Multi-agent reinforcement learning: Independent vs. Cooperative Agents

Presented by: Ardi Tampuu

Institute of Computer Science, University of Tartu

May 2015
This article by Ming Tan dates from 1993 and has been cited 813 times.

It’s a classic!
Humans and other living beings are capable to share knowledge and act together to achieve goals. So far (in year 1993) studying interactions was in baby-steps - usually one central controller guiding multiple agents.
This article aims to study if the same number cooperating agents outperform independent agents, given that they:

1. share sensation
2. share episodes
3. share knowledge
The game

The game the agents will try to learn is a hunter-prey game:

Figure 1: A 10 by 10 grid world.

Figure 2: A visual field of depth 2.
Learning Method

The agents learn by Q-learning algorithm and select actions stochastically, not greedily.

\[ p(a_t | s_t) = \frac{e^{Q(s_t, a_t)/T}}{\sum_{a \in \text{actions}} e^{Q(s_t, a)/T}} \]

The main output to pay attention is the time it takes to capture the prey.
Learning is good

First of all, learning improves performance:

Table 1: Average Number of Steps to Capture a Prey: Random vs. Independently Learning Hunters.

<table>
<thead>
<tr>
<th>N-of-prey/N-of-hunters</th>
<th>1/1</th>
<th>1/2</th>
<th>1/2 (joint task)</th>
<th>2/2 (joint task)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random hunters</td>
<td>123.8</td>
<td>50.47</td>
<td>354.45</td>
<td>224.92</td>
</tr>
<tr>
<td>Learning hunters</td>
<td>25.32</td>
<td>12.21</td>
<td>119.17</td>
<td>103.61</td>
</tr>
</tbody>
</table>
Case 1.1: There are two agents, but only 1 can capture prey. The other one is a scout. The hunter always knows the scout’s location and can calculate prey’s location if scout sees prey.

Sounds very logical, but the problem is that number of possible perceptual states increased from 26 to 442 and thus learning could have failed.
Case 1.2: There are two agents, both can capture prey.

Table 3: Two Independent Agents vs. Two Mutual Scouting Agents.

<table>
<thead>
<tr>
<th></th>
<th>Visual Depth</th>
<th>Training</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent agents</td>
<td>2</td>
<td>20.38 (+0.57)</td>
<td>24.04 (+1.60)</td>
</tr>
<tr>
<td>Mutual-scouting agents</td>
<td>2</td>
<td>25.20 (+0.79) (worse)</td>
<td>24.52 (+1.24) (same)</td>
</tr>
<tr>
<td>Independent agents</td>
<td>3</td>
<td>14.65 (+0.53)</td>
<td>16.04 (+0.56)</td>
</tr>
<tr>
<td>Mutual-scouting agents</td>
<td>3</td>
<td>14.02 (+0.75) (same)</td>
<td>12.98 (+0.65) (better)</td>
</tr>
<tr>
<td>Independent agents</td>
<td>4</td>
<td>12.21 (+0.65)</td>
<td>11.53 (+0.61)</td>
</tr>
<tr>
<td>Mutual-scouting agents</td>
<td>4</td>
<td>11.05 (+0.56) (better)</td>
<td>8.83 (+0.78) (better)</td>
</tr>
</tbody>
</table>

If the information is too low quality, state space increases and learning slows down more than the benefits are worth.
Sharing is policies is good

Case 2.1: There are two agents, both can capture prey. They do not share perceptual info, but they share their policies.

Telling each other what we have learned is good - more likely to stumble upon useful situations.
Case 2.2: There are two agents, both can capture prey. When one captures prey, he sends his whole solution path to the other (very primitive and biased, but useful).

Having more experience is good - more likely to stumble upon useful situations.
Cooperating is good

Case 3: There are two agents, both have to be close to the prey to capture it.

Figure 7: Typical runs for the 2-prey/2-hunter joint task.

Figure 8: Typical runs for the 1-prey/2-hunter joint task.
Thank you for your attention!