Intent Recognition in a Generalized Framework for Collaboration

Rupam Bhattacharyya*, Shyamanta M. Hazarika

Abstract

For a collaborative assistive device, human intent recognition (IR) is one of the first and foremost requirements. Formalizing the complex process of human IR in a compact yet expressive mathematical model holds promise. We put forward a Hierarchical Finite State Machine (H-FSM) for human IR within a generalized framework for collaborative assistive devices. Visual and contextual observations drive the H-FSM to different levels of granularity.

1. Introduction

The design of next generation assistive robots has to incorporate models of human rationale and cognitive responses. Human intent recognition (IR) is one of the first and foremost requirements. IR is "the process of becoming aware of the intentions of other agents, inferring them through observed actions or effects on the environment". Ability to integrate predicted effects of self and other's actions is key to successful collaboration. The mental model and model based on theory of mind (ToM) with similarity in data representation and computational processes involved holds promise. This can be viewed as shared representation which encodes facts like: 1. relative position between assistive robot and human(s); 2. responsiveness of the co-workers; and 3. status of various shared sub-tasks.

Human-human interaction can trigger successful cooperation in case of similar partners; this seems to re-establish the famous "chameleon effect". Collaborative assistive robot must mimic the actions of human. Mirror neurons can be utilized as a benchmarking testing suite for the collaborative assistive robot. Our framework is intended to give a structured approach for designing such a robotic assistant.

We put forward a Hierarchical Finite State Machine (H-FSM) for human IR within a generalized framework. The hierarchical nature inherent in IR is exploited through the H-FSM. Unpredictability of human actions in various situations is our motivation. While performing an action with a particular intent, human can...
readily switch to a completely different sequence reflecting another intent. Assistive robots need to be aware of such frequent changes so that it can update the shared goal. Following the principle of ToM, robotic assistant incorporates what the human is thinking in modeling its own action.

2. The Generalized Framework

A generalized framework for collaborative assistive devices is shown in Figure 1. The framework follows the underlying principle of mental model and ToM. Most of human-robot collaborative applications represent dynamical systems. Our framework is termed 'generalized' as: a. It describes the overall concept to design a collaborative assistive device. b. It requires efforts from various fields (for its components) as depicted in Table 1.

![Fig. 1. Generalized framework and its components](image)

<table>
<thead>
<tr>
<th>Research Field</th>
<th>Concept involved</th>
<th>Corresponding component in the framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychology</td>
<td>Theory of Mind (ToM), Mental model and Chameleon effect</td>
<td>Computational model for Intent Recognition</td>
</tr>
<tr>
<td>Neuroscience and Signal Processing</td>
<td>Mirror neurons</td>
<td>Testing</td>
</tr>
<tr>
<td>Computer Science (cognitive robotics and cognitive computer vision)</td>
<td>Mathematical modeling of the concepts involved in Theory of Mind (ToM) and Mental model.</td>
<td>Computational model for Intent Recognition</td>
</tr>
<tr>
<td></td>
<td>Developing therapeutic control algorithms.</td>
<td>Robotic agent (controller)</td>
</tr>
<tr>
<td></td>
<td>Mimicking the human actions to incorporate effects of “chameleon effect” in the collaborative assistive device</td>
<td>Computational model for Intent Recognition</td>
</tr>
</tbody>
</table>

**Hierarchical FSM**

The single hierarchical FSM can represent and simulate entire human IR scenario in a particular domain like navigating in an office. The classical H-FSM is modified so that the H-FSM fits into human IR scenario without the complex notations of entry/exit nodes, boxes or super nodes and the labeling function. Our H-FSM stays in between the classical FSM and classical H-FSM in terms of complexity. Without changing meaning of symbols and addition of new symbols of classical FSM, we can effectively model a complex dynamic system through our H-FSM. We consider the environment to be fully observable and thus the agent equipped with vision based capability can identify which state it is in.
The single hierarchical FSM can represent and simulate entire human IR scenario in a particular domain like navigating in an office. The classical H-FSM is modified so that the H-FSM fits into human IR scenario without the complex notations of entry/exit nodes, boxes or super nodes and the labeling function. Our H-FSM stays in between the classical FSM and classical H-FSM in terms of complexity. Without changing meaning of symbols and addition of new symbols of classical FSM, we can effectively model a complex dynamic system through our H-FSM. We consider the environment to be fully observable and thus the agent equipped with vision based capability can identify which state it is in.

Example scenario. Consider part of common household as shown in Figure 3; a patient armed with the wheelchair is to navigate through the indoor. The high level goals will be to move to different locations. Considering only implicitly communicated intention action performed by the patient and scene change information is the only input. Figure 4 is the hand simulation of the example scenario. Shaded state / states enclosed within dotted rectangle represent possibility of further expansion. Sequence (S1–S2–S3–S4) represents M or Level-0 FSM; whereas (S1–S11–S12) represents “zooming in” M Level-1 FSM.

Case: 1 Human in state S1 executes the transition (S1, f) and (S11, f). Human reaches the state where coffee table and parts of sofa is visible to both human and robot. Robot updates its shared goal to be “Coffee table” based on the recognized intent from the “white board”. Shaded state can spawn other levels of FSM (not shown in the figure). S11 is an ordinary state in the sense that the chosen action by the human (in our scenario) prevents S11 to unfold itself. Shaded states from S11 represent another two FSMs, M (Level-2) that can be further “zoomed-in” if the actions labeled in the transition edge are performed by the human.

Case: 2 Human in state S1 executes the transition (S1, r), (S2, f), (S3, l) and (S4, w). Another person is in front of the human (in the wheelchair); recognized intent is as in Table 2 (Intent is updated in the “white board”). The states S2 and S1 are ordinary states in the sense that the chosen action by the human (in our scenario) prevents S2 and S1 to unfold. The shaded states from S2 and S1 represent another two FSMs, M (Level-1) that can be “zoomed-in” if the corresponding actions labeled in the transition edge are performed by the human.
### Table 2. Example Scenario

<table>
<thead>
<tr>
<th>Current state</th>
<th>Scene information (known objects)</th>
<th>Allowed action set</th>
<th>Action taken</th>
<th>Scene change information</th>
<th>Recognized intent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Living room ($S_1$)</td>
<td>Coffee table, sofa, side table.</td>
<td>left(l), right(r), front(f), back(b), wait(w)</td>
<td>front(f)</td>
<td>Coffee table and parts of sofa (no other objects)</td>
<td>human is going towards the Coffee table</td>
</tr>
<tr>
<td>Hallway ($S_4$)</td>
<td>Person (he/she is approaching the wheelchair)</td>
<td>left(l), right(r), back(b), wait(w)</td>
<td>wait(w)</td>
<td>Person is still in front of the wheelchair</td>
<td>human want to communicate with him/her</td>
</tr>
</tbody>
</table>

4. Conclusion and future work

Our framework would encourage others to use emergent co-ordination and implicit communication. Main challenges include: a. How to model newly learned intentions? b. How to enable mimicry (by the robotic agent) in a shared human-robot workspace? c. Analysis of cycle detection and reachability issues. d. How to prioritize actions? e. Augmentation of system states with domain specific information. Addressing above challenges and thereafter evaluation of the H-FSM within a perceptual framework would be a definitive step for intent recognition in a generalized framework for collaboration. This is part of on-going research.

Acknowledgements

Financial support from MHRD as Centre of Excellence on Machine Learning Research and Big Data Analysis is gratefully acknowledged. Many thanks to Zubin Bhuyan for insightful discussions of the Theory of Mind.

References