Challenges for Social Embodiment

Elisabeth André
Universität Augsburg
Universitätsstr. 6a
Augsburg, Germany
andre@informatik.uni-augsburg.de

ABSTRACT

Current research in the area of social signal processing focuses on offline analysis of previously recorded human social cues. Approaches to exploit social signal processing techniques in naturalistic environments where agents socially interact with humans are rare and typically focus on isolated aspects, such as the creation of appropriate head nods or gaze behaviors. This position paper aims to identify challenges and research objectives for the area of social signal processing in order to encourage applications with more advanced forms of social embodiment in interactive settings.

Categories and Subject Descriptors

H5.2 [**Information Interfaces and Presentation**]: User Interfaces

General Terms

Human Factors

Keywords

Social Signal Processing, Multimodal Interaction, Social Robots, Virtual Agents

1. INTRODUCTION

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, hardly any 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 this paper is to identify topics for a future research agenda in order to promote a more integrated view of techniques for the realization of autonomous agents that interact with humans in naturalistic social settings.

2. EXISTING CHALLENGES

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. Nevertheless, technologies for realizing individual components of a social agent have reached a great level of advancement. Progress in the area of social signal interpretation has been considerably fostered by a number of international challenges, such as the series of the AVEC: Audio/Visual Emotion challenges[1]. To enable the comparison of different approaches, the organizers provided various annotated corpora of human behaviors for which clearly defined test procedures had to be performed. Based on the consideration that realistic data include more than clearly expressed social cues, some challenges deliberately included naturalistic behaviors that cannot always be uniquely assigned to a particular user state or attitude. Nevertheless, current challenges have been concentrating on offline analysis. Experience has shown, however, that the recognition rates that have been obtained in offline mode cannot be kept up in online mode, see [16].

The next logical step would be to organize a challenge focusing on social signal processing techniques in an interactive scenario where a social agent - a robot or an animated character - has to engage in a conversation with a human over an extended period of time. Candace Sidner and Charles Rich [15] coined the term always-on relational agents to describe the vision of a robotic or virtual character that lives as a permanent member in a human household. In order to be able to build up a long-term social relationship with the human user, such agents need to maintain a large repertoire of activities that may be jointly conducted by the agent and the human user, such as playing cards or talking about the weather. In the ideal case, the agent's conversational skills would be indistinguishable from those of a human user.

Meeting this challenge is the objective of the Loebner Prize competition which its founder, Hugh Loebner, refers to as the first formal instantiation of the Turing Test. During the Loebner Prize competition[2], several judges communicate

RFMIR'14, November 16, 2014, Istanbul, Turkey.
Copyright 2014 ACM 978-1-4503-0615-7
http://dx.doi.org/10.1145/2666253.2666265.

[1] `http://sspnet.eu/avec2014/`

[2] `http://www.loebner.net/Prizef/loebner-prize.html`

with various chat bots and human interlocutors and rank their conversational skills. The chat bot which is considered most human-like according to this ranking wins the annual prize. However, even the dialogue contributions of the winning chat bots appear rather fragmented and schematic. At the last competition in 2013, none of the finalists was able to trick the judges. Thus, the actual challenge has not yet been achieved so far. Some people doubt the scientific value of the Loebner Prize competition because it does not promote the development of sophisticated Artificial Intelligence Technologies, but rather counts on fooling people, see [18]. Even though the chat bots are supposed to engage in a social dialogue with a human, the chat bots' social and affective skills are not explicitly addressed in the Loebner contest. Therefore, it is rather unlikely that the challenge will generate significant new insights in the area of social signal processing. Nevertheless, it might provide some inspirations for the identification of a Grand Challenge for the area of Social Signal Processing.

A more recent reformulation of a Turing Test was proposed in a recent paper by Barbara Grosz [10]: "Is it imaginable that a computer (agent) team member could behave, over the long term and in uncertain, dynamic environments, in such a way that people on the team will not notice it is not human." Barbara Grosz focuses on collaboration between humans and agents requiring, among other things, sophisticated techniques for plan recognition, information sharing and the division of labor. Even though she points out that collaboration always requires some form of social intelligence, the relevance of social cues in human-agent interaction is not explicitly addressed.

3. TOPICS FOR A RESEARCH AGENDA

The paper Barbara Grosz [10] provides an excellent starting point for the definition of a Grand Challenge for the area of Social Signal Processing because it emphasizes the agent's abilities to autonomously interact with humans in a social setting. Given the fact that the much less ambitious Loebner Prize Challenge has not been met so far, it would be unrealistic to assume that we will be able to build an embodied agent with social skills that are (nearly) indistinguishable from those of a human in the near future. However, to make progress towards this vision, future challenges should give more emphasis to integrative social skills that include both the analysis of social cues as well as believable responses to them. Here we list a number of sub challenges that should be addressed in order to move towards this goal:

- *Investigation of Novel Modalities, such as Olfactory and Tactile Modalities, Smoothly Integrated with Traditional Modalities*
  First attempts have been made to enhance social agents by touch. Bickmore and colleagues [3] implemented a system consisting of a screen with an animated face installed on top of a mannequin. Human touch was simulated by squeezing the user's hands using an air bladder at the mannequin's hand. Gaffary and colleagues [9] made use of a virtual character to convey facial expressions and an air jet to simulate the haptic modality. While the experiments of both groups provided interesting insights regarding the complementary functions of haptic and visual cues, a smoother integration of the haptic modality with traditional modalities would be necessary to leverage haptic in future social signal processing applications.

- *Integration of Social Cue Analysis with Semantic and Pragmatic Analysis*
  Work done in the Semaine project [17] 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 [4].

- *Contextualized Analysis of Social Signals in Real-Life Settings*
  In the area of Pervasive Computing, a number of wearable applications have been developed that detect aspects of social behaviors, see [19] for an overview. However, the repertoire of investigated features is rather limited focusing on the amount of conversation recorded by the smart phone's microphone [5], communication data [12] or proximity behaviors detected by Bluetooth patterns [7], or postures [8]. Most of the approaches concentrate on offline analysis. Typically, data of people are collected over a certain period of time and analyzed afterwards. A promising pathway for the future is the integration of research done in the area of Social Signal Processing and research done in the area of Pervasive Computing in order to leverage mobile applications, such as a social coach that is employed in a real-life setting [6]. Furthermore, information obtained from context analysis and activity recognition might compensate for ambiguities in the interpretation of social cues.

- *Experience-Based Learning and Adapting for Diversity*
  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 [2]. To this end, sophisticated mechanisms for the simulation of self-regulatory social behaviors will be required. Furthermore, social interactions will have to be personalized to persons of different gender, personality and cultural background. For example, cultural norms and values determine whether it is appropriate to show emotions in a particular situation [14] and how they are interpreted by others [13].

While offline learning is prevalent in current systems exploiting social signal processing techniques, future work should explore the potential of online learning in order to enable continuous social adaptation processes.

- *Finding the Right Level of Sensitivity*
The further perfectionization of techniques for the analysis of social signals might lead to agents that respond to human signals in an oversensitive manner [1]. 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 [11] 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.

4. REFERENCES

[1] E. André. Exploiting unconscious user signals in multimodal human-computer interaction. *ACM Trans. Multimedia Comput. Commun. Appl.*, 9(1s):48:1–48:5, Oct. 2013.

[2] R. Aylett, G. Castellano, B. Raducanu, A. Paiva, and M. Hanheide. Long-term socially perceptive and interactive robot companions: challenges and future perspectives. In H. Bourlard, T. S. Huang, E. Vidal, D. Gatica-Perez, L.-P. Morency, and N. Sebe, editors, *Proceedings of the 13th International Conference on Multimodal Interfaces, ICMI 2011, Alicante, Spain, November 14-18*, pages 323–326. ACM, 2011.

[3] T. W. Bickmore, R. Fernando, L. Ring, and D. Schulman. Empathic touch by relational agents. *T. Affective Computing*, 1(1):60–71, 2010.

[4] W. Bosma and E. André. Exploiting emotions to disambiguate dialogue acts. In *Proceedings of the 9th International Conference on Intelligent User Interfaces*, IUI '04, pages 85–92, New York, NY, USA, 2004. ACM.

[5] K.-h. Chang, D. Fisher, J. Canny, and B. Hartmann. How's my mood and stress?: An efficient speech analysis library for unobtrusive monitoring on mobile phones. In *Proceedings of the 6th International Conference on Body Area Networks*, BodyNets '11, pages 71–77, 2011.

[6] I. Damian, C. S. S. Tan, T. Baur, J. Schöning, K. Luyten, and E. André. Exploring social augmentation concepts for public speaking using peripheral feedback and real-time behavior analysis. In *International Symposium on Mixed and Augmented Reality (ISMAR), München, September 10-12th*, 2014.

[7] T. M. Do and D. Gatica-Perez. Human interaction discovery in smartphone proximity networks. *Personal Ubiquitous Comput.*, 17(3):413–431, Mar. 2013.

[8] S. Feese, B. Arnrich, G. Troster, B. Meyer, and K. Jonas. Detecting posture mirroring in social interactions with wearable sensors. In *Proceedings of the 2011 15th Annual International Symposium on Wearable Computers*, ISWC '11, pages 119–120, Washington, DC, USA, 2011. IEEE Computer Society.

[9] Y. Gaffary, J.-C. Martin, and M. Ammi. Perception of congruent facial and haptic expressions of emotions. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '14, pages 135–135, New York, NY, USA, 2014. ACM.

[10] B. J. Grosz. What question would turing pose today? *AI Magazine*, 33(4):73–81, 2012.

[11] A. Hoekstra, H. Prendinger, N. Bee, D. Heylen, and M. Ishizuka. Highly realistic 3d presentation agents with visual attention capability. In A. Butz, B. D. Fisher, A. Krüger, P. Olivier, and S. Owada, editors, *Smart Graphics, 7th International Symposium, SG 2007, Kyoto, Japan, June 25-27, 2007, Proceedings*, volume 4569 of *Lecture Notes in Computer Science*, pages 73–84. Springer, 2007.

[12] R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong. Moodscope: Building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '13, pages 389–402, New York, NY, USA, 2013. ACM.

[13] D. Matsumoto. Cultural influences on the perception of emotion. *Journal of Cross-Cultural Psychology*, 20(1):92–105, 1989.

[14] D. Matsumoto. Cultural similarities and differences in display rules. *Motivation and Emotion*, 14(3):195–214, 1990.

[15] C. Rich and C. L. Sidner. Collaborative discourse, engagement and always-on relational agents. In *AAAI Fall Symposium: Dialog with Robots, Papers from the 2010 AAAI Fall Symposium, Arlington, Virginia, USA, November 11-13, 2010*, volume FS-10-05 of *AAAI Technical Report*. AAAI, 2010.

[16] S. Scherer, M. Glodek, F. Schwenker, N. Campbell, and G. Palm. Spotting laughter in natural multiparty conversations: A comparison of automatic online and offline approaches using audiovisual data. *TiiS*, 2(1):4, 2012.

[17] M. Schröder, E. Bevacqua, R. Cowie, F. Eyben, H. Gunes, D. Heylen, M. ter Maat, G. McKeown, S. Pammi, M. Pantic, C. Pelachaud, B. Schuller, E. de Sevin, M. F. Valstar, and M. Wöllmer. Building autonomous sensitive artificial listeners. *T. Affective Computing*, 3(2):165–183, 2012.

[18] S. M. Shieber. Lessons from a restricted turing test. *Commun. ACM*, 37(6):70–78, June 1994.

[19] A. Vinciarelli, R. Murray-Smith, and H. Bourlard. Mobile social signal processing: vision and research issues. In M. de Sá, L. Carriço, and N. Correia, editors, *Proceedings of the 12th Conference on Human-Computer Interaction with Mobile Devices and Services, Mobile HCI 2010, Lisbon, Portugal, September 7-10*, ACM International Conference Proceeding Series, pages 513–516. ACM, 2010.