Multimedia Information Extraction Roadmap: Integration of Affective Computing and Multimedia Information Extraction

Elisabeth André

Augsburg University, Germany

In the following, I focus on one particular challenge: the integration of affective computing and multimedia information extraction.

Critical Technical Challenges

What are the critical technical challenges in multimedia information extraction (MMIE)?

Work by Picard and others has created considerable awareness for the role of affect in human computer interaction. As key ingredients of affective computing Picard identifies recognizing, expressing, modelling, communicating, and responding to emotional information (Picard 2003). In the context of information extraction, methods of affective computing can be applied to enhance classical information extraction tasks by emotion and sentiment detection. In addition, the experience of users interacting with an information extraction system might be improved by recognizing and appropriately responding to the user's affective state (see also Belkin's keynote "Some(what) Grand Challenges for Information Retrieval" at ECIR 2008 which discusses the use of affective user interfaces to information retrieval systems). The availability of robust methods for emotion recognition is an important step in the development of information extraction systems that are sensitive to the user's emotional state. In addition, further research is required to shed light on the question of how to respond to the user's emotional state when he or she is interacting with an information extraction system. Finally, it is still unclear how to evaluate the potential benefits of an affect-aware information extraction system.

Existing Approaches

What are the important existing methods, techniques, data and tools that can be leveraged?

With labelled data collected from different modalities, such as speech, facial expressions, and postures, most studies rely on supervised pattern classification approaches for automatic emotion recognition, see Vogt et al. (2008) for an overview on emotion recognition from speech and Pantic & Rothkrantz (2003) for an overview on emotion recognition from facial expressions. Cues that have been investigated include postures, for example, see Kapoor and Picard (2005), gestures, for example, see De Silva et al. (2006), facial expressions, for example, see Chen and Huang (2000), acoustic and prosodic features of speech, for example see Batliner et al. (2006). Existing approaches in the area of text-based sentiment analysis may be categorized into: simple keyword spotting using emotion words, such as "happy", lexical affinity approaches that are based on affinity probabilities between words and emotions (e.g. the word "injury" indicates in 80% of the cases a negative emotion), statistical approaches using corpora annotated with emotional information and semantic-based approaches relying on hand-crafted models or knowledge-based approaches, see Liu and colleagues (2003) for an overview. In order to improve the recognition accuracy obtained from uni-modal emotion recognition systems, a number of studies attempted to exploit the advantage of using multimodal information, especially by fusing audio-visual information, see for example the work De Silva and Ng (2000) and Chen and Huang (2000).

Methods and Techniques

There is a vast body of literature on the automatic recognition of emotions. Instead of providing a literature overview, we refer to the web page of the Humaine Association which includes a comprehensive collection of literature on emotion research. The Humaine Association is a follow-up of the FP6 Network of Excellence Humaine, which aimed to lay the foundations for European development of systems that can register, model and influence human emotional and emotion-related states and processes - 'emotion-oriented systems'.

DataSets

A collection of databases suitable for affect sensing may be found under: http://emotion-research.net/wiki/Databases

Tools

A collection of various kinds of tools in the area of emotion research can be found in the toolbox folder of the Humaine association http://emotion-research.net/toolbox

A software methodology for the development and the engineering of affective multimodal interfaces is currently being developed within the CALLAS project (http://www.callas-newmedia.eu/). It will come with a collection of components that can be used to generate emotionally-aware user interfaces.

Remaining Gaps

What key technology gaps remain that require focused research?

More research needs to be conducted in order to advance the state of the art in emotion recognition by focusing on spontaneous non-acted affective states in natural environments for which acceptable recognition rates could not yet been obtained. Specific research topics include: synergistic fusion mechanism that take into account multiple channels as well as the situative context, the recognition of emotion blends, the determination of confidence values for recognized emotional states and the integration of cognitive models of emotion to improve the classification process. In the area of text-based emotion or sentiment detection, we need to move from a pure lexical analysis to deeper semantic and pragmatic processing. Furthermore, linguistic features, such as the structure of an utterance, should be augmented by acoustic features in order to improve sentiment detection.

References

Batliner, A., Fischer, R. Huber, J. Spilker, and E. Nöth, “How to find trouble in communication,” *Speech Communication*, vol. 40, pp. 117–143, 2003.

L. S. Chen and T. S. Huang. Emotional expressions in audiovisual human computer interaction. In ICME-2000, pages 423–426, 2000.

L. C. De Silva and P. C. Ng. Bimodal emotion recognition. In IEEE International Conf. on Automatic Face and Gesture Recognition, pages 332–335, March 2000.

Kapoor, A., & Picard, R.W. (2005). Multimodal affect recognition in learning environments. In MULTIMEDIA ’05: Proceedings of the 13th annual ACM international conference on Multimedia, New York, NY, USA, ACM Press. 677–682.

Liu, H., Liebermann, H. & Selker, T. (2003. A model of textual affect sensing using real-world knowledge, Proceedings of the 8th international conference on Intelligent user interfaces, pp. 125 – 132.

Pantic, M., Rothkrantz, L.J.M. Toward an affect-sensitive multimodal human-computer interaction. Proceedings of the IEEE , Volume: 91 Issue: 9 , Sept. 2003. Page(s): 137 –1390.

Picard, R. (2003). Affective computing: challenges. Int. J. Human-Computer Studies 59, pp. 55–64

Thurid Vogt, Elisabeth André and Johannes Wagner, "Automatic Recognition of Emotions from Speech: a Review of the Literature and Recommendations for Practical Realisation", Affect and Emotion in Human-Computer Interaction , Christian Peter, Russell Beale, Springer, Heidelberg, Germany, in press