EMOSEDE: Emotion and Sentiment Detection

HUSO 2016, The Second International Conference on Human and Social Analytics (IARIA conference)

Els Lefever
LT³, Language and Translation Technology Team
Ghent University
Ghent, Belgium
Els.lefever@ugent.be

***Abstract*—In the past decade, we have witnessed the rise of the Web 2.0 technologies, which allow the general public to add text to message boards, blogs, reviewing sites or social networking sites such as Twitter or Facebook. At the same time, research in the field of sentiment analysis has thrived, especially on social media data. Knowing what other people think has always played a major role in our decision making process. This holds for individuals looking for opinions about specific products or services, but to a larger extent even for companies and organizations who want to know the sentiments of the general public towards their products, brands, policies, political viewpoints, etcetera.**

In addition to sentiment analysis, which classifies text as positive or negative, researchers have started to automatically identify more fine-grained emotions in text, resulting in the related field of affect analysis or emotion detection. In crisis situations, for instance, companies are not only interested in detecting negative or positive messages, they also want to know whether people are angry, supportive, sad or disgusted in order to fine-tune their crisis communication strategy.

This special session will focus on topics related to sentiment and emotion analysis, stylometry and the influence of emotional states on natural language generation.

Keywords-sentiment analysis, emotion detection, personality detection, affective language, natural language processing

I. INTRODUCTION

The aim of the EMOSEDE 2016 special session is to bring together researchers in Natural Language Processing working on sentiment analysis and emotion detection from text.

The computational study of sentiment and its related concepts such as opinions, attitudes and beliefs is known as **automatic sentiment analysis** or opinion mining [1]. The first studies on automatic sentiment analysis focused on newswire text [2, 3, 4]. Interest in the field has flourished, however, with the expansion of the Internet and the birth of Web 2.0 applications. These include social networks like Facebook and Twitter, colloborative platforms (wikis, folksonomies), blogs, product review sites, etcetera. These social media generate a vast amount of opinionated data, which offers valuable insights into the public opinion. Applying sentiment analysis is an efficient way to process this huge amount of user-generated information by assigning a "positive" or "negative" label to chunks of texts.

More recently, the field of **emotion classification** has thrived. Instead of focusing on the polarity (positive, negative) of text, researchers have started to automatically extract specific emotions from text. Whereas the first studies (e.g. [5]) rely on Ekman's basic six emotions [6], viz. anger, disgust, joy, fear, sorrow, surprise, it has become clear that the Ekman classification does not account for all emotions expressed in social media posts. As a result, researchers have defined more fine-grained emotion schemes [7].

For this special session we focused on the following topics related to emotion and sentiment detection: resources and annotations for subjectivity, sentiment and emotion detection, emotional and personality profiles, data-driven and knowledge-based methods for sentiment and emotion analysis, aspect-based sentiment analysis methods and the relation between emotion and figurative language (irony, metaphor, parody). We were very happy to receive submissions addressing different topics related to sentiment and emotion detection.

II. SUBMISSIONS

The first paper is presented by Orphée De Clercq as a keynote talk and is entitled "The many aspects of fine-grained sentiment analysis. An overview of the task and its main challenges". As suggested by the title, this is a survey paper that presents the task of aspect-based sentiment analysis (ABSA) in great detail. The author defines and motivates the task, explains the three major components of the task, being the (1) aspect term extraction, (2) aspect term categorization and (3) aspect term polarity classification, and gives an extensive overview of the state of the art and currently available data sets. The last section of the paper reflects on the open issues for ABSA. Firstly, sentiment analysis systems are trained on domain-specific data sets, which makes them lose performance when applied to other domains. Secondly, user-generated content is very different from standard text. It contains a lot of orthographic variation (flooding, spelling mistakes, etc) and creative language use (irony, metaphors), explaining the drop in performance of standard text analysis tools when applied to social media data. This paper can certainly serve as a good introduction for researchers interested in the sentiment analysis task.

The second paper introduces a corpus analysis for the automatic detection of crisis emotions on social media: "Towards a framework for the automatic detection of crisis emotions on social media: a corpus analysis of the tweets posted after the crash of Germanwings Flight 9525" (Hoste et al.). The authors first highlight the importance of sentiment and emotion analysis of social media for effective crisis communication. Twitter, for instance, has become a very popular communication means for seeking and defusing crisis-related information. In the experimental part, the authors investigate to what extent automatic sentiment analysis techniques can be used for detecting crisis emotions on Twitter. They also provide an analysis of the manually labelled emotions in the corpus and find that *sympathy*, *anger* and *contempt* are the emotions that were most frequently expressed in the data. An important direction for future research lies in the classification of contextual information, such as the object and characteristics of the sender of the crisis emotion.

The third paper also presents research on automatic emotion detection: "Analysing Emotions in Social Media Coverage on Paris Terror Attacks: A Pilot Study" (Van Hee & Lefever). The paper introduces a pilot study in which machine learning techniques are applied to automatically detect emotions in Facebook posts after the Paris attacks. To evaluate the feasibility of emotion classification, the authors compiled and manually annotated a corpus of Dutch Facebook posts with the six basic emotions of Ekman. The experimental results show that emotion classification is not a trivial task and that the results suffer from data sparseness. In addition, six emotions appeared not to be sufficient; half of the annotated emotions could not be matched to any of Ekman's basic emotions. Another important insight is that the use of irony heavily impacts the sentiment information that is used in the emotion classification process.

"What does the bird say? Exploring the link between personality and language use in Dutch tweets" (Vandenhoven & De Clercq) touches upon the topic of computational stylometry for personality prediction. The authors compiled a database containing tweets of twenty Dutch-speaking users that also filled out a personality test. They analyzed the language use by means of syntactic analysis (part-of-speech tagging) and sentiment and personality-charged words. Despite the limited data set, interesting correlations were found between certain personality traits and language use. A statistical significant correlation was for instance detected between the personality trait *openness* and social terms (e.g. *family, friends*) as well as sensory and perceptual processes (e.g. *see, touch, listen*). These initial experiments clearly show the potential of using NLP techniques to define a personality profile of social media users.

The final paper in this session, "Producing Affective Language. Content Selection, Message Formulation, and Computational Modelling" (Goudbeek et al.), introduces a project aiming at investigating how emotional states influence language production. Two experiments are conducted: first the authors assessed whether disgusted speakers are more prone to align with their dialogue partners, second the authors developed a corpus of emotionally-laden soccer reports to investigate whether the reports differ depending on whether they come from the winning or losing team. The preliminary findings of this study will be used to build an affective natural language generation system.

III. Conclusion

The EMOSEDE special session includes a broad range of topics related to automated sentiment and emotion analysis. It contains both survey papers as well as pilot studies introducing interesting ideas for future work in this thriving research domain.

Acknowledgment

I would like to thank the organisers of HUSO 2016 for their tireless efforts and for accepting EMOSEDE as a special session. I also thank the members of the programme committee for their hard work with the reviews and feedback. Last, but not least, I am very thankful to the authors for their very interesting contributions.

References

[1] B. Pang, L. Lee, S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," Proceedings of the Conference on Empricial Methods in Natural Language Processing, pp. 79–86, 2002.

[2] S. Das, M. Chen, Yahoo! for amazon: Extracting market sentiment from stock message boards, in: In Asia Pacific Finance Association Annual Conf. (APFA), Vol. 35, 2001, p. 43.

[3] P. D. Turney, Thumbs Up or Thumbs Down?: Semantic Orientation Ap- plied to Unsupervised Classification of Reviews, Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02, Association for Computational Linguistics, Stroudsburg, PA, USA, 2002, pp. 417–424.

[4] J. Wiebe, T. Wilson, C. Cardie, Annotating expressions of opinions and emotions in language, Computer Intelligence 39 (2) (2005) 165–210.

[5] P. Ekman, "An argument for basic emotions," Cognition & Emotion, pp. 169–200, Vol. 6 (3–4), 1992.

[6] C. Strapparava, R. Mihalcea, "Learning to Identify Emotions in Text," Proceedings of the 2008 ACM symposium on Applied computing, pp. 1556–1560, 2008.

[7] Y. Jin, B.F. Liu, D. Anagondahalli, L. Austin, "Scale development for measuring publics' emotions in organizational crises," Public Relations Review, pp. 509–518, Vol. 40, 2014.