

COMPARING METHODS FOR TWITTER SENTIMENT ANALYSIS

This work extends the set of works which deal with the popular problem of sentiment analysis in Twitter. It investigates the most popular document ("tweet") representation methods which feed sentiment evaluation mechanisms. In particular, we study the bag-of-words, n-grams and n-gram graphs approaches and for each of them we evaluate the performance of a lexicon-based and 7 learning-based classification algorithms (namely SVM, Naïve Bayesian Networks, Logistic Regression, Multilayer Perceptrons, Best-First Trees, Functional Trees and C4.5) as well as their combinations, using a set of 4451 manually annotated tweets. The results demonstrate the superiority of learning-based methods and in particular of n-gram graphs approaches for predicting the sentiment of tweets. They also show that the combinatory approach has impressive effects on n-grams, raising the confidence up to 83.15% on the 5-Grams, using majority vote and a balanced dataset (equal number of positive, negative and neutral tweets for training). In the n-gram graph cases the improvement was small to none, reaching 94.52% on the 4-gram graphs, using Orthodromic distance and a threshold of 0.001.

Psomakelis E., Tserpes K., Anagnostopoulos D. and Varvarigou T. (2014), "**Comparing Methods for Twitter Sentiment Analysis**", In Proceedings of the International Conference on Knowledge Discovery and Information Retrieval (IC3K 2014), ISBN 978-989-758-048-2, pp. 225-232. DOI: [10.5220/0005075302250232](https://doi.org/10.5220/0005075302250232)