic Junction. Right Top: Success, contrastive, and complexity losses for our method. Right Bottom: Success, autoencoder loss for `ae-comm` with supervised pretraining.

ACKNOWLEDGMENTS

This work was supported by AFOSR award FA9550-18-1-0097 DARPA: HR001120C0036 AFRL/AFOSR FA9550-18-1-0251 ARL W911NF-19-2-0146.

REFERENCES

[1] Xili Dai, Shengbang Tong, Mingyang Li, Ziyang Wu, Michael Psenka, Kwan Ho Ryan Chan, Pengyuan Zhai, Yaodong Yu, Xiaojun Yuan, Heung-Yeung Shum, et al. 2022. CTRL: Closed-Loop Transcription to an LDR via Minimaxing Rate Reduction. *Entropy* 24, 4 (2022), 456. 1

[2] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. 2010. The pascal visual object classes (voc) challenge. *International journal of computer vision* 88, 2 (2010), 303–338. 3

[3] Benjamin Eysenbach and Sergey Levine. 2021. Maximum Entropy RL (Provably) Solves Some Robust RL Problems. In *International Conference on Learning Representations*. 1

[4] Benjamin Eysenbach, Tianjun Zhang, Ruslan Salakhutdinov, and Sergey Levine. 2022. Contrastive Learning as Goal-Conditioned Reinforcement Learning. *arXiv preprint arXiv:2206.07568* (2022). 1

[5] Jakob N Foerster, Yannis M Assael, Nando de Freitas, and Shimon Whiteson. 2016. Learning to communicate with Deep multi-agent reinforcement learning. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*. 2145–2153. 1

[6] Benjamin Freed, Rohan James, Guillaume Sartoretti, and Howie Choset. 2020. Sparse Discrete Communication Learning for Multi-Agent Cooperation Through Backpropagation. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 7993–7998. https://doi.org/10.1109/IROS45743.2020.9341079 1

[7] Seth Karten, Siva Kailas, Huao Li, and Katia Sycara. 2023. On the Role of Emergent Communication for Social Learning in Multi-Agent Reinforcement Learning. *arXiv preprint arXiv:2302.14276* (2023). 1

[8] Seth Karten, Mycal Tucker, Siva Kailas, and Katia Sycara. 2022. Towards True Lossless Sparse Communication in Multi-Agent Systems. *Workshop on Deep Reinforcement Learning at Conference on Neural Information Processing Systems (NeurIPS)* (2022). 1

[9] Seth Karten, Mycal Tucker, Huao Li, Siva Kailas, Michael Lewis, and Katia Sycara. 2023. Interpretable Learned Emergent Communication for Human-Agent Teams. *IEEE Transactions on Cognitive and Developmental Systems* (2023), 1–1. https://doi.org/10.1109/TCDS.2023.3236599 1

[10] Brenden M Lake, Tal Linzen, and Marco Baroni. 2019. Human few-shot learning of compositional instructions. *arXiv preprint arXiv:1901.04587* (2019).

[11] Angeliki Lazaridou and Marco Baroni. 2020. Emergent multi-agent communication in the deep learning era. *arXiv preprint arXiv:2006.02419* (2020). 1

[12] Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. 2016. Multi-agent cooperation and the emergence of (natural) language. *arXiv preprint arXiv:1612.07182* (2016). 1

[13] David Lewis. 1969. *Convention*. Harvard University Press, Cambridge, MA. 1

[14] Toru Lin, Jacob Huh, Christopher Stauffer, Ser Nam Lim, and Phillip Isola. 2021. Learning to Ground Multi-Agent Communication with Autoencoders. *Advances in Neural Information Processing Systems* 34 (2021). 3

[15] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. 2017. Multi-agent actor-critic for mixed cooperative-competitive environments. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 6382–6393. 1

[16] Amar R Marathe, Kristin E Schaefer, Arthur W Evans, and Jason S Metcalfe. 2018. Bidirectional communication for effective human-agent teaming. In *International Conference on Virtual, Augmented and Mixed Reality*. Springer, 338–350. 1

[17] Georgios Papoudakis, Filippos Christianos, Arrasy Rahman, and Stefano V Albrecht. 2019. Dealing with non-stationarity in multi-agent deep reinforcement learning. *arXiv preprint arXiv:1906.04737* (2019). 1

[18] 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). 1

[19] Amanpreet Singh, Tushar Jain, and Sainbayar Sukhbaatar. 2018. Learning when to Communicate at Scale in Multiagent Cooperative and Competitive Tasks. In *International Conference on Learning Representations*. 1, 3

[20] Sainbayar Sukhbaatar, Rob Fergus, et al. 2016. Learning multiagent communication with backpropagation. *Advances in neural information processing systems* 29 (2016), 2244–2252. 1

[21] Mycal Tucker, Julie Shah, Roger Levy, and Noga Zaslavsky. 2022. Towards Human-Agent Communication via the Information Bottleneck Principle. *arXiv preprint arXiv:2207.00088* (2022). 3

[22] Changxi Zhu, Mehdi Dastani, and Shihan Wang. 2022. A Survey of Multi-Agent Reinforcement Learning with Communication. *arXiv preprint arXiv:2203.08975* (2022). 1