
Sentiment analysis in the Twitter stream

Alina Sinelnikova, Eirini Ntoutsis, and Hans-Peter Kriegel¹

Institute for Informatics, Ludwig-Maximilians University (LMU), Munich
sinelnikovaaa@gmail.com, {ntoutsis, kriegel}@dbs.ifi.lmu.de

Abstract. Nowadays, more and more people publish their opinions online so everyone has the possibility to catch up the thoughts of millions of people without even knowing them. This way, consumers have the enormous power to influence each other by sharing their brand experiences, either positive or negative. Twitter is the most famous micro-blogging service and an opinion-rich resource that allows people to broadcast their opinions about politics, products, movies etc in real time. With 200 million tweets generated on a daily basis, there is a need for opinion mining and sentiment analysis in order to help business analysts in the decision making process.

In this work, we deal with the challenges posed by the Twitter stream, namely size, unbalanced classes, changing class distributions, as well as with the specific limitations of the Twitter language, namely, colloquial language, tweet length and the difficult nature of the sentiment analysis problem due to the subjectivity of the tweets. For the study, we use a dataset of predefined topics from the Twitter API monitored over a period of three months. We experimented with a variety of classifiers such as Multinomial Naive Bayes, Adaptive Hoeffding Tree, Stochastic Gradient Descent, a hybrid Hoeffding Tree and Naive Bayes classifier and ensembles of classifiers. For the evaluation, we used both holdout and prequential methods. As a forgetting mechanism we used a sliding window. We evaluated the different methods and also the impact of the different preprocessing steps. We implemented a sentiment analysis tool that connects our methods to the Twitter API and identifies and monitors the changes in the sentiment distribution of the current opinions regarding some user defined topic.

References

- PANG, B. and LEE, L. (2008): Opinion Mining and Sentiment Analysis. *Found. Trends Inf. Retr.*, 2 1554-0669.
- BIFET, A. and Holmes, G. and Pfahringer, B.(2011): MOA-TweetReader: Real-Time Analysis in Twitter Streaming Data. In: T. Elomaa, J. Hollmén and H. Mannila (Eds.): *Discovery Science*. Springer, 46–60.

Keywords

DATA STREAMS, SENTIMENT ANALYSIS, TWITTER.