Learning to Communicate with Deep Multi-Agent RL

Jakob N. Foerster*, Yannis M. Assael*, Nando de Freitas, Shimon Whiteson

https://arxiv.org/abs/1605.06676
Motivation

“In the beginning was the Word, and the Word was with God…” [1]

“...as one people speaking the same language they have begun to do this, then nothing they plan to do will be impossible..” [2]

“..debates on these questions have been so fiery that in 1866 the French Academy of Sciences banned publications about the origin of human language.” [3]

[1] John et al., The Bible, Judea, 90-110 A.D.
[2] Yawest et al., Genesis, Judea, 600-500 B.C.
Background and Setting

Cooperative

Partially Observable

Multi Agent RL

DQN

Also: Centralised Learning vs Decentralised Execution
Methods

Differentiable Inter-Agent Learning & Reinforced Inter-Agent Learning

Jakob Foerster
Methods

a) RIAL - RL based Communication

b) DIAL - Differentiable Communication

*DRU*(m) = \begin{cases} \text{Logistic}(\mathcal{N}(m, \sigma)), & \text{if training, else} \\ \mathbb{1}\{m > 0\} \end{cases}
Methods - DRU

Centralised Training

Decentralised Execution

\[
\sigma = 0.5 \quad \sigma = 1.0 \quad \sigma = 2.0
\]

Output $\hat{n}$

Input $m$

\[
\sigma = 0.5
\]

Output $\hat{n}$

Input $m$
Methods - Architecture

Outputs -

Hidden state -

Inputs -

2 layers GRUs for memory

Note: With and without parameter sharing across agents!
Experiments - Switch

Action:
- On
- None
- Tell

Prisoner, in IR

Switch:
- On
- Off

Day 1
- On

Day 2
- On

Day 3
- Off

Day 4
- On

Riddle:
- Multi-Agent: N agents with 1-bit communication channel
- State: N-bit array: has i-th prisoner been to IR
- Action: ‘Tell’ / ‘None’
- Reward: +1 (freedom) / 0 (episode expires) / -1 (all die)
- Observation: ‘None’ OR switch

Freedom!
Experiments - Switch Complexity Analysis

For \( n \) Minions:

\[ 4n(3^T + 1 - 3)/2 \]

For 4 Minions:

\[ 4354288 \]
Experiments - Switch

3 Minions

4 Minions

Jakob Foerster
Experiments - Switch Strategy

3 Minions Strategy
Experiments - MNIST Games

SHELDON COOPER

presents

FUN WITH MNIST
Experiments - MNIST Game

Reward:
- +1 (correct guess)
- 0 (wrong guess)

Communication:
- 1-bit per step ($m$)
Experiments - MNIST Results and Strategy

Results

<table>
<thead>
<tr>
<th>Norm. R (Optimal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIAL</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># Epochs</th>
</tr>
</thead>
<tbody>
<tr>
<td>10k</td>
</tr>
</tbody>
</table>

Strategy

<table>
<thead>
<tr>
<th>True Digit</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>
Experiments - Impact of Noise

![Graph showing the impact of noise on activation probability](image-url)
Future Work
Conclusions

- AI agents can discover communication protocols through Deep RL
- Gradients used in DIAL allow for faster learning
- Parameter sharing can accelerate learning
- Adding noise to the channel => bimodal distribution
- Protocols can be extracted and understood
- A lot of exciting work to be done here

PS: Code will be published online
thank you!
Q&A