Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Sentiment analysis system adaptation for multilingual processing: The case of tweets

Alexandra Balahur [*], José M. Perea-Ortega

European Commission, Joint Research Center (JRC), Institute for the Protection and Security of the Citizen (IPSC), Via E. Fermi 2749, T.P. 267, I-21027 Ispra, VA, Italy

ARTICLE INFO

Article history:
Received 6 December 2013
Received in revised form 1 October 2014
Accepted 8 October 2014
Available online xxxx

Keywords:
Subjectivity analysis
Sentiment analysis
Multilingual resources
Social media mining
Chat analysis

ABSTRACT

Nowadays opinion mining systems play a strategic role in different areas such as Marketing, Decision Support Systems or Policy Support. Since the arrival of the Web 2.0, more and more textual documents containing information that express opinions or comments in different languages are available. Given the proven importance of such documents, the use of effective multilingual opinion mining systems has become of high importance to different fields. This paper presents the experiments carried out with the objective to develop a multilingual sentiment analysis system. We present initial evaluations of methods and resources performed in two international evaluation campaigns for English and for Spanish. After our participation in both competitions, additional experiments were carried out with the aim of improving the performance of both Spanish and English systems by using multilingual machine-translated data. Based on our evaluations, we show that the use of hybrid features and multilingual, machine-translated data (even from other languages) can help to better distinguish relevant features for sentiment classification and thus increase the precision of sentiment analysis systems.

1. Introduction

In the past decade, the quantity of user-generated contents on the Internet has been growing exponentially. Social Media platforms, such as Facebook, Twitter, Flickr and LinkedIn, as well as commercial sites, like Amazon and Booking.com offer their users the possibility to share their experiences and opinions on topics ranging from economics, to politics, products, VIPs and globally-critical events. The value of such unbiased, real-time user-generated content has been shown to be tremendous, with applications in Marketing, Decision Support Systems, Politics and Public Policy support, disaster and crisis management, etc. Since the high volume of opinionated information makes its manual processing virtually impossible, systems have been developed to treat texts and process the opinions they contain automatically, in the context of the Subjectivity and Sentiment Analysis tasks, within the field of Natural Language Processing (NLP).

Subjectivity and Sentiment Analysis typically aim at detecting subjective, "private" states (i.e. opinions, emotions, sentiments, evaluations, beliefs, and speculations) in texts (Pang & Lee, 2008; Pang, Lee, & Vaithyanathan, 2002; Wiebe, 2000) and subsequently classifying them according to their polarity.

A lot of research has concentrated on developing methods for subjectivity and sentiment analysis in different types of texts and with different applications in mind. Nonetheless, the majority of the research has concentrated on texts written

[*] Corresponding author.
E-mail addresses: alexandra.balahur@jrc.ec.europa.eu (A. Balahur), jose-manuel.perea-ortega@jrc.ec.europa.eu (J.M. Perea-Ortega).

http://dx.doi.org/10.1016/j.ipm.2014.10.004
0306-4573/© 2014 Elsevier Ltd. All rights reserved.

in English, since it is the language with most processing tools and annotated resources. While for some applications analyzing only opinions written in English is enough, for others, such as news monitoring, being able to detect and comparatively analyze opinions expressed in different sources is an important requirement.

This paper presents the experiments carried out for the *Sentiment Analysis in Tweets* task of SemEval 2013[1] (Balahur, 2013) and the modifications performed to the English system for the participation in the TASS[2] competition (Balahur & Perea-Ortega, 2013). SemEval 2013 Task 2 was entitled "Sentiment Analysis in Tweets" and is focused on English. TASS is an experimental evaluation workshop for sentiment analysis and online reputation management systems developed with a focus on Spanish. We present the approaches followed in these competitions and their adaptation to new languages, as well as additional experiments carried out to improve the results obtained.

In the SemEval 2013 task 2, participating systems had to classify snippets from or entire tweets according to their polarity – into positive, negative and neutral (or objective). In the TASS 2013 edition, we only participated in the first task, entitled *"sentiment analysis at global level"*. In this task, the participants were asked to assess the global polarity of short texts extracted from Twitter by using 5 levels of sentiment (very positive, positive, neutral, negative and very negative), plus discriminate them from the objective ones. To tackle this task, we applied an approach based on machine learning by trying different feature combinations, using dictionary-based features and adding external data for training obtained through machine translation. The main motivation for the experiments in the TASS competition was to evaluate the manner in which our approach (applied for English and combinations of data from different languages) could perform for Spanish. It followed previous efforts to test the viability of using machine-translated data for sentiment analysis in newspaper articles (Balahur & Turchi, 2014, 2012) and tweets (Balahur & Turchi, 2013) and showed promising results. The evaluations in the two competitions and in the subsequent experiments show that the use of supervised learning with additional dictionary features and external training data obtained from machine translated texts might be considered good strategies to generate learning data for polarity classification systems.

The rest of the paper is organized as follows: the following section deals with the state of the art in sentiment analysis. The main features of the proposed approaches are presented in Section 3. Section 4 describes the data used for learning while the different experiments carried out are detailed in Section 6. Finally, the results obtained and the conclusions are discussed in Section 7 and Section 8, respectively.

2. State of the art

The work presented herein is related to research in NLP on short informal text classification and multilingual text classification.

As regards short informal text classification, Go, Bhayani, and Huang (2009) performed one of the first studies involving sentiment analysis applied to tweets. The authors introduced emoticons (e.g. ":)", ": (", etc.) as markers of positive and negative tweets. Following their initial findings, Read (2005) employed the method to generate a corpus of sentiment-annotated tweets. They considered that positive tweets were the ones containing positive emoticons (e.g. " :)"), and negative tweets were the ones with negative emoticons (e.g. " : ("). In their subsequent experiments, they introduce different supervised approaches (SVM, Naïve Bayes and Maximum Entropy) and various sets of features and conclude that the simple use of unigrams leads to good results, but it can be slightly improved by the combination of unigrams and bigrams.

In the same line of thinking, Pak and Paroubek (2010) also generated a corpus of tweets for sentiment analysis, by selecting positive and negative tweets based on the presence of specific emoticons. Subsequently, they perform different experiments to classify sentiment in the obtained corpus and conclude that the best settings include the use of a Naïve Bayes classifier with unigrams and part-of-speech tags.

Another approach on sentiment analysis in tweets is that of Zhang, Ghosh, Dekhil, Hsu, and Liu (2011). Here, the authors adopt a hybrid approach, combining supervised learning with the knowledge on sentiment-bearing words, which they extract from the DAL sentiment dictionary (Whissell, 1989). Their pre-processing stage includes the removal of retweets, translation of abbreviations into original terms and deleting of links, a tokenization process, and part-of-speech tagging. They employ various supervised learning algorithms to classify tweets into positive and negative, using *n*-gram features with SVM and syntactic features with Partial Tree Kernels, combined with the knowledge on the polarity of the words appearing in the tweets. The authors conclude that the most important features are those corresponding to sentiment-bearing words. Finally, Jiang, Yu, Zhou, Liu, and Zhao (2011) classify sentiment expressed on previously-given "targets" in tweets. They add information on the context of the tweet to its text (e.g. the event that it is related to). Subsequently, they use SVM and General Inquirer and perform a three-way classification (positive, negative, neutral).

In SemEval 2013, a task was organized on sentiment analysis in tweets (Wilson et al., 2013). Here, the best-performing systems used additional dictionaries that were built from large data sets and word-emotion association dictionaries built from millions of tweets. From here, we can see that the use of dictionaries to improve the features used in supervised learning is a good strategy.

[1] http://www.cs.york.ac.uk/semeval-2013/task2/.

[2] Workshop on Sentiment Analysis at SEPLN, http://www.daedalus.es/TASS2013.

Please cite this article in press as: Balahur, A., & Perea-Ortega, J. M. Sentiment analysis system adaptation for multilingual processing: The case of tweets. *Information Processing and Management* (2014), http://dx.doi.org/10.1016/j.ipm.2014.10.004