Summary for AVEC 2018: Bipolar Disorder and Cross-Cultural Affect Recognition

Fabien Ringeval, Björn Schuller, Michel Valstar, Roddy Cowie, Maja Pantic

To cite this version:


HAL Id: hal-01994197
https://hal.archives-ouvertes.fr/hal-01994197
Submitted on 25 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Summary for AVEC 2018: Bipolar Disorder and Cross-Cultural Affect Recognition

Fabien Ringeval  
Université Grenoble Alpes, CNRS  
Grenoble, France  
fabien.ringeval@imag.fr

Björn Schuller†  
University of Augsburg  
Augsburg, Germany  
schuller@ieee.org

Michel Valstar  
University of Nottingham  
Nottingham, UK  
michel.valstar@nottingham.ac.uk

Roddy Cowie  
Queen’s University Belfast  
Belfast, UK

Maja Pantic†  
Imperial College London  
London, UK

ABSTRACT

The eighth Audio-Visual Emotion Challenge and workshop AVEC 2018 was held in conjunction with ACM Multimedia’18. This year, the AVEC series addressed major novelties with three distinct sub-challenges: bipolar disorder classification, cross-cultural dimensional emotion recognition, and emotional label generation from individual ratings. The Bipolar Disorder Sub-challenge was based on a novel dataset of structured interviews of patients suffering from bipolar disorder (BD corpus), the Cross-cultural Emotion Sub-challenge relied on an extension of the SEWA dataset, which includes human-human interactions recorded ‘in-the-wild’ for the German and the Hungarian cultures, and the Gold-standard Emotion Sub-challenge was based on the RECOLA dataset, which was previously used in the AVEC series for emotion recognition. In this summary, we mainly describe participation and conditions of the AVEC Challenge.

1 INTRODUCTION

This year’s Audio-Visual Emotion Challenge and workshop (AVEC 2018) has been organised in conjunction with the 26th ACM International Conference on Multimedia, MM 2018, held in Seoul, Korea, 22 – 26 October 2018.

The AVEC 2018’s theme is ‘Bipolar Disorder and Cross-Cultural Affect’ and it is the eighth competition event aimed at comparison of multimedia processing and machine learning methods for automatic audiovisual emotion and health analysis, with all of the participants competing under strictly the same conditions in this first of its kind series [4–6, 8–12]. However, further similar endeavours that have been undertaken since AVEC started in 2011 are to be noted [3].

As before, the goal of the Challenge is to compare the relative merits of the approaches for audiovisual recognition of emotion and health under well-defined and strictly comparable conditions, and establish to what extent fusion of the approaches is possible and beneficial. The main underlying motivation is the need to advance emotion recognition and health estimation, for multimedia retrieval to a level where behaviours expressed during human-human, or human-agent interactions, can be reliably sensed in real-life conditions, as this is exactly the type of data that the new generation of affect-oriented multimedia and human-machine/human-robot communication interfaces have to face in the real world.

We called for participation in three Sub-challenges: (i) Bipolar Disorder Sub-challenge (BDS): participants were required to classify audiovisual recordings of structured interviews of patients suffering from bipolar disorder – recorded periodically from the day of admittance to discharge – into remission, hypo-mania, or mania, as defined by the Young Mania Rating Scale [13]; (ii) Cross-cultural Emotion Sub-challenge (CES): participants were asked to predict the level of three emotional dimensions (arousal, valence, and liking) time-continuously from audiovisual recordings of dyadic interactions captured ‘in-the-wild’, i.e., recorded in various places such as home or work place, and with arbitrary personal equipment, in a cross-cultural setting with only labels of subjects with German culture available to infer the labels of subjects with Hungarian culture; (iii) Gold-standard Emotion Sub-challenge (GES): participants were required to generate a reliable gold-standard, i.e., a single time series preserving variance of emotion, from individual ratings of emotional dimensions (arousal and valence), that was then evaluated by a baseline multimodal (audio, video, physiology) emotion recognition system from recordings of dyadic interactions.
As benchmarking database, a novel dataset of bipolar disorder was used for the BDS, the BD corpus [2]. It includes audiovisual recordings of structured interviews performed by 47 Turkish speaking subjects aged 18-53. All those subjects suffered from bipolar disorder and were recruited from a mental health service hospital where they were diagnosed by clinicians following DSM-5’s inclusion criteria [1]. For the CES, an extended version of the Sentiment Analysis in the Wild (SEWA) database, with new data collected from 32 pairs of Hungarian participants aged 18-60+ in the same conditions as for the 32 others pairs of German participants with same age range, was used as a blind test set for the first ever cross-cultural (German → Hungarian) emotion recognition competition task. For the GES, we used the RECOLA dataset [7], which contains audiovisual and physiological recordings of dyadic interactions from 27 French speaking subjects aged 18-25.

Besides participation in the Challenge we called for papers addressing the overall topics of this workshop. In the following sections, we 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 2018.

2 CHALLENGE CONDITIONS

As in previous years, we required to sign an end user license agreement to access the data. After downloading the data, participants could directly start their own experiments with the train and development sets. In addition, standard feature sets were provided for audiovisual data, along with scripts available in a public repository², which participants were free to use for reproducing both the baseline features and recognition systems. Once they found their best method they had to 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.

3 PARTICIPATION

The call for participation and papers attracted registrations of 57 teams from all over the world, with 41, 44, and 45 teams participating in the BDS, CES, and GES, respectively. 11 teams submitted results for the BDS, 8 teams for the CES, and 4 teams for the GES. Finally, 23 paper submissions were received, which were assigned three reviewers, each, and reviewed independently. AVEC 2018 reviewing was double blind, and acceptance was based on relevance to the workshop, novelty, technical quality, and performance on the test partition. The program committee accepted 11 papers in addition to the independently reviewed 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 audiovisual emotion recognition and health analysis in real-life settings.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the Horizon 2020 Programme through the Innovative Action No. 645094 (SEWA), and the Research Innovative Action No. 645378 (ARIA-VALUSPA), and No. 688835 (DE-ENIGMA), and from the University of Fribourg, Switzerland, through the commercial licenses of the RECOLA project. The authors further thank the sponsors of the challenge – audEERING GmbH and the Association for the Advancement of Affective Computing (AAAC) – and the technical program committee members for the quality of their review.

REFERENCES


¹http://sewaproject.eu
²https://github.com/AudioVisualEmotionChallenge/AVEC2018