Towards Affect-Awareness for Social Robots

Samuel Spaulding and **Cynthia Breazeal**
Media Lab
Massachusetts Institute of Technology
Cambridge, MA

Abstract

Recent research has demonstrated that emotion plays a key role in human decision making. Across a wide range of disciplines, old concepts, such as the classical "rational actor" model, have fallen out of favor in place of more nuanced models (e.g., the frameworks of behavioral economics and emotional intelligence) that acknowledge the role of emotions in analyzing human actions. We now know that context, framing, and emotional and physiological state can all drastically influence decision making in humans. Emotions serve an essential, though often overlooked, role in our lives, thoughts, and decisions. However, it is not clear how and to what extent emotions should impact the design of artificial agents, such as social robots.

In this paper I argue that enabling robots, especially those intended to interact with humans, to sense and model emotions will improve their performance across a wide variety of human-interaction applications. I outline two broad research topics (affective inference and learning from affect) towards which progress can be made to enable "affect-aware" robots and give a few examples of applications in which robots with these capabilities may outperform their non-affective counterparts. By identifying these important problems, both necessary for fully affect-aware social robots, I hope to clarify terminology, assess the current research landscape, and provide goalposts for future research.

Introduction

Artificial Intelligence, as a field, is the science and engineering of replicating intelligent behavior. Naturally, this leads to the follow up question: "What is intelligent behavior?" Over the past decades, this definition has changed substantially, as artificial intelligence researchers have realized that 'intelligence' encompasses a range of abilities both broader and deeper than originally anticipated. Over time, the field has shifted away from its original emphasis on symbolic manipulation and logic. In its place, the modern approach to AI has emphasized pattern recognition, embodied cognition, and general-purpose learning from data.

In parallel, researchers studying *human* intelligence have come to recognize that earlier emphasis on spatial reasoning and language capabilities are only part of the picture. Emotional Intelligence, its impact, and its measurement are now recognized and studied as critical components of intelligence in their own right. Emotions, it turns out, are inextricably linked to our bodies, our thoughts, and our core experience of being human. Moreover, it appears that, far from being an irrational aberration, the influence of emotion on decision making may in fact be an *advantageous adaptation*. Medical history is replete with accounts of patients who have suffered lesions to areas of the brain associated with emotional processing. These patients, while perfectly capable *physically*, suffer from crippling indecision over even very simple tasks and decisions(Bechara, Damasio, and Damasio 2000).

Science fiction writers have long hypothesized what emotionally sensitive robots might do or feel (whether they do, in fact, 'feel' is a separate topic), but until the end of the 20th century, such topics remained purely speculative. In recent years, however, major advances in affective technology, human-robot interaction, and computational cognitive modeling have provided researchers with the tools to transform the dream of affect-aware robots into reality.

The field of Affective Computing (Picard 2000) was born less than 20 years ago and has grown into a highly active research topic with impact in many other research areas. Major advances in unintrusive affective sensing technology, such as algorithms for facial expression analysis (El Kaliouby and Robinson 2005), wristbands that sense electrodermal activity, and pressure sensitive chairs and computer peripherals have enabled the collection of large quantities of real-time affective data without significantly disrupting interaction with technology.

Within HRI, researchers have begun to acknowledge the roles of emotion in social interaction and build towards enabling social robots that can sense and act on affective and emotional information. (Paiva, Leite, and Ribeiro 2014). This work is still exploratory, and, thus far, there does not yet exist a widely-accepted emotional/affective framework within which researchers studying affect-aware robots work. This paper sketches two abilities that affect-aware robots will have to possess, as well as some associated research challenges.

Robots capable of affective inference

At the most basic level, an affect-aware robot should be capable of *affective inference*: it should be able to interpret signals of human affect and draw some inferences from them

about the state of the world or the mental state of a human. For example, a robot capable of affective inference working in the domain of educational tutoring might analyze facial expressions looking for signs of frustration, confusion, or delight. From its affective sensing, the robot might then infer what educational concepts a child has mastered, based on the questions answered, in order to track his/her curricular progress more accurately (Spaulding 2015).

Affective inference is, importantly, not the same as *affect recognition* (i.e., the ability to label emotions, for instance, by analyzing facial expressions and classifying them as 'happy' or 'sad' or 'surprised'). Though inferential methods are often used for affect recognition, the sense in which I use "affective inference" in this paper relies on taking the recognition problem further, in being able to *ground* that label in some additional knowledge that reflects that nature of the world. A robot capable of affective inference should possess some knowledge (either learned or programmed) about what 'happy' facial expressions indicate about the world or a human user. Robots capable of affective inference will be better able to interpret and understand their users' actions, build better models of their users, and better predict how those users may act in the future.

Robots capable of learning from affect

Building on top of ability to draw inferences from affective information is the capability to *learn from affective data*. Because emotions form an important part of the state of the world, it is necessary to consider how emotion could be used to help shape or guide robot actions. Generally speaking, almost all learning algorithms use features extracted from raw sensory data to identify patterns and structure. We know that affective features are highly salient for humans: we unconsciously mirror others' affective states, we choose what to say or do, in part, based on our understanding of others' affective state, and many of our most fundamental social learning processes are mediated by affective signals. Thus, one path towards developing robots capable of social learning or social interaction is studying how to enable a robot to learn from affective signals by adapting modern machine learning methods to handle affective features.

Affective signals can be considered a type of human-generated training signal. Learning from human-generated reward is an active research topic within the HRI and Intelligent Agent communities. However, this work typically deals with consciously-awarded human-generated reward, often in the form of "clicker training." A significant amount of research in this area is dedicated to understanding and characterizing the nature of the human-generated reward signal to design learning algorithms better suited for these classes of reward functions.

One challenge in enabling robots to learn from affect is that, in some cases, affective reward signal (e.g., reward for smiling/frowing or other genuine facial expressions, or reward based on galvanic skin response) is not consciously given. In this respect, using affective response as reward stands in contrast to other, more studied, channels of human-generated reward. It is likely, therefore, that unconscious affective response constitutes an entirely new class of reward

function - one in which previous characterizations of human-generated reward functions may not apply. Fortunately, there are entire disciplines dedicated to understanding and characterizing human affective responses, albeit not necessarily in the context of a human-robot interaction. Researchers should look to cognitive and affective sciences for models characterizing affective response and affective learning.

Another challenge to affective learning lies in identifying aspects of the reward signal common to the task state versus aspects that are unique or personalized to an individual human teacher. Groups of humans eventually seem to converge on a small number of distinct strategies in novel agent-training tasks. However, for everyday tasks that require high levels of social interactivity, humans express themselves emotionally in highly personal and context-dependent styles. Whether a computationally tractable set of affective expressive styles can be identified and made useful to an affectively learning robot remains an open question.

Even so, it seems likely that action policies learned from affective signals in the context of one person will not directly transfer to another person, or even the same person in a different context. Personalized and long-term learning models are active research topics, but there has been little work investigating how to adapt methods from these fields to an affective reward signal. A deeper understanding of what aspects of the affective signals are generalizable or common components across populations or contexts is another crucial step towards developing robots that can learn from affect.

Conclusion

We are now beginning to recognize the importance of emotion in daily life and understand how affect leaves its trace on nearly every aspect of our experience, from how we interpret social actions and events to how we literally perceive the physical world around us (Riener et al. 2011). Advances in cognitive neuroscience are sweeping away the old view of a unified, rational consciousness, yet the majority of learning and action models for socially interactive robots are still grounded in the principles of classically rational decision-making. The maturation of affective computing technology now provides us with the tools and opportunity to design and build emotionally intelligent robots. Affect-aware robots, designed to interpret, understand, and act on what we now know to be a fundamental aspect of human experience, will bring us closer to a future in which our technology responds intuitively, comfortably, and fluently.

References

Bechara, A.; Damasio, H.; and Damasio, A. R. 2000. Emotion, decision making and the orbitofrontal cortex. *Cerebral cortex* 10(3):295–307.

El Kaliouby, R., and Robinson, P. 2005. Real-time inference of complex mental states from facial expressions and head gestures. In *Real-time vision for human-computer interaction*. Springer. 181–200.

Paiva, A.; Leite, I.; and Ribeiro, T. 2014. Emotion modeling for social robots. *The Oxford Handbook of Affective Computing* 296.

Picard, R. W. 2000. *Affective computing*. MIT press.

Riener, C. R.; Stefanucci, J. K.; Proffitt, D. R.; and Clore, G. 2011. An effect of mood on the perception of geographical slant. *Cognition and Emotion* 25(1):174–182.

Spaulding, S. 2015. Developing affect-aware robot tutors. Master's thesis, Massachusetts Institute of Technology.