Dimensional and Continuous Analysis of Emotions for Multimedia Applications

[A Tutorial Overview]

Hatice Gunes
Queen Mary University London
School of Electronic Eng. & Computer Science
Mile End Road, London E1 4NS U.K.
hatice@eecs.qmul.ac.uk

Björn Schuller[1,2]
[1] JOANNEUM RESEARCH
DIGITAL, 8010 Graz, Austria
[2] TUM – MMK, 80333 Munich, Germany
schuller@IEEE.org

Categories and Subject Descriptors

H.1.2 [**User / Machine systems**]: Human information processing; H.5.5 [**Sound and Music Computing**]: Methodologies and techniques, Modeling, Signal analysis, synthesis, and processing; I.2.10 [**Vision and scene understanding**]: Video analysis; I.5.2 [**Design Methodology**]: Feature evaluation and selection, Classifier design and evaluation; I.5.4 [**Applications**]: Signal processing

General Terms

Human Factors, Signal processing, Feature evaluation and selection, Classifier design and evaluation, Experimentation

Keywords

Affective Computing, Dimensional Emotion Prediction, Continuous Analysis, Audio/Visual Emotion Recognition, Sentiment Analysis, Music Mood Classiffication, Tutorial

1. MOTIVATION

Emotion has attracted the interest of the multimedia community in two forms, namely as *emotion in media* and *emotion evoked by media*. Emotion in media refers to people in video or audio documents where they interact and communicate subtle emotions and affective states by means of language, vocal intonation, facial expression, head movement, body movement and posture. This causes multimedia content to be heavily loaded with emotion in speech [4], music [6], sound [5], text [4], and video [2]. The wide-spread nature of such affective content has also brought about interest in emotion and mood classification in multimedia, e.g., mood classification in music, sentiment and opinion mining in text, etc. Emotion evoked by media focuses on the analysis of emotions evoked in human observers by sounds played and images displayed. This area of research is also closely linked to implicit human-centered tagging where the aim is to automatically tag multimedia data based on analysing user's nonverbal reactions (e.g., laughter or accelerated heart beat) to such data [3]. Affective computing aims to equip multimedia computing devices (such as personal PCs, smart phones, media players) with the means to retrieve, interpret, understand, and respond to emotions, affect and moods – similar to the way humans rely on their senses to assess each other's communicative and affective state. In the last 15 years researchers in the affective computing and related fields have invested increased effort into creating automatic (media) processing systems, and exploring how to best model, analyze and interpret the subtlety, complexity and continuity of human affective and communicative behavior, and affective cues encoded in music, image, and text, and how to make these findings applicable to naturalistic and realistic application settings (e.g., multimedia retrieval, interaction with robots, virtual agents (tutors), and games; single and multi-user smart environments, e.g., in a car; implicit (multimedia) tagging; clinical and biomedical studies, e.g., autism, depression, pain; etc.) [1]. Such efforts have been bearing fruit – the affective computing field has undergone a rapid growth and has become a highly active area of research and practice. There are now various conferences and workshops dedicated to expand various research ideas in these fields (e.g., the conference series on Affective Computing and Intelligent Interaction (ACII)). This resulted in a number of valuable publications (e.g., IEEE Tran. on Affective Computing, and various textbooks) and portals (e.g., the HUMAINE Portal[1]). Despite such notable developments, the application aspect of affective computing has largely been neglected – it has not managed to go out 'into the wild'. This is mainly due to the fact that human and media encoded emotions are complex constructs with fuzzy boundaries. A small number of discrete categories (e.g., happiness and sadness) may not reflect the subtlety and complexity of the emotions conveyed by such rich sources of information. Therefore, researchers in relevant fields have recently invested increased effort exploring how to best model, analyze and interpret the subtlety, complexity and continuity (represented along a temporal continuum) of affective behavior in terms of latent dimensions (e.g., arousal, power and valence) and appraisals, rather than in terms of a small number of discrete emotion categories (e.g., happiness, sadness, surprise, disgust, fear and anger) [1]. This is particularly needed for multimedia applications where enabling automatic and continuous emotion analysis would be extremely beneficial for personalized and emotion-sensitive multimedia content analysis and processing, implicit human-centered tagging, multimedia event understanding, search and retrieval, multimedia interaction and digital art installations, etc.

---

[1] *http://www.emotion-research.net*

2. AIMS

In the light of the above, this tutorial aims to become the initial but crucial step towards bringing together researchers and academics from the very relevant yet disconnected fields of research and practice: affective computing and multimedia. The tutorial aims to give a comprehensive introduction to automatic, dimensional and continuous analysis of emotions and affective signals, and provide indicators and examples of how the current developments in this field can be utilized to enhance multimedia applications.

More specifically, the tutorial aims to provide answers to the following questions: 1) Why has the field shifted towards dimensional and continuous analysis and interpretations of affective displays recorded in naturalistic settings? 2) What are the affect models used, and the affect signals measured? 3) How has the current automatic measurement technology been developed? 4) How can the recent advancements in this field be utilized to enhance multimedia applications? The tutorial will focus on providing a broad overview of recent algorithms and methodology, and predict potentially upcoming trends for relevant multimedia applications.

3. BIOGRAPHIES

Dr Hatice Gunes is a Lecturer (Assistant Prof.) at the School of Electronic Eng. & Computer Science, Queen Mary University of London (QMUL), UK. Prior to joining QMUL, she was a postdoctoral researcher at Imperial College London, UK, and an Associate of University of Technology Sydney (UTS), Australia. She received her Ph. D. degree in Computer Science from UTS in 2007. Her research interests lie in the areas of affective computing, visual information processing, and machine learning, focusing on automatic affective behaviour analysis, continuous prediction, and multicue and multimodal emotion recognition. Dr Gunes has published more than 50 technical papers in these areas that received more than 600 citations (h-index = 15). She has acted as the main organiser of the 1st workshop on automatic affect analysis in continuous and multidimensional space (EmoSPACE), organised in conjunction with IEEE FG'11 and is a Guest Editor of Special Issues in Int'l Journal of Synthetic Emotions and Image and Vision Computing Journal. She has also served as a member of the Editorial Advisory Board for the Affective Computing & Interaction Book (IGI Global, 2011), as an area chair for ACII 2013, as a general workshop chair of BCS HCI 2012, as an invited speaker at the Int'l Workshop on Social Signal Processing (WSSP 2011) and the Summer School on Affective Computing & Social Signal Processing (ACSSP 2010). From 2004 to 2007, for her PhD research in affective computing, she was a recipient of the Australian Government Int'l Postgraduate Research Scholarship (IPRS) awarded to top quality international postgraduate students. Dr Gunes, together with co-authors, has also received awards for Outstanding Paper (IEEE FG 2011), Quality Reviewer (IEEE ICME 2011), Best Demo (IEEE ACII 2009), and Best Student Paper (VisHCI 2006). In 2012 she received funding from the British Council under the UK-Turkey HE Partnership Programme and is currently the co-chair of the Int'l Workshop on Affective Computing for Mobile HCI. She is member of the IEEE, ACM, and HUMAINE Association.

Dr Björn W. Schuller received his diploma in 1999 and his doctoral degree for his study on Automatic Speech and Emotion Recognition in 2006, both in electrical engineering and information technology from TUM (Munich University of Technology) where he also finished his habilitation thesis in 2012. There, he also finished his habilitation thesis on Intelligent Audio Analysis. At present, he is with JOANNEUM RESEARCH in Graz/Austria and tenured as Senior Lecturer heading the Intelligent Audio Analysis Group at TUM's Institute for Human-Machine Communication since 2006. From 2009 to 2010 he was with the CNRS-LIMSI Spoken Language Processing Group in Orsay/France dealing with affective and social signals in speech. In 2010 he was also a visiting scientist in the Imperial College London's Department of Computing in London/UK working on audiovisual behaviour recognition. Dr Schuller is president-elect of the HUMAINE Association and member of the ACM, IEEE and ISCA and (co-)authored 4 books and more than 300 publications leading to more than 2 800 citations (h-index = 28). He serves as co-founding member and secretary of the steering committee, associate editor, and guest editor of the IEEE Tran. on Affective Computing, associate and repeated guest editor for the Computer Speech and Language, associate editor for the IEEE Tran. on Systems, Man and Cybernetics: Part B Cybernetics and the IEEE Tran. on Neural Networks and Learning Systems, and guest editor for the IEEE Intelligent Systems Magazine, Speech Communication, Image and Vision Computing, Cognitive Computation, and the EURASIP Journal on Advances in Signal Processing, and as workshop and challenge organizer including the first of their kind INTERSPEECH 2009 Emotion, 2010 Paralinguistic, 2011 Speaker State, and 2012 Speaker Trait Challenges and the 2011 and 2012 Audio/Visual Emotion Challenge and Workshop and programme committee member of more than 40 international workshops and conferences. He coordinates the European Community funded ASC-Inclusion STREP project that provides interactive emotion games for ASC children and was invited expert in the W3C Emotion Markup Language Incubator Groups.

4. REFERENCES

[1] GUNES, H., AND SCHULLER, B. Categorical and Dimensional Affect Analysis in Continuous Input: Current Trends and Future Directions. *Image & Vision Computing* (2012). 26 pages.

[2] NICOLAOU, M., GUNES, H., AND PANTIC, M. Output-Associative RVM Regression for Dimensional and Continuous Emotion Prediction. *Image & Vision Computing 30*, 3 (March 2012), 186–196.

[3] PANTIC, M., AND VINCIARELLI, A. Implicit Human-Centered Tagging. *IEEE Signal Processing Magazine 26*, 6 (2009), 173–180.

[4] SCHULLER, B., BATLINER, A., STEIDL, S., AND SEPPI, D. Recognising Realistic Emotions and Affect in Speech: State of the Art and Lessons Learnt from the First Challenge. *Speech Communication 53*, 9/10 (2011), 1062–1087.

[5] SCHULLER, B., HANTKE, S., WENINGER, F., HAN, W., ZHANG, Z., AND NARAYANAN, S. Automatic Recognition of Emotion Evoked by General Sound Events. In *Proc. IEEE ICASSP* (2012), pp. 341–344.

[6] SCHULLER, B., WENINGER, F., AND DORFNER, J. Multi-Modal Non-Prototypical Music Mood Analysis in Continuous Space: Realiability and Performances. In *Proc. ISMIR* (2011), pp. 759–764.