

Title	マイクロブログにおける皮肉表現を対象とした感情分析
Author(s)	TUNGTHAMTHITI, PIYOROS
Citation	
Issue Date	2016-09
Type	Thesis or Dissertation
Text version	ETD
URL	<a href="http://hdl.handle.net/10119/13826">http://hdl.handle.net/10119/13826</a>
Rights	
Description	Supervisor:白井 清昭, 情報科学研究科, 博士

Sentiment analysis of sarcasm in microblogging is important in a range of natural language processing (NLP) applications such as text mining and opinion mining. However, this is a challenging task, as the real meaning of a sarcastic sentence is the opposite of the literal meaning. Furthermore, microblogging messages are short and usually written in a free style that may include misspellings, grammatical errors, and complex sentence structures. This thesis proposes a novel method of sentiment analysis on microblogging that enables us to identify orientation and intensity of the sentiment expressed in the tweets, especially in the sarcastic tweets.

First, we introduce a novel method to identify sarcasm in tweets. It is an ensemble of two supervised classifiers: one is Support Vector Machine (SVM) with N-gram features, the other is SVM with our proposed features. Our features represent intensity of sentiment and contradiction of sentiment derived by a naive sentiment analysis of the tweet. In the sentiment contradiction feature, coherence among multiple sentences in the tweet is also considered, which is automatically identified by our proposed method based on unsupervised clustering algorithm. Furthermore, a way to expand concepts of unknown sentiment words is presented to compensate for insufficiency of a sentiment lexicon. Our method also considers punctuation and special symbols, which are frequently used in Twitter. Results of experiments using two datasets show that our proposed system outperforms baseline systems. The accuracy of sarcasm identification on two datasets is 83% or 76%.

Next, we propose a sentiment analysis system designed for handling sarcastic tweets. To train the model to guess the polarity and intensity of the sentiment in the sarcastic tweets, we used a rich set of features, that are our proposed features used for sarcasm recognition as well as the features grounded on several linguistic levels proposed by the previous work. A decision tree with these features is trained to classify the tweets into an 11-scale score in range of -5 to +5. The system is evaluated on the dataset released by the organizers of the SemEval 2015 task 11. The results show that our method largely outperforms the systems proposed by the participants of the task on sarcastic and ironic tweets.

Finally, we propose a method for developing a sentiment analysis tool that can guess the fine-grained sentiment score for various types of the tweets. The system consists of two steps. At the first step, the given tweets are classified if they are sarcastic by our sophisticated sarcasm recognition method. At the second step, our sentiment analysis system designed for the sarcastic tweets is used to guess the sentiment scores of the tweets that are judged as sarcasm in the first step. On the other hand, for the tweets judged as non-sarcasm, the three existing sentiment analyzers are applied to guess the sentiment score. The results of the experiments show that our proposed two-steps sentiment analysis system outperforms any single sentiment analyzers on a data set consisting of both sarcastic and non-sarcastic tweets.

In addition, as for the application of the proposed method, our technique to recognize the sarcasm is integrated to an existing target-dependent sentiment analysis system. We also show that the integration can improve the performance via the experiments using a relatively small data set consisting of three targets.

**Keywords:** Sarcasm, Microblogging, Sentiment analysis, Coherence, Concept knowledge, Machine learning, Clustering