Socially-Sensitive Technologies for HRI: Challenges and Perspectives

Elisabeth André
Universität Augsburg
andre@hcm-lab.de

Abstract
Starting the recent years, a significant amount of effort has been dedicated to explore the potential of social signal processing in human interaction with embodied conversational agents and social robots. While there is a proliferation of studies that investigate specific aspects of embodied social interaction under laboratory conditions, less attention has been paid to the design and realization of naturalistic social settings in which artificial agents autonomously interact with human users. To bring social agents to the people’s daily environment, user-agent communication should be properly situated in the context of the application at hand rather than isolated as a laboratory experiment. The objective of the paper is to identify relevant knowledge and reasoning capabilities to enhance the social sensitivity of artificial agents.

1 Towards More Naturalistic Social Settings

Societal challenges, such as an ageing population, have created the need for a new generation of robots that are able to smoothly interact with people in their daily environment. Technologies for realizing individual components of a social agent have reached a great level of advancement. However, due to the complexity of social behaviors that have to be simulated by artificial agents, there is a big gap between the vision of an artificial agent with human-like social skills and currently available implementations of it.

A significant amount of effort is still required to move from pre-scripted short-term interaction to interactive scenarios where a robot has to engage in a conversation with a human over an extended period of time. [Rich and Sidner, 2010] coined the term always-on relational agents to describe the vision of an artificial agent with human-like social skills and currently available implementations of it.

2 Towards a Deeper Understanding of Social Cues

Social signal analysis is known to be a very hard problem and a real bottleneck in social human-agent interaction. To enhance an agent’s sensitivity, a deeper understanding of the user’s situation is required. To illustrate this, we present a number of examples:

- Integration of Social Cue Analysis with Semantic and Pragmatic Analysis

  Work done in the Semaine project [Schröder et al., 2012] has shown that simple backchannel signals, such as “I see”, may suffice to create the illusion of a sensitive listener. However, to engage humans over a longer period of time, a deeper understanding of the dialogue would be necessary. While a significant amount of work has been done on the semantic/pragmatic processing in the area of Natural Language Processing, work that accounts for a close interaction between the communication streams required for semantic/pragmatic processing and social signal processing is rare. The integration of social signal processing with semantic and pragmatic analysis may help resolve ambiguities. Especially short utterances tend to be highly ambiguous when solely the linguistic data is considered. An utterance like “right” may be interpreted as a confirmation as well as a rejection, if intended cynically, and so may the absence of an utterance. Preliminary studies have shown that the consideration of
social cues may help improve the robustness of semantic and pragmatic analysis, see [Bosma and André, 2004].

- **Incorporating an Affective Theory of Mind**
  Interpreting emotional signs separated from any social context information is extremely hard if not impossible. For example, a smile is not always a sign of happiness. People also tend to smile when feeling embarrassment [Keltner, 1995]. Furthermore, how emotions are perceived depends on the social relationship between the interlocutors [Hareli and Hess, 2012]. For example, a person may interpret a smile of a competitor rather as gloating than happiness. Many recognition systems nowadays are not able to take these subtle differences into account. Rather they would map a smile onto the emotional state happiness. The problem is that they do not explicitly represent the causes of detected affect. To enable a deeper understanding of the user’s emotions, a robot should be equipped with a Theory of Mind (ToM) capability that will enable it to assess a situation not only from its own, but also from the user’s perspective. To this end, the MARSSI model has been developed that relates appraisal rules and emotion regulation rules to social cues shown by a user [Gebhard et al., 2018].

- **Modeling the Dynamics of Interpersonal Cues**
  In human-human conversation, the interlocutors dynamically coordinate and adjust their verbal and nonverbal behaviors to each other in order to demonstrate engagement in the conversation. To determine the level of engagement of a human in a dialogue with a robot, it does not suffice to analyze the individual behavior patterns of the human. Rather the dynamics of bidirectional behavior patterns have to be taken into account. Examples of bidirectional behavior patterns include the establishment of shared attention, the regulation of the dialogue flow by turn taking and the generation of backchannel feedback [Mehlmann et al., 2014]. Depending on the situative context, such behavior patterns may vary considerably and the social signals need to be interpreted accordingly.

- **Finding the Right Level of Sensitivity in Human-Robot Interaction**
  Further improvements of techniques for the analysis of social signals might lead to agents that respond to human signals in an oversensitive manner [André, 2013]. Agents that show a reaction to any social signal will most likely irritate users. Furthermore, their behavior might confuse users because the adaption was based on social signals the users were not aware of. Obviously, not every social signal cue from the user should trigger a response from the agent. The problem of deciding which user behavior should be interpreted as system input is called the “Midas Touch Problem”. Hoekstra and colleagues [Hoekstra et al., 2007] present a number of strategies to mitigate the “Midas Touch Problem” for an application with two agents that adapt their presentations to the user’s level of attentiveness. In their work, eye gaze was the only user cue that was interpreted by the agents. Thus, the question arises of how to determine the right level of sensitivity for a multitude of social signals in interactive conversational settings.

- **Socially-Aware Learning for Robots**
  Humans adapt their social behaviors during interactions based on explicit or implicit cues they receive from the interlocutor. In order to establish longer lasting relationships between artificial companions and human users, artificial companions need to be able to adjust their behavior on the basis previous interactions. That is, they should remember previous interactions and learn from them [Aylett et al., 2011]. To this end, sophisticated mechanisms for the simulation of self-regulatory social behaviors will be required. An example of a socially-aware robot that adapts its linguistic style dynamically to the user’s preferences based on implicitly provided feedback has been presented by [Ritschel et al., 2017]. However, a significant amount of work still needs to be done to take advantage of implicitly and explicitly provided feedback in order to overcome the noise of continuously provided social signals and the sparsity of occasionally provided verbal feedback to a robot.

References


