Impact of Interaction Design on Human Satisfaction Teaching Reinforcement Learning Agents in Partially Observable Domains

Divya Srivastava, Spencer Frazier, Karen M Feigh
Introduction

What is Interactive Machine Learning?
• Humans play an active role during the learning process of machine learning agents
• Proven to converge quicker on desired behavior than traditional ML techniques with no human interaction

What are the GOALS of IML?
• We want nonexperts to be able to interact with and train agents easily

How do we do this?
• Design with human experience in mind!
• Design ML agents that are
  • more intuitive for people to understand
  • easily customizable
  • robust
Research Questions

- How do different methods of interaction between a teacher and a RL agent affect the teacher’s experience?
- How can we design interaction algorithms that foster a positive experience for non-expert teachers when teaching RL agents?
Background

Four Variations of Interaction Algorithm

<table>
<thead>
<tr>
<th>1-Step:</th>
<th>5-Step:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advice is followed for one</td>
<td>Advice is followed for 5</td>
</tr>
<tr>
<td>time step in the direction</td>
<td>time steps in the direction</td>
</tr>
<tr>
<td>provided by teacher.</td>
<td>provided by teacher.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time Delay:</th>
<th>Probabilistic:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advice is followed for one</td>
<td>Agent chooses whether to</td>
</tr>
<tr>
<td>time step in the direction</td>
<td>follow advice based on a</td>
</tr>
<tr>
<td>provided by teacher after a</td>
<td>probability.</td>
</tr>
<tr>
<td>2 second delay.</td>
<td></td>
</tr>
</tbody>
</table>

Factors That Influence Human Teacher’s Experience

- Perceived Intelligence
- Transparency
- Frustration
  - Compliance with Input
  - Robustness
  - Complexity
  - Powerlessness
  - Flexibility
  - Probabilistic
  - Effort
  - Accuracy of ASR
Research Questions

- How do different methods of interaction between a teacher and a RL agent affect the teacher’s experience?
  - How can we design interaction algorithms that foster a positive experience for non-expert teachers when teaching RL agents?
- How does this experience change based on the observability of the domain?
Task/Domain

**Task**

Help the agent navigate through the maze and approach a non-playable character

**Study 1:**
- No penalties in the maze

**Study 2:**
- Stationary penalties ("water") in the maze
- Task failed and agent receives negative reward if agent runs into water

Teacher’s View of Agent in Maze

Agent’s FPV
24 & 30 participants with no prior ML experience for each respective study

Order of agents trained was balanced
Metrics

Objective ML Measures
• Number of steps to complete episode
• Number of advice given by teacher per episode
• Cumulative reward per episode
• Training time per episode

Human Experience Measures
• Frustration
• Transparency
• Immediacy
• Perceived Performance
• Perceived Intelligence
Results – ML Metrics

As the domain becomes more complex, the number of steps it takes to complete an episode is roughly the same for the 1-Step and Time Delay variation.

Introducing probability results in an increasing number of steps per episode.
Results – ML Metrics

The advice given by the teacher will vary depending on the domain.
Results – ML Metrics

- As the domain becomes more complex, the cumulative reward decreases across all agents.
- The 1-Step method continuously accumulates the highest reward.

Cumulative Reward per Episode

Study 0: Fully Observable with Penalties
Study 1: Partially Observable
Study 2: Partially Observable with Penalties
Results – ML Metrics

- The 1-Step method consistently had the best training time* followed by the 5-Step method

*in the partially observable domain, the Time Delay agent took the least amount of time to train per episode. This may be because the amount of advice given to the Time Delay agent was lower than the amount given to the 1-Step agent, and thus the agent fell back on its action-selection process and equated the 1-Step agent in training time
• The Probabilistic method consistently rated worse than other interaction methods
• The Time Delay interaction rates very similarly to the 1-Step and 5-Step interaction methods in partially observable domains
• Frustration levels with the agents were lower for the partially observable domains, but much higher overall once penalties introduced.
A closer look at Perceived Intelligence

- Once penalties are introduced into the partially observable environment, the Time Delay method is considered equally intelligent to the 1-Step method.
- The 1-Step interaction method was considered the most intelligent in the domain without penalties.
- The 5-Step method is considered one of, if not the most, intelligent agent in the domains with penalties.

Above: average rating of perceived intelligence after training each agent

Above: Rankings of perceived most intelligent agent after training all four agents
## Features

### Study 0: Fully Observable
### Study 1: Partially Observable
### Study 2: Partially Observable with Penalties

<table>
<thead>
<tr>
<th>Feature</th>
<th>Study 0</th>
<th>Study 1</th>
<th>Study 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frustration</td>
<td>Intelligence</td>
<td>Frustration</td>
</tr>
<tr>
<td>Compliance</td>
<td>71</td>
<td>54</td>
<td>67</td>
</tr>
<tr>
<td>Effort</td>
<td>9</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>Frustration</td>
<td>0</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>Immediacy</td>
<td>38</td>
<td>54</td>
<td>33</td>
</tr>
<tr>
<td>Improvement Through Time</td>
<td>0</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>Less Instruction</td>
<td>0</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>Memory</td>
<td>0</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Randomness</td>
<td>21</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>Repeating Myself</td>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Transparency</td>
<td>4</td>
<td>0</td>
<td>33</td>
</tr>
</tbody>
</table>

Above: Percentage of participants who mentioned in their long responses certain features that contributed to the user experience.

- Compliance, Immediacy, Randomness, and Transparency found to be the four most mentioned features that impact human experience across all studies.
Conclusions

In partially observable domains, it appears to be both allowable and even recommended to:

- minimize unpredictability in the agent’s behavior
- to incorporate a short time delay to give the teacher some time to make a decision regarding the agent’s next action.

The decision of generalizing movement through time is left to the algorithm designer since the user perception and objective performance of these interaction methods vary in partially observable domains.
Thank you for listening!

Contact Information:

Divya Srivastava, divya.srivastava@gatech.edu
Spencer Frazier, sfrazier7@gatech.edu
Karen M Feigh, karen.feigh@gatech.edu