Good News vs. Bad News: What are they talking about?

Olga Kanishcheva
Intelligent Computer Systems Department, National Technical University "KhPI"
kanichshevaolga@gmail.com
Victoria Bobichev
Department of Informatics and Systems Engineering
Technical University of Moldova
victoria.bobicev@ia.utm.md

Abstract

Today's massive news streams demand the automate analysis which is provided by various online news explorers. However, most of them do not provide sentiment analysis. The main problem of sentiment analysis of news is the differences between the writers and readers attitudes to the news text. News can be good or bad but have to be delivered in neutral words as pure facts. Although there are applications for sentiment analysis of news, the task of news analysis is still a very actual problem because the latest news impacts people's lives daily.

In this paper, we explored the problem of sentiment analysis for Ukrainian and Russian news, developed a corpus of Ukrainian and Russian news and annotated each text using one of three categories: positive, negative and neutral. Each text was marked by at least three independent annotators via the web interface, the inter-annotator agreement was analyzed and the final label for each text was computed. These texts were used in the machine learning experiments. Further, we investigated what kinds of named entities such as Locations, Organizations, Persons are perceived as good or bad by the readers and which of them were the cause for text annotation ambiguity.

1 Introduction

In news, sentiments are conveyed in a subtle manner without the use of explicit sentiment bearing words, and their detection requires contextual knowledge.

Most research conducted in the field of sentiment analysis has been done for the English language (Zhang and S. Skiena (2010), Liu B. and Zhang L. (2012)), some more for European Union languages whereas far less research has been completed for East European languages including Ukrainian and Russian.

Sentiment analysis of user generated content has been the focus of many researches; however, mass media news articles deserve the attention of these researches as well. Sentiment analysis of the news helps make their image more transparent, as possible biases in different news sources can be uncovered. The lack of research in this field was the motivation for the current work.

In comparison with the other domains, like product reviews, sentiment polarized words are used less frequently and sentiments are conveyed by complex structures and contextual knowledge as a way for journalists to seem more objective than they actually are. Many newspapers at least want to give an impression of objectivity and journalists desist from using obviously positive or negative vocabulary. This causes the sentiment classification task to become very challenging as we need to find domain-specific methods to handle this complexity.

The paper is organized as follows: in the next Section we describe related work. In Section 3 we described the corpus, its annotation and inter-rater agreement. In Section 4 we present the sentiment analysis experiments. Section 5 interprets the results; this part contains conclusions and future work.

2 Related Work

In recent years, sentiment analysis has been developing faster; it is connected with the growth of online texts and social networks. Some surveys about this area were presented in works Zhang and Skiena (2010), Liu and Zhang (2012).

Proceedings of Recent Advances in Natural Language Processing, pages 325–333,
Varna, Bulgaria, Sep 4–6 2017.
https://doi.org/10.26615/978-954-452-049-6_044