Modelling users' affect in job interviews: technological demo

Kaśka Porayska-Pomsta, Keith Anderson, Ionut Damian, Tobias Baur, Elisabeth André, Sara Bernardini, Paola Rizzo

Modelling Users' Affect in Job Interviews: Technological Demo

Kaśka Porayska-Pomsta[1], Keith Anderson[2], Ionut Damian[3], Tobias Baur[3], Elisabeth André[3], Sara Bernardini[4], and Paola Rizzo[1]

[1] London Knowledge Lab, Institute of Education
23-29 Emerald Street, London WC1N 3QS, UK
{K.Porayska-Pomsta,P.Rizzo}@ioe.ac.uk
[2] Tandemis Limited
108 Blackheath Hill, London SE10 8AG, UK
keith@tandemis.co.uk
[3] Human Centered Multimedia, Augsburg University
Universitätsstr. 6a, 86159 Augsburg, Germany
{damian,baur,andre}@hcm-lab.de
[4] Department of Informatics, King's College London, UK, WC2R 2LS
sara.bernardini@kcl.ac.uk

Abstract. This demo presents an approach to recognising and interpreting social cues-based interactions in computer-enhanced job interview simulations. We show what social cues and complex mental states of the user are relevant in this interaction context, how they can be interpreted using static Bayesian Networks, and how they can be recognised automatically using state-of-the-art sensor technology in real-time.

1 Introduction

A recruiter ascertains a job seeker's fit to a given job based largely on the social cues, i.e. the job seeker's conscious or unconscious behaviours. This demo shows an approach to recognising and interpreting relevant social cues of interviewees' that is based on a combination of social signal processing and Bayesian Networks (BNs). The work is conducted within the TARDIS project (http://tardis.lip6.fr/) – a scenario-based serious game where virtual agents act as recruiters.

2 Relevant Work

Recent work in the field of behavioural analysis has shown that social cues significantly influence the outcome of job interviews [1, 2]. Several computer-based environments for social skills acquisition have been proposed (e.g.[3], [4]).

In the field of signal processing, most research has focused on the recognition of emotions from speech [5] or facial expressions [6], with the analysis of postures

[7] and gestures [8] having received little attention, especially in the context of the interactive scenarios considered in TARDIS.

To help the target users acquire the relevant social skills, our system needs to recognise a comprehensive set of the users' social cues. Such cues provide the basis for inferring users' complex mental states and for selecting appropriate responses by virtual recruiters.

3 Analysis of Social Cues

We conducted a preliminary study with users, during which 10 real job seekers and 5 practitioners engaged in mock interviews. The data was video recorded and annotated manually. The analysis focused specifically on the frequency of social-cue occurrence across the different interactions.
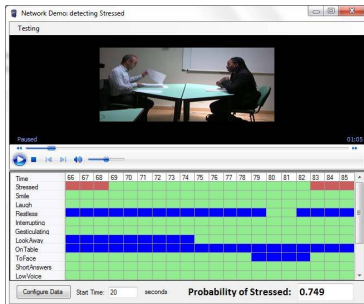

Fig. 1. Video playback facility, with the annotation data and Bayesian network output in the bottom-right corner of the screen
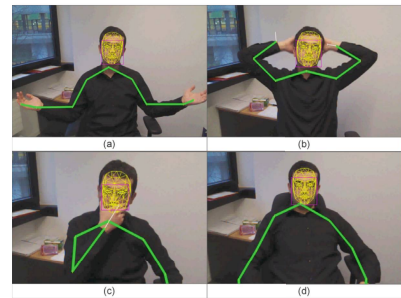

Fig. 2. Postures the system can recognize: (a) arms open, (b) hands behind head, (c) hand close to head, (d) leaning backwards

Episodes as in Fig. 1 were annotated for fine-grained behaviours, e.g. looking away, lack of direct eye contact, smiling, etc. The annotations were mapped onto seven complex mental states, identified by the practitioners during post-hoc walkthroughs and semi-structured interviews, as relevant to this context: *stressed, embarrassed, ill-at-ease, bored, focussed, hesitant, relieved*. The annotations were fed into a C# application which calculates the frequencies of the co-occurrences of groups of social cues with each mental state. This data is used to infer the BNs' probability tables, representing the complex mental states. Another C# application evaluates networks and consists of two components: (1) simultaneous playback of video, annotations used to the train the BNs, and BNs real-time output for a given mental state; (2) a facility that feeds the annotations through the network and produces statistics about the network's correct and incorrect classifications. A demonstration of these applications will be given.

4   Automatic Recognition of Social Cues

We investigated how the social cues identified during the user study can be recognised automatically, using a combination of sensors and software algorithms that yield best results in terms of accuracy, low intrusion, reliability, set-up time and cost. As a first step, we implemented the recognition of the following six social cues: *Hands to face; Looking away; Postures (arms open/ crossed and hands behind head); Leaning back/forward; Voice activity (interrupting the interviewer, short answers and long silence).* – see Fig. 2. These feed directly into the BNs.

We chose the Microsoft Kinect as the main sensor for social cue recognition because it is low-cost, it does not require any time-consuming configuration, it is relatively robust against lighting conditions, it incorporates a microphone and an RGB camera in addition to the depth camera, and it is minimally intrusive. For recording and pre-processing human behaviour data, our system relies on the SSI framework [9], the demonstration of which will be also provided.

References

1. Curhan, J., Pentland, A.: Thin slices of negotiation: predicting outcomes from conversational dynamics within the first 5 minutes (2007)
2. Arvey, R.D., Campion, J.E.: The employment interview: A summary and review of recent research. Personnel Psychology 35(2), 281–322 (1982)
3. Bernardini, S., Porayska-Pomsta, K., Smith, T.J., Avramides, K.: Building autonomous social partners for autistic children. In: Nakano, Y., Neff, M., Paiva, A., Walker, M. (eds.) IVA 2012. LNCS, vol. 7502, pp. 46–52. Springer, Heidelberg (2012)
4. Vala, M., Sequeira, P., Paiva, A., Aylett, R.: Fearnot! demo: a virtual environment with synthetic characters to help bullying. In: Proc. 6th Intl. Joint Conf. on Autonomous Agents and Multiagent Systems, AAMAS 2007, pp. 271:1–271:2. ACM, New York (2007)
5. Vogt, T., Andre, E.: Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition. In: IEEE Intl. Conf. on Multimedia and Expo, ICME 2005, pp. 474–477 (July 2005)
6. Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(1), 39–58 (2009)
7. Kapoor, A., Picard, R.W.: Multimodal affect recognition in learning environments. In: Proc. 13th Annual ACM Intl. Conf. on Multimedia, MULTIMEDIA 2005, pp. 677–682. ACM, New York (2005)
8. Kleinsmith, A., Bianchi-Berthouze, N.: Form as a cue in the automatic recognition of non-acted affective body expressions. In: D'Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (eds.) ACII 2011, Part I. LNCS, vol. 6974, pp. 155–164. Springer, Heidelberg (2011)
9. Wagner, J., Lingenfelser, F., André, E.: The social signal interpretation framework (SSI) for real time signal processing and recognition. In: Proc. Interspeech 2011 (2011)