LEARNING HUMAN-ROBOT TASK WITH REINFORCEMENT LEARNING

Xuan-Anh Nguyen and Hsien-I Lin
Graduate Institute of Automation Technology
National Taipei University of Technology
2017/11/5
Outline

- Introduction
- Problem Statement
- HRC with Human Reward Assignment
- Experiment
- Conclusions
Introduction

Source: http://www.youtube.com
Problem Statement

• **Advantages:**
  - More flexible manufacturing practices
  - Eliminate the safety cage
  - Increase performance

• **Challenges:**
  - Safety
  - High level cognitive skills
  - Mutual adaptation
Reinforcement Learning

- Set of states: $S$
- Set of actions: $A$
- Reward function $R$
Reinforcement Learning

- The agent task: to find an optimal policy, mapping states to actions, that maximize long-run measure of the reinforcement.

maximize:

\[ V^\pi(s) \equiv E[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \ldots] = E[\sum_{i=0}^{\infty} \gamma^i r_{t+i}] \]

\[ \pi^*(s) = \arg \max_a \left[ r(s, a) + \gamma V^\star(\delta(s, a)) \right] \]
Multi-agent RL

• A human-robot system => Multi-agent system

One robot + One human

Robot action set aR

Human action set aH

System action: \( a = [aR, aH] \)
Human reward assignment

- Benefits:
  - End users can specify correct behavior
  - Without requiring programming skill

Source: http://www.ScienceNordic.com
Human reward assignment

- Compared to learning from human demonstration:
  - Easier to give a feedback than make demonstrations.
Human reward assignment

- **Environment**
- **Robot**
  - Action
  - State, Reward

- **Human**
  - Feedback
  - Reward

- **Environment**
  - Action
  - State

- **Credit assignment**
  - Reward

- **Robot**
  - Action
  - State
Human reward assignment

\[ t = 0 \]

All \( s \in S, a^1 \in A^1, a^2 \in A^2 \)

\( Q^1_t(s, a^1, a^2) = 1, Q^2_t(s, a^1, a^2) = 1 \)

initialize \( s_0 \)

Choose action \( a^1_t \) based on \( \pi^1(s_t) \)

Observe \( r, a^2_t \) and \( s_{t+1} \)

\[
Q^1_{t+1}(s, a^1, a^2) = (1 - \alpha_t) Q^1_t(s, a^1, a^2) + \alpha_t \left[ r_t + \beta \pi^1(s_{t+1}) Q^1_t(s_{t+1}) \pi^2(s_{t+1}) \right]
\]

\[
Q^2_{t+1}(s, a^1, a^2) = (1 - \alpha_t) Q^2_t(s, a^1, a^2) + \alpha_t \left[ r_t + \beta \pi^1(s_{t+1}) Q^2_t(s_{t+1}) \pi^2(s_{t+1}) \right]
\]

\[ t = t + 1 \]
# Human reward assignment

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Set <code>current_state = START_STATE</code></td>
</tr>
<tr>
<td>2.</td>
<td>Loop until <code>current_state == GOAL_STATE</code></td>
</tr>
<tr>
<td>3.</td>
<td>Execute robot action <code>ar</code> according to current policy <code>π</code></td>
</tr>
<tr>
<td>4.</td>
<td>Observe human action <code>ah</code></td>
</tr>
<tr>
<td>5.</td>
<td>Set <code>next_state</code> to the resultant state</td>
</tr>
<tr>
<td>6.</td>
<td>Human enters reward <code>r</code></td>
</tr>
<tr>
<td>7.</td>
<td>Update robot policy <code>π</code> using <code>Q(λ)</code></td>
</tr>
<tr>
<td>8.</td>
<td>Record <code>(current_state, ar, next_state)</code></td>
</tr>
<tr>
<td>9.</td>
<td><code>Current_state = next_state</code></td>
</tr>
</tbody>
</table>
Experiment

A  B  C

https://roboticsandautomationnews.com
Experiment
Experiment

Training phase
# Experiment

<table>
<thead>
<tr>
<th>Subjective measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1: “YuMi was helpful in accomplishing the task”</td>
</tr>
<tr>
<td>Q2: “The collaboration is natural”</td>
</tr>
<tr>
<td>Q3: “YuMi and I work efficiently together”</td>
</tr>
<tr>
<td>Q4: “YuMi and I work fluently together”</td>
</tr>
<tr>
<td>Q5: “YuMi and I contributed equally to the completion of the task”</td>
</tr>
<tr>
<td>Q6: “YuMi was able to keep track of the task progress”</td>
</tr>
<tr>
<td>Q7: “YuMi is intelligent”</td>
</tr>
<tr>
<td>Q8: “YuMi perceived accurately what my goals are”</td>
</tr>
<tr>
<td>Q9: “Our team performance improved over time”</td>
</tr>
<tr>
<td>Q10: “YuMi was committed to the task”</td>
</tr>
</tbody>
</table>
Conclusions

- We introduced a method for human-robot system that leverages human cognitive ability.
- A pick-and-place human-robot task was experimented.
- Our method allows a human to quickly train a robot to follow his intention and collaborative efficiently with him in a joint task.
Thank you for your listening!