On Emergent Communication in Competitive Multi-agent Teams

Paul Pu Liang
with: Jeffrey Chen, Ruslan Salakhutdinov, Louis-Philippe Morency, Satwik Kottur

piang@cs.cmu.edu
@piang279
Grounded Language

Task 1: Single Supporting Fact
Mary went to the bathroom.
John moved to the hallway.
Mary travelled to the office.
Where is Mary? A: office

Task 2: Two Supporting Facts
John is in the playground.
John picked up the football.
Bob went to the kitchen.
Where is the football? A: playground

Task 3: Three Supporting Facts
John picked up the apple.
John went to the office.
John went to the kitchen.
John dropped the apple.
Where was the apple before the kitchen? A: office

Task 4: Two Argument Relations
The office is north of the bedroom.
The bedroom is north of the bathroom.
The kitchen is west of the garden.
What is north of the bedroom? A: office
What is the bedroom north of? A: bathroom

[Weston et al., ICLR 2016]  
[Das et al., CVPR 2018]
Emergent Communication

symbolic data

<table>
<thead>
<tr>
<th>CAT</th>
<th>SOFA</th>
<th>CAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>distractors</td>
<td>target</td>
<td></td>
</tr>
</tbody>
</table>

pixel data

\[22, 2, 0\] the third one

speaker listener

Lazaridou et al., ICLR 2017, Lazaridou et al., ICLR 2018

Bordes et al., ICLR 2017
Task and Talk

(a) Instances

- **shape**
  - triangle
  - square
  - circle
  - star

- **color**
  - blue
  - green
  - red
  - purple

- **style**
  - filled
  - dashed
  - dotted
  - solid

(b) Tasks

- (color, shape)
- (shape, color)
- (style, color)
- (color, style)
- (shape, style)
- (style, shape)

[Q Bot]

[A Bot]

[Kottur et al., EMNLP 2017]
On Emergent Communication in Competitive Multi-agent Teams

Kottur et al., EMNLP 2017
Task and Talk

(a) Instances

(b) Tasks

[Kottur et al., EMNLP 2017]
Task and Talk

(a) Instances

<table>
<thead>
<tr>
<th>shape</th>
<th>color</th>
<th>style</th>
</tr>
</thead>
<tbody>
<tr>
<td>triangle</td>
<td>blue</td>
<td>filled</td>
</tr>
<tr>
<td>square</td>
<td>green</td>
<td>dashed</td>
</tr>
<tr>
<td>circle</td>
<td>red</td>
<td>dotted</td>
</tr>
<tr>
<td>star</td>
<td>purple</td>
<td>solid</td>
</tr>
</tbody>
</table>

(b) Tasks

<table>
<thead>
<tr>
<th>(color, shape)</th>
<th>(shape, color)</th>
<th>(style, color)</th>
<th>(color, style)</th>
<th>(shape, style)</th>
<th>(style, shape)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(color, shape)</td>
<td>(shape, color)</td>
<td>(style, color)</td>
<td>(color, style)</td>
<td>(shape, style)</td>
<td>(style, shape)</td>
</tr>
</tbody>
</table>

Q Bot

Q1: Y

A Bot

A1: 2

Prediction (purple, filled)

Result ✓

[Kottur et al., EMNLP 2017]
Competitive Emergent Communication

Prediction (purple, filled)
Result

Cooperation

Competition
Task, Talk, and Compete

(a) Instances

- Team 1:
  - Q1: Y
  - A1: 2
  - Q1: X
  - A2: 2

Prediction (purple, filled)
Result ✔

(b) Tasks

- Team 2:
  - Q1: Z
  - A1: 2
  - Q1: Y
  - A2: 1

Prediction (green, star)
Result ✗
Sources of Competition

Team 1
(color, style)
Q1: Y
A1: 2
Q1: X
A2: 2
Prediction (purple, filled)
Result ✓

Team 2
(color, shape)
Q1: Z
A1: 2
Q1: Y
A2: 1
Prediction (green, star)
Result ✗

(1) Reward sharing

For fair comparison, the total reward for correct answers is added to Equation 7. This encourages teams to compete and assess their performance relative to other teams. Starting a game and the various sources of competition...

We modify the reward $s_t$, $q_{A^1}(1)$.

This encourages teams to compete and form the agents exactly why they were penalized. They can now use the other agent's dialog overhearing and/or task sharing. Intuitively, this process. Stage 1) Each team is trained to cooperate within a team, in addition to learning how to compete with the other team. Stage 2) The incorrect team observes state $s_t$, $q_{A^1}$, $A_2$, while the incorrect team suffers a penalty $G_{10}$. For wrong ones. When there asymmetry in the results the correct team has a larger bonus $G_{500}$.

The incorrect team observes state $s_t$, $q_{A^1}$, $A_2$, while the incorrect team suffers a penalty $G_{10}$. For wrong ones. When there asymmetry in the results the correct team has a larger bonus $G_{500}$.

...formation of convergence.

In the case of dialogue overhearing and/or task sharing, the team 2. All information shared from team 2 is considered confidential information about the tactics, knowledge, and how to compete with the other team.

Asynchronous Updates: The teams are trained completely independently, then the gradient algorithm ($\hat{G}$) would lead to a data domain mismatch problem (1).

Finally, we fully augment both $s_t$, $q_{A^1}$ and $s_t$, $q_{A^2}$, $A_1$. This encourages teams to compete and assess their performance relative to other teams.

...by which $\hat{G}$ is added to Equation 7, and similarly for $Q_{-2}$. For correct answers $Q_{-1}$ is added to Equation 7. This encourages teams to compete and assess their performance relative to other teams.

...for the team 2. All information shared from team 2 is considered confidential information about the tactics, knowledge, and how to compete with the other team.

Asynchronous Updates: The teams are trained completely independently, then the gradient algorithm ($\hat{G}$) would lead to a data domain mismatch problem (1).

Finally, we fully augment both $s_t$, $q_{A^1}$ and $s_t$, $q_{A^2}$, $A_1$. This encourages teams to compete and assess their performance relative to other teams.
Sources of Competition

**Team 1**

- **Prediction** (purple, filled)
- **Result** ✓

**Team 2**

- **Prediction** (green, star)
- **Result** ✗

---

### 3.1 Metrics

- **Reward sharing**
  - **(1) Reward sharing** ✓
  - **(2) Dialog overhearing** ✗

---

**Task Sharing:**

- **Q1: Y** A1: 2
- **Q1: X** A2: 2

**Result** ✓

**Task, Talk & Compete:**

- **A1: 2** (green, star)
- **A2: 1**

**Prediction** (green, star)

**Result** ✗

---

**Confidential Review Copy. DO NOT DISTRIBUTE.**
Sources of Competition

Team 1

- Prediction (purple, filled)
- Result

Q1: Y
A1: 2
Q1: X
A2: 2

(1) Reward sharing

Q2: Y
A1: 2
Q1: X
A2: 2

(2) Dialog overhearing

Y 2 X 2
Z 2 Y 1

(3) Task sharing

- (color, style)
- (color, shape)

Team 2

- Prediction (green, star)
- Result

Q1: Z
A1: 2
Q1: Y
A2: 1

- (color, shape)
Neural Architecture
Neural Architecture

Q-bot

\[ S_t^Q = [G, q_1, a_1, \ldots, q_{t-1}, a_{t-1}] \rightarrow q_t \in V_Q \]

A-bot

\[ S_t^A = [I, q_1, a_1, \ldots, q_{t-1}, a_{t-1}, q_t] \rightarrow a_t \in V_A \]
Neural Architecture

Trained to maximize expected reward using reinforce algorithm
+R for correct, -10R for incorrect
Reward Sharing

<table>
<thead>
<tr>
<th></th>
<th>Team 2 ✓</th>
<th>Team 2 ✗</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team 1 ✓</td>
<td>(+R, +R)</td>
<td>(+R, −100R)</td>
</tr>
<tr>
<td>Team 1 ✗</td>
<td>(−100R, +R)</td>
<td>(−10R, −10R)</td>
</tr>
</tbody>
</table>

Paul Pu Liang

On Emergent Communication in Competitive Multi-agent Teams
Dialog Overhearing

On Emergent Communication in Competitive Multi-agent Teams
Task Sharing

Team 1 (color, shape) $G^{(1)}$

ListenNet $\rightarrow S_0^Q(1)$

ListenNet $\rightarrow S_t^Q(1)$

ListenNet $\rightarrow q_t(1)$

SpeakNet $\rightarrow I_t(1)$

ListenNet $\rightarrow S_t^A(1)$

ListenNet $\rightarrow a_t(1)$

Rounds of dialog

Team 2 (style, color) $G^{(2)}$

ListenNet $\rightarrow S_0^Q(2)$

ListenNet $\rightarrow S_t^Q(2)$

ListenNet $\rightarrow q_t(2)$

SpeakNet $\rightarrow I_t(2)$

ListenNet $\rightarrow S_t^A(2)$

ListenNet $\rightarrow a_t(2)$

Rounds of dialog

On Emergent Communication in Competitive Multi-agent Teams
Experimental Setup

Baseline: [Kottur et al., EMNLP 2017]
Rewards: (+R, -100R) reward structure
Params: double number of parameters
Double: 2 teams trained independently

Cooperative baselines

![Diagram](diagram.png)

Prediction (purple, filled)
Result

Q1: Y
Q1: X
A1: 2
A2: 2

(color, style)
Experimental Setup

Baseline: [Kottur et al., EMNLP 2017]
Rewards: (+R, -100R) reward structure
Params: double number of parameters
Double: 2 teams trained independently

RS: reward sharing
DO: dialog overhearing
TS: task sharing

Cooperative baselines

Competitive methods

Prediction (purple, filled)
Result ✓
Results

Competition improves generalization (test accuracy)
Faster rates of convergence

DO + TS: 75.8%
Double: 57.8%
Params: 53.3%
Rewards: 49.7%
Baseline: 45.6%

[Kottur et al., EMNLP 2017]
Results: task sharing and dialog overhearing

Sharing messages via overhearing dialog improves performance
Composing sources of competition improves performance

Baseline: 45.6%
Competitive methods
DO: 65.7%
DO + TS: 75.8%
Cooperative baselines
TS: 53.1%
Results: adding reward sharing

Composing sources of competition improves performance

Baseline: 45.6%

Competitive methods

- RS + TS: 68.9%
- RS + DO: 78.3%
- RS + DO+ TS: 77.2%
Measuring information in language

Higher IC scores when trained with competition
Correlated with task performance

Baseline
Cooperative baselines

DO + TS
Competitive methods

0.675

0.806

0.814

0.834

[Jaques et al., ICML 2019]
Conclusion

Code: https://github.com/pliang279/Competitive-Emergent-Communication

pliang@cs.cmu.edu
@pliang279