Real-Time Policy Learning with User Feedback for British Sign Language (BSL) Users in an Example of Navigation-Based Task

Boris Mocialov, Patricia A. Vargas, Graham H. Turner {mocialov, p.a.vargas, g.h.turner} @hw.ac.uk
Robotics Lab at School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh Centre for Robotics

Overview

Motivation:
- User-guided learning accelerates policy-learning process
- Individual adaptation increases operational longevity
- Specific target group with specific requirements

Challenges:
- Choice of communication medium and appropriate perception
- Finding tasks that the target group would benefit from

Applications:
- Specialised equipment control
- Wide-use of natural multi-modal communication

Problem Statement

Knowing a set of response primitives and not performing explicit classifications:
Can we, using known primitives, (re-)learn to respond to the BSL in a way that suits the user?

Approach

Prototype 1: Confidence Matrix with Mutual Exclusive Actions [2]

<table>
<thead>
<tr>
<th>Iterations</th>
<th>User</th>
<th>Agent</th>
<th>User Feedback</th>
<th>Action 1</th>
<th>Action 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iteration 1</td>
<td>Sign 1</td>
<td>Action 2</td>
<td>Wrong</td>
<td>Sign 1</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sign 2</td>
<td>-0.1</td>
</tr>
<tr>
<td>Iteration 2</td>
<td>Sign 2</td>
<td>Action 1</td>
<td>Wrong</td>
<td>Sign 1</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sign 2</td>
<td>-0.1</td>
</tr>
<tr>
<td>Iteration 3</td>
<td>Sign 1</td>
<td>Action 1</td>
<td>Correct</td>
<td>Sign 1</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sign 2</td>
<td>-0.2</td>
</tr>
</tbody>
</table>

Result:
- Slower convergence (or failure to converge) with more classes and introduced noise

Limitations:
- Mutually exclusive actions
- Static environment with no concept drift
- Noise-less perception

Example Task

Aim:
- Agent must learn associations between BSL commands and agent’s primitive actions in 2D navigation-based task

Evaluation:
- Similarity between generated trajectory and desired trajectory with respect to:
  - Order of visited landmarks (L1, L2, etc.)
  - Side we pass landmarks on (West, East, North, South, etc.)

Future Work

1. Study methods for real-time policy learning in detail
2. Modify existing method or create new one to suit the problem
3. Overcome limitations by applying the new method
4. Design a set of tasks that suits the target group
5. Run experiments with the target group using designed tasks

Acknowledgements

This research has been supported by the Heriot-Watt University, School of Engineering & Physical Sciences James Watt Scholarship through Edinburgh Centre for Robotics and the Engineering and Physical Sciences Research Council (EPSRC) as part of the CDT in Robotics and Autonomous Systems at Heriot-Watt University and The University of Edinburgh.

Contact Details
Email: bm4@hw.ac.uk
Web: http://www.edinburgh-robotics.org/students/boris-mocialov
Twitter, Instagram, Facebook, LinkedIn: @mocialov
Phone: +44(0)7821178255