AVEC 2015 – The 5th International Audio/Visual Emotion Challenge and Workshop

Fabien Ringeval
University of Passau
Chair of Complex & Intelligent Systems
Passau, Germany

Björn Schuller*
University of Passau
Chair of Complex & Intelligent Systems
Passau, Germany

Michel Valstar
University of Nottingham
Mixed Reality Lab
Nottingham, UK

Roddy Cowie
Queen's University Belfast
School of Psychology
Belfast, UK

Maja Pantic†
Imperial College London
Intelligent Behaviour Understanding Group
London, UK

ABSTRACT

The fifth Audio-Visual Emotion Challenge and workshop AVEC 2015 was held in conjunction ACM Multimedia'15. Like the previous editions of AVEC, the workshop/challenge addresses the detection of affective signals represented in audio-visual data in terms of high-level continuous dimensions. A major novelty was further introduced this year by the inclusion of the physiological modality – along with the audio and the video modalities – in the dataset. In this summary, we mainly describe participation and its conditions.

Categories and Subject Descriptors

I [**Pattern Recognition**]: Applications

General Terms

Theory

Keywords

Affective Computing, Multimodality, Challenge

1. INTRODUCTION

This year's Audio-Visual Emotion Challenge and workshop (AVEC 2015) has been organised in conjunction with the 23rd ACM International Conference on Multimedia, held in Brisbane, Australia, 26 – 30 October 2015 (ACM-MM'15).

The AVEC 2015's theme is 'Uniting Audio, Video and Physiological Data' and it is the fifth competition event aimed at comparison of multimedia processing and machine learning methods for automatic audio, visual and – for the first time ever – physiological emotion analysis, with all participants competing under strictly the same conditions in this first of its kind series [6, 5, 8, 9]. However, further similar endeavours are to be noted in the meanwhile, e.g., [1, 7].

The goal of the Challenge is to provide a common benchmark test set for multimodal information processing and to bring together the audio, video and physiological emotion recognition communities, to compare the relative merits of the three approaches to emotion recognition under well-defined and strictly comparable conditions and establish to what extent fusion of the approaches is possible and beneficial. A second motivation is the need to advance emotion recognition systems to be able to deal with fully naturalistic behaviours in large volumes of un-segmented, non-prototypical and non-preselected data, as this is exactly the type of data that both multimedia and human-machine/human-robot communication interfaces have to face in the real world.

We were calling for teams to participate in a Challenge of fully-continuous emotion detection from audio, or video, or physiological data, or any combination of these three modalities. As benchmarking database the RECOLA multimodal corpus of remote and collaborative affective interactions was used [4]. Even though this database does not feature human-machine but rather human-human interaction, we strongly believe that the latter is the most interesting type of communication to study for the development of systems that will interact with humans, as we want such systems achieving realistic human like behaviours in the near future. Emotion needed to be recognised in terms of continuous time, continuous valued dimensional affect in two dimensions: arousal and valence.

Besides participation in the Challenge we were calling for papers addressing the overall topics of this workshop, in particular works that address the differences between audio, video and physiological processing of emotive data, and the issues concerning combined audio-visual-physiological 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 turn AVEC 2015 into reality.

*The author is further affiliated with Department of Computing, Imperial College London, London, UK.

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

MM'15, October 26–30, 2015, Brisbane, Australia
ACM 978-1-4503-3459-4/15/10.
http://dx.doi.org/10.1145/2733373.2806408.

2. PARTICIPATION

The call for participation and papers attracted registrations of 31 teams and 136 team members from all over the world. 13 teams submitted results for the emotion recognition challenge – overall, at the moment of writing over 38 submissions of results to the challenge were received. Finally, 15 paper submissions were received. AVEC 2015 reviewing was double blind, and acceptance was based on relevance to the workshop, novelty, and technical quality. The programme committee accepted 9 papers in addition to the baseline paper as oral presentation (the acceptance rate thus equalling 60%), which were assigned three reviewers, each, and reviewed independently. We hope that these proceedings will serve as a valuable reference for researchers and developers in the area of multimodal emotion recognition.

3. CHALLENGE CONDITIONS

A baseline paper explaining the dataset, the challenge evaluation procedure, baseline features and baseline results was made available during the ongoing challenge [3]. In order to ensure the correct use of the recordings provided in the RECOLA database, in compliance with the European and Swiss federal laws for data privacy and ethics, participants were required to sign an end user license agreement (EULA). Data were stored on a secured server (https) and a unique combination of user name and password was provided to each registered team, upon reception and validation of the completed EULA. 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 to the organisers by email and scores were returned within 24 hours during typical working days.

Each participant had up to five submission attempts for both arousal and valence. Badly formatted results were not counted towards one of these five submissions. Further, the top-two performers of the Challenge were asked to submit their program to the organisers at University of Passau to verify the results, both on the original test set and extra hold-out data. Delivery was as an executable or, e. g., encrypted Matlab or similar code, and they were asked to work with the organisers in validating their results. The organisers provided for each dimension (i. e., arousal and valence) and data partition (i. e., train and development) the root mean squared error (RMSE), the Pearson's correlation coefficient and the concordance correlation coefficient (CCC) [2], which is used to rank participants. Predictions returned on the individual sequences of a same data partition were concatenated before computing the correlation based coefficients. Ranking of the participants was obtained by averaging the best CCC achieved on arousal and valence, independently of their submission.

4. PROGRAM AND COMMITTEE

The workshop is a full-day event held on 26 October 2015 starting with a keynote speech by Roland Goecke of University of Canberra, followed by an introduction to the challenge, a series of paper presentations (oral), a demo session, a panel discussion and finally an overview of the challenge results and an announcement of the winners of the Challenge. The organisers – the authors of this summary – would like to thank all participants and in particular also the highly dedicated program committee of this year: *Felix Burkhardt*, Deutsche Telekom, Germany, *Rama Chellappa*, University of Maryland, USA, *Fang Chen*, NICTA, Australia, *Mohamed Chetouani*, UPMC, France, *Jeffrey Cohn*, University of Pittsburgh, USA, *Laurence Devillers*, LIMSI-CNRS, France, *Julien Epps*, UNSW, Australia, *Anna Esposito*, Second University of Naples, Italy, *Roland Goecke*, University of Canberra, Australia, *Jarek Krajewski*, University of Wuppertal, Germany, *Marc Méhu*, Webster Vienna Private University, Austria, *Louis-Philippe Morency*, USC, USA, *Stefan Scherer*, USC, USA, *Stefan Steidl*, University of Erlangen-Nuremberg, Germany, *Jianhua Tao*, Chinese Academy of Science, China, *Matthew Turk*, University of California, USA and *Stefanos Zafeiriou*, Imperial College London, UK.

Acknowledgments

The research leading to these results has received funding from the EC's Seventh Framework Programme through the ERC Starting Grant No. 338164 (iHEARu), and the EU's Horizon 2020 Programme through the Innovative Action No. 644632 (MixedEmotions), No. 645094 (SEWA) and the Research Innovative Action No. 645378 (ARIA-VALUSPA). The authors would further like to thank the sponsors of the challenge, the Association for the Advancement of Affective Computing (AAAC) and the audEERING UG. The responsibility lies with the authors.

5. REFERENCES

[1] A. Dhall, R. Goecke, J. Joshi, K. Sikka, and T. Gedeon. Emotion Recognition In The Wild Challenge 2014: Baseline, Data and Protocol. In *Proc. ICMI*, pages 461–466, Istanbul, Turkey, November 2014. ACM.

[2] L. Li. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45(1):255–268, March 1989.

[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 (AVEC), ACM MM*, Brisbane, Australia, October 2015.

[4] F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne. Introducing the RECOLA Multimodal Corpus of Remote Collaborative and Affective Interactions. In *Proc. FG*, Shanghai, China, April 2013.

[5] B. Schuller, M. Valstar, F. Eyben, R. Cowie, and M. Pantic. AVEC 2012 – The Continuous Audio/Visual Emotion Challenge. In *Proc. ICMI*, pages 449–456, Santa Monica (CA), USA, 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 *Proc. ACII 2011*, volume II, pages 415–424, Memphis (TN), USA, October 2011. Springer.

[7] M. Sjöberg, Y. Baveye, H. Wang, V. Quand, B. Ionescu, E. Dellandréa, M. Schedl, C.-H. Demarty, and L. Chen. The MediaEval 2015 Affective Impact of Movies Task. In *Proc. MediaEval 2015 Workshop*, Wurzen, Germany, 2015.

[8] 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 *Proc. MM*, pages 1085–1086, Barcelona, Spain, October 2013. ACM.

[9] M. Valstar, B. Schuller, J. Krajewski, R. Cowie, and M. Pantic. AVEC 2014: the 4th international audio/visual emotion challenge and workshop. In *Proc. MM*, pages 1243–1244, Orlando (FL), USA, November 2014. ACM.