Summary for AVEC 2016 – Depression, Mood, and Emotion Recognition Workshop and Challenge

Michel Valstar
University of Nottingham
School of Computer Science

Jonathan Gratch
University of Southern
California, ICT

Björn Schuller[*]
University of Passau
Chair of Complex & Intelligent
Systems

Fabien Ringeval
Université Grenoble Alpes
Laboratoire d'Informatique

Roddy Cowie
Queen's University Belfast
Department of Psychology

Maja Pantic[†]
Imperial College London
Intelligent Behaviour
Understanding Group

ABSTRACT

The sixth Audio-Visual Emotion Challenge and workshop AVEC 2016 was held in conjunction ACM Multimedia'16. This year the AVEC series addresses two distinct sub-challenges, multi-modal emotion recognition and audio-visual depression detection. Both sub-challenges are in a way a return to AVEC's past editions: the emotion sub-challenge is based on the same dataset as the one used in AVEC 2015, and depression analysis was previously addressed in AVEC 2013/2014. In this summary, we mainly describe participation and its conditions.

Keywords

Affective Computing; Social Signal Processing; Depression

1. INTRODUCTION

This year's Audio-Visual Emotion Challenge and workshop (AVEC 2016) has been organised in conjunction with the 24th ACM International Conference on Multimedia, held in Amsterdam, The Netherlands, 15 – 19 October 2016.

The AVEC 2016's theme is 'Depression, Mood, and Emotion Recognition' and it is the sixth competition event aimed at comparison of multimedia processing and machine learning methods for automatic audio, visual and audio-visual emotion analysis, with all of the participants competing under strictly the same conditions in this first of its kind series [3, 5, 6, 9, 10]. However, further similar endeavours that

[*]The author is further affiliated with Imperial College London, Department of Computing, London, U.K.

[†]The author is further affiliated with Twente University, EEMCS, Twente, The Netherlands.

have been undertaken since AVEC started in 2011 are to be noted, e. g., [1].

The AVEC 2016 edition features besides true multimodal emotion recognition in its emotion recognition sub-challenge, where besides audio and video participants also have access to physiological data. In addition, we take emotion analysis to a higher-level behaviour understanding by looking at mood disorders, in particular the serious and important problem of automatic detection of depression. Adding objective measures to what is otherwise an entirely subjective process of diagnosing and monitoring depression promises to be an invaluable support to the mental health profession besides the usage in media retrieval systems.

As before, the goal of the Challenge is to provide a common benchmark test set for individual multimodal information processing and to bring together the audio and video emotion recognition communities (extended with the physiological analysis community for the emotion recognition sub-challenge), to compare the relative merits of these approaches to emotion and depression recognition under well-defined and strictly comparable conditions and to establish to what extent fusion of the approaches is possible and beneficial. A second motivation is the need to advance emotion recognition for multimedia retrieval to a level where *behaviomedical* systems [7] are able to deal with large volumes of non-prototypical naturalistic behaviour in reaction to known stimuli, as this is exactly the type of data that diagnostic tools and other applications would have to face in the real world.

We called for participation in two Sub-Challenges: (i) fully-continuous emotion detection from audio, from video, and physiological data or from any combination thereof, and (ii) classification of depression. As benchmarking database an enhanced subset of the AVEC 2015 RECOLA database is used [4]. Emotion needed to be recognised in terms of continuous time, *and* continuous valued dimensional affect in two dimensions: valence and arousal. For depression, the Distress Analysis Interview Corpus - Wizard of Oz (DAIC-WOZ) database was used. This is part of a larger corpus, the Distress Analysis Interview Corpus (DAIC) [2], that contains clinical interviews designed to support the diagnosis of psychological distress conditions such as anxiety, depression, and post-traumatic stress disorder. Depression needed

to be estimated in terms of binary classes depressed/non-depressed, based on a widely accepted clinical self-report questionnaire, PHQ8.

Besides participation in the Challenge we called for papers addressing the overall topics of this workshop, in particular works that address the differences between audio and video processing of emotive data, and the issues concerning combined audio-visual emotion recognition.

In the following sections, we will describe the participation in this year and outline the conditions for participation in particular in the competitive challenge event. We further acknowledge those that helped realise AVEC 2016.

2. CHALLENGE CONDITIONS

A baseline paper explaining the dataset, the challenge evaluation procedure, baseline features and baseline results was made available while the challenge was ongoing [8]. Due to the sensitive nature of the data, a strict registration for access to the challenge data policy was followed. As in previous years, a web-based database was used, however, we required to sign a new end user license agreement and for the depression sub-challenge no raw video was made available, only video features. After downloading the data participants could directly start their own experiments with the train and development sets. Once they found their best method they should write a paper for the Workshop. At the same time they could compute their results per instance of the test set. Participants' results needed to be sent as a single packed file per Sub-Challenge to the organisers by email and scores were returned within 24 hours during typical working days.

Each participant had up to five submission attempts per Sub-Challenge. Badly formatted results were not counted towards one of these five submissions. The organisers provided for each affective dimension the Concordance Correlation Coefficient (CCC), which is used to rank participants on the emotion sub-challenge. For the depression sub-challenge the F1 measure was used to rank participants. The RMS error of the actual PHQ8 scores was provided as well, which could be used by the authors to further discuss their results in the paper accompanying their submission.

3. PARTICIPATION

The call for participation and papers attracted registrations of 56 teams from all over the world. 13 teams submitted results for the Affect recognition Sub-Challenge, and 7 teams submitted results for the Depression recognition Sub-Challenge. Finally, 12 paper submissions were received, which were assigned three reviewers, each, and reviewed independently. AVEC 2016 reviewing was double blind, and acceptance was based on relevance to the workshop, novelty, and technical quality. The programme committee accepted 12 papers in addition to the baseline paper as oral presentation. Again, we hope that these proceedings will serve as a valuable reference for researchers and developers in the area of audio-visual emotion recognition and depression analysis.

Acknowledgments

4. REFERENCES

[1] A. Dhall, R. Goecke, J. Joshi, M. Wagner, and T. Gedeon. Emotion Recognition In The Wild Challenge 2013. In *Proceedings ICMI*, pages 509–516, Sydney, Australia, December 2013. ACM.

[2] J. Gratch et al. The distress analysis interview corpus of human and computer interviews. In *Proc. of LREC*, pages 3123–3128, Reykjavik, Iceland, May 2014. ELRA.

[3] F. Ringeval, B. Schuller, M. Valstar, S. Jaiswal, E. Marchi, D. Lalanne, R. Cowie, and M. Pantic. Av+ec 2015: The first affect recognition challenge bridging across audio, video, and physiological data. In *Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge*, pages 3–8. ACM, 2015.

[4] F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In *Proc. of EmoSPACE, FG*, Shanghai, China, 2013.

[5] B. Schuller, M. Valstar, F. Eyben, R. Cowie, and M. Pantic. AVEC 2012 – The Continuous Audio/Visual Emotion Challenge. In *Proceedings ICMI*, pages 449–456, Santa Monica, CA, October 2012. ACM.

[6] B. Schuller, M. Valstar, F. Eyben, G. McKeown, R. Cowie, and M. Pantic. AVEC 2011 - The First International Audio/Visual Emotion Challenge. In *Proceedings ACII*, volume II, pages 415–424. Springer, Memphis, TN, October 2011.

[7] M. Valstar. Automatic behaviour understanding in medicine. In *Proceedings of the 2014 Workshop on Roadmapping the Future of Multimodal Interaction Research including Business Opportunities and Challenges*, pages 57–60. ACM, 2014.

[8] M. Valstar, J. Gratch, B. Schuller, F. Ringeval, D. Lalanne, M. Torres Torres, S. Scherer, G. Stratou, R. Cowie, and M. Pantic. Avec 2016 – depression, mood, and emotion recognition workshop and challenge. In *Proc. of AVEC, ACM MM*. ACM, 2016.

[9] M. Valstar, B. Schuller, J. Krajewski, R. Cowie, and M. Pantic. Workshop summary for the 3rd international audio/visual emotion challenge and workshop (AVEC'13). In *Proceedings MM*, pages 1085–1086, Barcelona, Spain, October 2013. ACM.

[10] M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie, and M. Pantic. Avec 2014: 3d dimensional affect and depression recognition challenge. In *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*, pages 3–10. ACM, 2014.