Intent-Grounded Compositional Communication through Mutual Information in Multi-Agent Teams

Seth Karten and Katia Sycara
skarten@cs.cmu.edu

ABSTRACT

Fundamentally, emergent communication is a representation learning problem. Typically, it is phrased as a Lewis game, in which participants signal using observational information. 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. However, in MARL, the information sent, and the information received defines a different Markov network than defined in the traditional information bottleneck problem. Thus, in this work, we define, study, and show how to approach the unsupervised emergent communication objective, which defines a signaling trade-off between sending referential complexity and ordinal task-specific utility. 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. Through our methodology, each message is independently composed of a set of emergent concepts, which we show span the observations and intents. Additionally, messages are naturally compressed to the least number of bits. We test our novel methodology on referential and ordinal multi-agent tasks.

THEORETICAL DERIVATIONS

- Mutual information measures the dependence between variables in a distribution.

$$I(X;Y) = \mathbb{E}_{p(x,y)}\left[\log \frac{p(x|y)}{p(x)}\right] = \mathbb{E}_{p(x,y)}\left[\log \frac{p(y|x)}{p(y)}\right]$$

- Our method ensures the prediction of each token in a message contains independent information from other tokens.

$$I(m_1;\ldots;m_L|h) \leq \int \ldots \int g_m(*)dhdm_1 \ldots dm_L$$

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

- To induce input-guided complexity, we use the following regularization to produce low-entropy messages.

$$I(H;M) \leq \sum_l^L \int\int p(h)q(m_l|h)\log\frac{q(m_l|h)}{z(m_l)}dm_l\,dh$$

$$= \sum_l^L \mathbb{E}_{h\sim p(h)}\left[D_{KL}\left(q(m_l|h)||z(m_l)\right)\right]$$

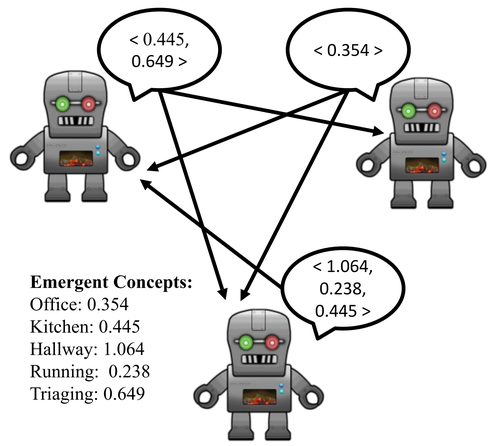
- We show that the following contrastive update optimizes the optimal critic for a goal-conditioned reward function.

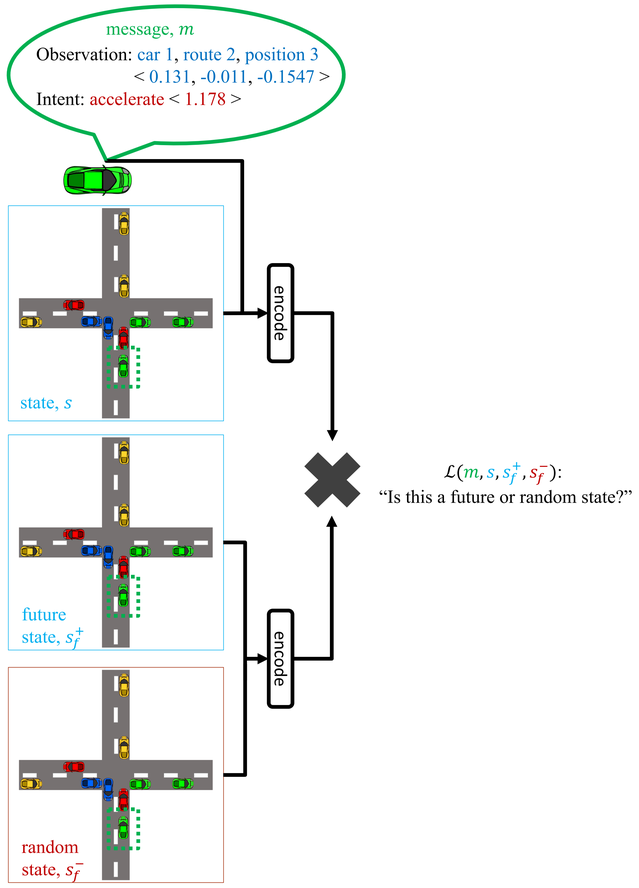
$$I(M^j, Y^i) \leq \log\left(\sigma(f(s,m,s_f^+))\right) + \log\left(1-\sigma(f(s,m,s_f^-))\right)$$

- Since message tokens contain independent information, our method uses the following self-supervised objective to limit sequence lengths with an end of sequence token.

$$H(m_{\text{EOS}}, m_l) = -\pi(m_{\text{EOS}})\log(\pi(m_l))$$

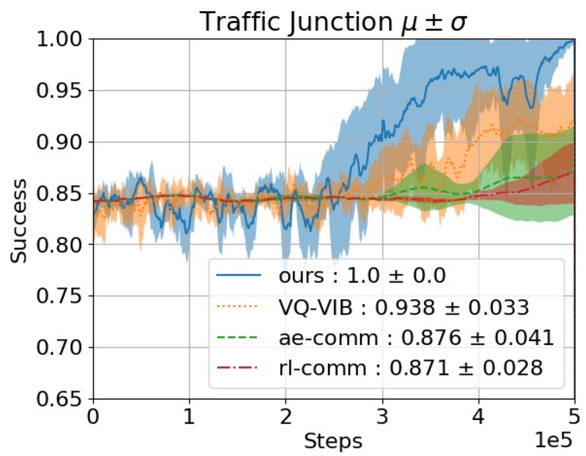
COMPOSITIONAL, CONTRASTIVE COMMUNICATION





EXPERIMENTS AND RESULTS

- Our compositional, contrastive method is able to find converge to 100% task success, leading to better solutions with less data.



- Unlike the autoencoder loss, our regularization objectives converge as the model converges to high performance on the task.

- A β ablation shows that the size of the message is minimized when compared to non-compositional messaging and without our regularization terms. Our method also creates a bijection between observations and intents and the emergent concepts, minimizing redundancy.

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

Algorithm 1 Compositional Message Gen.(h_t)

$T \leftarrow$ num_tokens
$m = \mathbf{0}$ ▷ $T \times d_m$, $d_m \leftarrow$ token_size
$Q \leftarrow$ Q_MLP(h_t)
$V \leftarrow$ V_MLP(h_t)
for $i \leftarrow 1$ to T **do**
  $K \leftarrow$ K_MLP(m)
  $\hat{h} = \text{softmin}(\frac{Q^\top \text{mean}(K,1)}{\sqrt{d_k}})^\top V$
  $m_i \sim \mathcal{N}(\hat{h}; \mu, \sigma)$
end for
return m

RELATED WORK

Baselines:

- *rl-comm* is a baseline communication method learned solely through policy loss.
- *ae-comm* uses an autoencoder to ground communication in input observations.
- *VQ-VIB* uses a variational autoencoder to ground discrete communication in input observations and a mutual information objective to ensure low entropy communication.

DISCUSSION

Contributions:

- Our method learns compressed, discrete concepts (observations and intents) through our derived regularization objectives.
- The compositional nature also naturally reduces the message size and converges to a bijection to the expected vocabulary size.

We also plan to expand into the following future directions:

- Offline RL for contrastive learning
- Conflict resolution with communication

FULL WORKSHOP PAPER