
Editorial

The abundance of opinions available on the World Wide Web represents an information repository of enormous intellectual and economic value. Automated methods to exploit this rich knowledge mine have become more and more relevant within the last decade and the availability of large amounts of data is an ideal premise for the application of empirical methods.

Although many researchers from different nations and institutes intensively work on the development of these techniques, many challenges have been left uncovered. The most pressing problems range from migrating sentiment analysis systems to new text types or domains, developing robust natural language applications that effectively exploit sentiment analysis, to the creation of resources that enable research in other languages than English. Moreover, a deeper understanding of subjective language beyond lexical keyword matching still needs to be acquired.

This special issue consists of a selection of papers presented at the 2nd Workshop on Practice and Theory of Opinion Mining and Sentiment Analysis (PATHOS) held in conjunction with GSCL-2013 in Darmstadt, Germany, on September 23rd, 2013. In order to ensure articles of a high quality, a second reviewing cycle was carried out on the revised submissions originally accepted and presented at the PATHOS workshop.

We briefly outline the topics addressed in those papers:

- Albertini et al. present an unsupervised method based on Growing Hierarchical Self-Organizing Maps to provide an alternative feature encoding. The aim of this encoding is to obtain a less sparse feature representation that typically arises with (traditional) bag of words applied on short documents. In the light of the growing importance of analyzing short texts from microblogging services, most prominently messages from Twitter, the task addressed by the authors is highly relevant to sentiment analysis. Their proposed encoding is evaluated against other competing methods (such as Autoencoders) and shown to outperform them.
- Another paper that focuses on learning-based methods is Marchand et al. who examine multi-polarity words, i.e. polar expressions that change their polarity across different domains. As the set of domains on which sentiment analysis can be applied is pretty large, learning-based approaches often face the problem that only labeled out-of-domain training data are available. Marchand et al. show that the deletion of multi-polarity words substantially improves classification performance when such training data are used and propose a method to detect such words. They assume a realistic setting in which no labeled information from the target domain is available.

- Ruppenhofer et al. describe the shared task on source and target extraction from political speeches which is to be organized in summer 2014. This article makes a welcome contribution to JLCL, being the flagship journal for research in German speaking countries, since it describes the first shared task that is exclusively concerned with sentiment analysis in German.
- Another work that focuses on sentiment analysis on a language other than English is presented by Veselovska et al. who introduce a subjectivity lexicon for Czech. The work describes the creation of the resource and its evaluation on polarity classification in four different domains and is an important example of resource creation for Czech.
- Degaetano-Ortlieb et al. report on a study in a rather different direction. It is a descriptive approach on sentiment analysis whose purpose is to uncover evaluative expressions with a focus on the notion of "importance" in the genre of scientific research articles. The study is carried out on a specially annotated corpus that allows an examination of very complex linguistic properties. We believe that such research enables a deeper understanding of sentiment and subjective language than can be gained by the predominant resources, such as textual corpora labeled for polarity and sentiment lexicons.
- The last article of this issue comes in a similar vein. Gu et al. present an exploratory study of using electroencephalography (EEG) for the prediction of lexical valence. This is a highly interdisciplinary work as it departs from traditional unimodal approaches of sentiment analysis that exclusively draw information for prediction from text. This work is an example of the emerging research area of multimodal analysis that has recently attracted wide attention in sentiment analysis.

We would like to thank the “Gesellschaft für Sprachtechnologie und Computerlinguistik” for accepting this volume to be published in the JLCL series. Special thanks also go to Thierry Declerck, the editor-in-chief of the journal, for supporting us with technical issues in creating this special issue. Moreover, we are indebted to our reviewers for their hard work. Last but not least, we thank the authors of the articles for their interesting contributions.

August 2014

Stefan Gindl, Robert Remus, Michael Wiegand