Künstl Intell (2016) 30:91–92
DOI 10.1007/s13218-015-0401-0

Multimodal Behavior Analytics for Interactive Technologies

Stefan Scherer[1]

Published online: 15 October 2015

Keywords Multimodal behavior analytics · Machine learning · Nonverbal behavior

1 Introduction

Human communication is multifaceted and information between humans is communicated on many channels in parallel. In order for a machine to become an efficient and accepted social companion, it is important that the machine understands interactive cues that not only represent direct communicative information such as spoken words but also nonverbal behavior. Hence, technologies to understand and put nonverbal communication into the context of the present interaction are essential for the advancement of human-machine interfaces [3, 4].

Multimodal behavior analytics—a transdisciplinary field of research—aims to close this gap and enables machines to automatically identify, characterize, model, and synthesize individuals' multimodal nonverbal behavior within both human-machine as well as machine-mediated human-human interaction.

The emerging technology of this field is relevant for a wide range of interaction applications, including but not limited to the areas of healthcare and education. Exemplarily, the characterization and association of nonverbal behavior with underlying clinical conditions, such as depression or post-traumatic stress, holds transformative potential and could change treatment and the healthcare systems efficiency significantly [6].

✉ Stefan Scherer
scherer@ict.usc.edu

[1] Institute for Creative Technologies, University of Southern California, Los Angeles, USA

Within the educational context the assessment of proficiency and expertise of individuals' social skills, in particular for those with learning disabilities or social anxiety, can help create individualized education scenarios [2, 8]. The potential of machine-assisted training for individuals with autism spectrum disorders (ASD) for example could have far reaching impacts on our society.

In the following, I highlight two behavior analytics approaches that were investigated in my PhD dissertation [3] and summarized in a multimodal framework for human behavior analysis [4].

2 Multimodal Behavior Analytics

Laughter Detection One of the most iconic human behaviors is laughter; it is universally understood by all cultures and yet is immensely versatile in its meaning and variable in its expression (e.g. inhaled, exhaled, snort-like laughs as well as laughter bearing various meanings, e.g. humorous, nervous, or social laughter [1]). Due to the relative importance of laughter and its potential impact on human-machine interaction, we investigated the capability of multimodal sequential classifiers to spot laughter in natural multiparty conversations [5]. Utilizing all available data channels, we extracted three independent feature streams, including frequency and spectrum based features from the audio stream, and coarse movement related features from the video stream. Utilizing multimodal sequence classifiers such as hidden Markov models (HMM) and echo state networks (ESN) we achieved considerable accuracies in recognizing this challenging human paralinguistic behavior (F_1 = 0.72 for the HMM, with 0.8 recall and 0.64 precision; F_1 = 0.63 for the ESN with 0.81 recall and 0.52 precision).

Voice Quality Recognition Voice quality, a term that refers to the timbre or coloring of the voice, serves many purposes in human-human communication. In particular, the dynamic use of voice qualities in spoken communication informs us about the attitude, mood, social status, and affective state of the speaker. Yet voice quality is very difficult to identify and often even confused by human experts. In order to investigate the usefulness of uncertain or fuzzy information provided by human experts, we analyzed the classification performance of fuzzy-input fuzzy-output support vector machines (F^2SVM) [7]. These F^2SVM outperformed other state of the art approaches significantly, by solely utilizing the information provided by the fuzzy annotations of the human experts during training on a subset of the voice quality data for which the majority vote of the human annotators always coincided with the actual target label. The F^2SVM classified the voice quality samples with an error rate of 13.88 % ($\sigma = 3.89$) in speaker independent classification experiment, and 17.66 % in the cross corpus experiment. This experiment indeed shows that the usage of fuzzy and uncertain information has the potential to improve classification results.

3 Concluding Remarks

To date, we have only scraped the surface of understanding human nonverbal communicative behavior utilizing these novel objective and quantitative behavior analytics approaches. Yet, this vibrant and highly multidisciplinary research that integrates the fields of psychology, machine learning, multimodal sensor fusion, and pattern recognition, emerges as an essential field of investigation for computer science. The thorough understanding of the underlying mechanisms of human behavior will advance the development of technology that tightly cooperates with human interactants and has the potential to optimize human performance and well-being alike.

References

1. Campbell N, Kashioka H, Ohara R (2005) No laughing matter. In Proceedings of Interspeech 2005, ISCA, pp 465–468
2. Chollet M, Wortwein T, Morency L-P, Shapiro A, Scherer S (2015) Exploring feedback learning strategies to improve public speaking: an interactive virtual audience framework. In accepted for publication at UbiComp 2015
3. Scherer S (2011) Analyzing the User's State in HCI: From Crisp Emotions to Conversational Dispositions. PhD thesis, Ulm University
4. Scherer S, Glodek M, Layher G, Schels M, Schmidt M, Brosch T, Tschechne S, Schwenker F, Neumann H, Palm G (2012) A generic framework for the inference of user states in human computer interaction: How patterns of low level communicational cues support complex affective states. J Multimodal User Interf Spec Issue Concept Framew Multimodal Social Signal Proces 6(3):117–141
5. Scherer S, Glodek M, Schwenker F, Campbell N, Palm G (2012) Spotting laughter in natural multiparty conversations: a comparison of automatic online and offline approaches using audiovisual data. ACM Transac Interact Intell Syst Spec Issue Affect Interact Nat Environ 2(1):4:1–4:31
6. Scherer S, Stratou G, Lucas G, Mahmoud M, Boberg J, Gratch J, Rizzo A, Morency L.-P (2014) Automatic audiovisual behavior descriptors for psychological disorder analysis. Image Vision Comput J Spec Issue Best Face Gesture 2013 32(10):648–658
7. Thiel C, Scherer S, Schwenker F(2007) Fuzzy-input fuzzy-output one-against-all support vector machines. In 11th international conference on knowledge-based and intelligent information and engineering systems (KES'07), volume 3 of Lecture Notes in Artificial Intelligence, Springer, Heidelberg, pp 156–165
8. Wortwein T, Morency L-P, Scherer S (2015) Automatic assessment and analysis of public speaking anxiety: A virtual audience case study. In: To appear in Proceedings of IEEE Affective Computing and Intelligent Interaction 2015 (ACII)

Stefan Scherer is a Research Assistant Professor at the University of Southern California (USC) and the USC Institute for Creative Technologies where he leads research projects funded by NSF and ARL. He received the degree of Dr. rer. nat. from Ulm University with the grade summa cum laude. His research aims to automatically identify, characterize, model, and synthesize individuals' multimodal nonverbal behavior within both human-machine as well as machine-mediated human-human interaction. (http://schererstefan.net)