

Emotion and Sentiment Analysis

ABSTRACTS

Editors:

J. Fernando Sánchez-Rada, Björn Schuller

Workshop Programme

23 May 2016

9:00 – 9.10	Introduction by Workshop Chair
9.10 – 10:30	Social Media
Cristina Bosco et al.	Tweeting in the Debate about Catalan Elections
Ian D. Wood and Sebastian Ruder	Emoji as Emotion Tags for Tweets
Antoni Sobkowicz and Wojciech Stokowiec	Steam Review Dataset - new, large scale sentiment dataset
10:30 – 11:00	Coffee break
11:00 – 13:00	Corpora and Data Collection
Ebuka Ibeke et al.	A Curated Corpus for Sentiment-Topic Analysis
Jasy Liew Suet Yan and Howard R. Turtle	EmoCues-28: Extracting Words from Emotion Cues for a Fine-grained Emotion Lexicon
Lea Canales et al.	A Bootstrapping Technique to Annotate Emotional Corpora Automatically
Francis Bond et al.	A Multilingual Sentiment Corpus for Chinese, English and Japanese
13:00 – 14:00	Lunch break
14:00 – 15:00	Personality and User Modelling
Shivani Poddar et al.	PACMAN: Psycho and Computational Framework of an Individual (Man)
Veronika Vinczel, Klára Hegedűs, Gábor Berend and Richárd Farkas	Telltale Trips: Personality Traits in Travel Blogs
15:00 – 16:00	Linked Data and Semantics
Minsu Ko	Semantic Classification and Weight Matrices Derived from the Creation of Emotional Word Dictionary for Semantic Computing
J. Fernando Sánchez-Rada et al.	Towards a Common Linked Data Model for Sentiment and Emotion Analysis
16:00 – 16:30	Coffee break
16:30 – 18:00	Beyond Text Analysis
Bin Dong, Zixing Zhang and Björn Schuller	Empirical Mode Decomposition: A Data-Enrichment Perspective on Speech Emotion Recognition
Rebekah Wegener, Christian Kohlschein, Sabina Jeschke and Björn Schuller	Automatic Detection of Textual Triggers of Reader Emotion in Short Stories
Andrew Moore, Paul Rayson and Steven Young	Domain Adaptation using Stock Market Prices to Refine Sentiment Dictionaries

Workshop Organizers

J. Fernando Sánchez-Rada*	UPM, Spain
Björn Schuller *	Imperial College London, United Kingdom
Gabriela Vulcu	Insight Centre for Data Analytics, NUIG, Ireland
Carlos A. Iglesias	UPM, Spain
Paul Buitelaar	Insight Centre for Data Analytics, NUIG, Ireland
Laurence Devillers	LIMSI, France

Workshop Programme Committee

Elisabeth André	University of Augsburg, Germany
Noam Amir	Tel-Aviv University, Israel
Rodrigo Agerri	EHU, Spain
Cristina Bosco	University of Torino, Italy
Felix Burkhardt	Deutsche Telekom, Germany
Antonio Camurri	University of Genova, Italy
Montse Cuadros	VicomTech, Spain
Julien Epps	NICTA, Australia
Francesca Frontini	CNR, Italy
Diana Maynard	University of Sheffield, United Kingdom
Sapna Negi	Insight Centre for Data Analytics, NUIG, Ireland
Viviana Patti	University of Torino, Italy
Albert Salah	Boğaziçi University, Turkey
Jianhua Tao	CAS, P.R. China
Michel Valstar	University of Nottingham, United Kingdom
Benjamin Weiss	Technische Universität Berlin, Germany
Ian Wood	Insight Centre for Data Analytics, NUIG, Ireland

Preface/Introduction

ESA 2016 is the sixth edition of the highly successful series of Corpora for Research on Emotion. As its predecessors, the aim of this workshop is to connect the related fields around sentiment, emotion and social signals, exploring the state of the art in applications and resources. All this, with a special interest on multidisciplinary, multilingualism and multimodality. This workshop is a much needed effort to fight the scarcity of quality annotated resources for emotion and sentiment research, especially for different modalities and languages.

This year's edition once again puts an emphasis on common models and formats, as a standardization process would foster the creation of interoperable resources. In particular, researchers have been encouraged to share their experience with Linked Data representation of emotions and sentiment, or any other application of Linked Data in the field, such as enriching existing data or publishing corpora and lexica in the Linked Open Data cloud.

Approaches on semi-automated and collaborative labeling of large data archives are also of interest, such as by efficient combinations of active learning and crowdsourcing, in particular for combined annotations of emotion, sentiment, and social signals. Multi- and cross-corpus studies (transfer learning, standardisation, corpus quality assessment, etc.) are further highly relevant, given their importance in order to test the generalisation power of models.

The workshop is supported by the Linked Data Models for Emotion and Sentiment Analysis W3C Community Group ¹, the Association for the Advancement of Affective Computing ² and the SSPNet ³ – some of the members of the organizing committee of the present workshop are executive members of these bodies.

As organising committee of this workshop, we would like to thank the organisers of LREC 2016 for their tireless efforts and for accepting ESA as a satellite workshop. We also thank every single member of the programme committee for their support since the announcement of the workshop, and their hard work with the reviews and feedback. Last, but not least, we are thankful to the community for the overwhelming interest and number of high-quality submissions. This is yet another proof that the emotion and sentiment analysis community is thriving. Unfortunately, not all submitted works could be represented in the workshop.

J.F. Sánchez-Rada, B. Schuller, G. Vulcu, C. A. Iglesias, P. Buitelaar, L. Devillers

May 2016

¹<http://www.w3.org/community/sentiment/>

²<http://emotion-research.net/>

³<http://sspnet.eu/>

Social Media

Monday 23 May, 9:10 – 10:30

Chair: J. Fernando Sánchez-Rada

Tweeting in the Debate about Catalan Elections

Cristina Bosco, Mirko Lai, Viviana Patti, Francisco M. Rangel Pardo and Paolo Rosso

The paper introduces a new annotated Spanish and Catalan data set for Sentiment Analysis about the Catalan separatism and the related debate held in social media at the end of 2015. It focuses on the collection of data, where we dealt with the exploitation in the debate of two languages, i.e. Spanish and Catalan, and on the design of the annotation scheme, previously applied in the development of other corpora about political debates, which extends a polarity label set by making available tags for irony and semantic oriented labels. The annotation process is presented and the detected disagreement discussed.

Emoji as Emotion Tags for Tweets

Ian D. Wood and Sebastian Ruder

In many natural language processing tasks, supervised machine learning approaches have proved most effective, and substantial effort has been made into collecting and annotating corpora for building such models. Emotion detection from text is no exception; however, research in this area is in its relative infancy, and few emotion annotated corpora exist to date. A further issue regarding the development of emotion annotated corpora is the difficulty of the annotation task and resulting inconsistencies in human annotations. One approach to address these problems is to use self-annotated data, using explicit indications of emotions included by the author of the data in question. We present a study of the use of unicode emoji as self-annotation of a Twitter user's emotional state. Emoji are found to be used far more extensively than hash tags and we argue that they present a more faithful representation of a user's emotional state. A substantial set of tweets containing emotion indicative emoji are collected and a sample annotated for emotions. The accuracy and utility of emoji as emotion labels are evaluated directly (with respect to annotations) and through trained statistical models. Results are cautiously optimistic and suggest further study of emoji usage.

Steam Review Dataset - new, large scale sentiment dataset

Antoni Sobkowicz and Wojciech Stokowiec

In this paper we present new binary sentiment classification dataset containing over 3,640,386 reviews from Steam User Reviews, with detailed analysis of dataset properties and initial results of sentiment analysis on collected data.

Corpora and Data Collection

Monday 23 May, 11:00 – 13:00

Chair: Viviana Patti

A Curated Corpus for Sentiment-Topic Analysis

Ebuka Ibeke, Chenghua Lin, Chris Coe, Adam Wyner, Dong Liu, Mohamad Hardyman Barawi and Noor Fazilla Abd Yusof

There has been a rapid growth of research interest in natural language processing that seeks to better understand sentiment or opinion expressed in text. However, most research focus on developing new models for opinion mining, with little efforts being devoted to the development of curated datasets for training and evaluation of these models. This work provides a manually annotated corpus of customer reviews, which has two unique characteristics. First, the corpus captures sentiment and topic information at both the review and sentence levels. Second, it is time-variant, which preserves the sentiment and topic dynamic information of the reviews. The annotation process was performed in a two-stage approach by three independent annotators, achieving a substantial level of inter-annotator agreements. In another set of experiments, we performed supervised sentiment classification using our manual annotations as gold-standard. Experimental results show that both Naive Bayes model and Support Vector Machine achieved more than 92

EmoCues-28: Extracting Words from Emotion Cues for a Fine-grained Emotion Lexicon

Jasy Liew Suet Yan and Howard R. Turtle

This paper presents a fine-grained emotion lexicon (EmoCues-28) consisting of words associated with 28 emotion categories. Words in the lexicon are extracted from emotion cues (i.e., any segment of text including words and phrases that constitute expression of an emotion) identified by annotators from a corpus of 15,553 tweets (microblog posts on Twitter). In order to distinguish between emotion categories at this fine-grained level, we introduce cue term weight and describe an approach to determine the primary and secondary terms associated with each emotion category. The primary and secondary terms form the foundation of our emotion lexicon. These terms can function as seed words to enrich the vocabulary of each emotion category. The primary terms can be used to retrieve synonyms or other semantically related words associated with each emotion category while secondary terms can be used capture contextual cues surrounding these terms.

A Bootstrapping Technique to Annotate Emotional Corpora Automatically

Lea Canales, Carlo Strapparava, Ester Boldrini and Patricio Martínez-Barco

In computational linguistics, the increasing interest of the detection of emotional and personality profiles has given birth to the creation of resources that allow the detection of these profiles. This is due to the large number of applications that the detection of emotion states can have, such as in e-learning environment or suicide prevention. The development of resources for emotional profiles can help to improve emotion detection techniques such as supervised machine learning, where the development of annotated corpora is crucial.

Generally, these annotated corpora are performed by a manual annotation process, a tedious and time-consuming task. Thus, research on developing automatic annotation processes has increased. Due to this, in this paper we propose a bootstrapping process to label an emotional corpus automatically, employing NRC Word-Emotion Association Lexicon (Emolex) to create the seed and generalised similarity measures to increase the initial seed. In the evaluation, the emotional model and the agreement between automatic and manual annotations are assessed. The results confirm the soundness of the proposed approach for automatic annotation and hence the possibility to create stable resources such as, an emotional corpus that can be employed on supervised machine learning for emotion detection systems.

A Multilingual Sentiment Corpus for Chinese, English and Japanese

Francis Bond, Tomoko Ohkuma, Luís Morgado da Costa, Yasuhide Miura, Rachel Chen, Takayuki Kuribayashi and Wenjie Wang

In this paper, we present the sentiment tagging of a multi-lingual corpus. The goal is to investigate how different languages encode sentiment, and compare the results with those given by existing resources. The results of annotating a corpus for both concept level and chunk level sentiment are analyzed.

Personality and User Modelling

Monday 23 May, 11:00 – 13:00

Chair: Björn Schuller

PACMAN: Psycho and Computational Framework of an Individual (Man)

Shivani Poddar, Sindhu Kiranmai Ernala and Navjyoti Singh

Several models have tried to understand the formation of an individual's distinctive character i.e. personality from the perspectives of multiple disciplines, including cognitive science, affective neuroscience and psychology. While these models (for eg. Big Five) have so far attempted to summarize the personality of an individual as a uniform, static image, no one model comprehensively captures the mechanisms which leads to the formation and evolution of personality traits over time. This mechanism of evolving personality is what we attempt to capture by means of our framework. Through this study, we leverage the Abhidhamma tradition of Buddhism to propose a theoretical model of an individual as a stochastic finite state machine. The machine models moment to moment states of consciousness of an individual in terms of a formal ontology of mental factors that constitute any individual. To achieve an empirical evaluation of our framework, we use social media data to model a user's personality as an evolution of his/her mental states (by conducting some psycho-linguistic inferences of their Facebook (FB) statuses). We further analyze the user's personality as a composition of these recurrent mental factors over a series of subsequent moments. As the first attempt to solve the problem of evolving personality explicitly, we also present a new dataset and machine learning module for analysis of mental states of a user from his/her social media data.

Telltale Trips: Personality Traits in Travel Blogs

Veronika Vinczel, Klára Hegedűs, Gábor Berend and Richárd Farkas

Here we present a corpus that contains blog texts about traveling. The main focus of our research is the personality trait of the person hence we do not just annotate opinions in the classical sense but we also mark those phrases that refer to the personality type of the author. We illustrate the annotation principles with several examples and we calculate inter-annotator agreement rates. In the long run, our main goal is to employ personality data in a real-world application, e.g. a recommendation system.

Linked Data and Semantics

Monday 23 May, 14:00 – 16:00

Chair: Ian D. Wood

Semantic Classification and Weight Matrices Derived from the Creation of Emotional Word Dictionary for Semantic Computing

Minsu Ko

This paper introduces a general creation method for an emotional word dictionary (EWD) which contains a semantic weight matrix (SWM) and a semantic classification matrix (SCM) which will be used as an efficient foundation for opinion mining. These two matrices are combined into a single n by 7 matrix called as a classification and weight matrix (CWM) in a machine-processable format. Such a matrix would also have applications in the field of semantic computing. This paper also details investigations which were performed in order to gather information on the efficiency of using CWM based on categorizing synonymous relations and frequencies. The multilingual extensibility of the EWD will benefit semantic processing of opinion mining as a generic linguistic resource which has an emotional ontology structure and linked data.

Towards a Common Linked Data Model for Sentiment and Emotion Analysis

J. Fernando Sánchez-Rada, Björn Schuller, Viviana Patti, Paul Buitelaar, Gabriela Vulcu, Felix Burkhardt, Chloé Clavel, Michael Petychakis and Carlos A. Iglesias

The different formats to encode information currently in use in sentiment analysis and opinion mining are heterogeneous and often custom tailored to each application. Besides a number of existing standards, there are additionally still plenty of open challenges, such as representing sentiment and emotion in web services, integration of different models of emotions or linking to other data sources. In this paper, we motivate the switch to a linked data approach in sentiment and emotion analysis that would overcome these and other current limitations. This paper includes a review of the existing approaches and their limitations, an introduction of the elements that would make this change possible, and a discussion of the challenges behind that change.

Beyond Text Analysis

Monday 23 May, 16:30 – 18:00

Chair: J. Fernando Sánchez-Rada

Empirical Mode Decomposition: A Data-Enrichment Perspective on Speech Emotion Recognition

Bin Dong, Zixing Zhang and Björn Schuller

To deal with the data scarcity problem for Speech Emotion Recognition, a novel data enrichment perspective is proposed in this paper by applying Empirical Mode Decomposition (EMD) on the existing labelled speech samples. In doing this, each speech sample is decomposed into a set of Intrinsic Mode Functions (IMFs) plus a residue by EMD. After that, we extract features from the primary IMFs of the speech sample. Each single classification model is trained first for the corresponding IMF. Then, all the trained models of the IMFs plus that of the original speech are combined together to classify the emotion by majority vote. Four popular emotional speech corpora and three feature sets are used in an extensive evaluation of the recognition performance of our proposed novel method. The results show that, our method can improve the classification accuracy of the prediction of valence and arousal with different significance levels, as compared to the baseline.

Automatic Detection of Textual Triggers of Reader Emotion in Short Stories

Rebekah Wegener, Christian Kohlschein, Sabina Jeschke and Björn Schuller

This position paper outlines our experimental design and platform development for remote data collection, annotation and analysis. The experimental design captures reader response to 5 short stories, including reading time, eye and gaze tracking, pupil dilation, facial gestures, combined physiological measures, spoken reflection, comprehension and reflection. Data will be gathered on a total corpus of 250 short stories and over 700 readers. We describe both the experiment design and the platform that will allow us to remotely crowd-source both reader response and expert annotation, as well as the means to analyse and query the resulting data. In the paper we outline our proposed approach for gaze-text linkage for remote low-quality webcam input and the proposed approach to the capture and analysis of low arousal affect data. The final platform will be open-source and fully accessible. We also plan to release all acquired data to the affective computing research community.

Domain Adaptation using Stock Market Prices to Refine Sentiment Dictionaries

Andrew Moore, Paul Rayson and Steven Young

As part of a larger project where we are examining the relationship and influence of news and social media on stock price, here we investigate the potential links between the sentiment of news articles about companies and stock price change of those companies. We describe a method to adapt sentiment word lists based on news articles about specific companies, in our case downloaded from the Guardian. Our novel approach here is to adapt word lists in sentiment classifiers for news articles based on the relevant stock price change of a company at the time of web publication of the articles. This adaptable word list approach is compared against the financial lexicon from Loughran and McDonald (2011) as well as the more general MPQA word list (Wilson et al., 2005). Our experiments investigate the need for domain specific word lists and demonstrate how general word lists miss indicators of sentiment by not creating or adapting lists that come directly from news about the company. The companies in our experiments are BP, Royal Dutch Shell and Volkswagen.