Sentiment Classification for Product Reviews using Ensemble Models

Panagiotis Diakos, Katerina Kapagiannidi, Arsinta Rantza
Department of Marketing and Communication, School of Business, Athens University of Economics and Business, Patision 81, Athens 11362, Greece

Panagiotis P. Repoussis
Department of Marketing and Communication, School of Business, Athens University of Economics and Business, Patision 81, Athens 11362, Greece and
School of Business, Stevens Institute of Technology, Castle Point on Hudson, Hoboken, NJ 07030, USA
{ prepousi@aueb.gr; prepouss@stevens.edu }

Abstract

This work develops and applies ensemble models for sentiment analysis of product reviews. In statistics and machine learning, ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone [1]. Sentiment analysis is a natural language processing task that deals with determining the sentiment orientation of a review [2]. The goal of this work is to train and test simple and more advanced models using different classification algorithms as well as to compare the resulting accuracy. The test bed for our experiments is a dataset containing 2000 single product reviews crawled from the Amazon.com website. Each review contains multiple sentences and it is classified with 1 (positive) or 0 (negative) label. All data mining tools used to conduct this research are provided by RapidMiner.

Text mining is initially performed and a term-by-document matrix is generated by converting all text to lowercase; tokenizing; removing stopwords and tokens with fewer than three characters; stemming; and finally removing terms that only occurred once in the data set. Next, each sentence is parsed into *n*-gram candidate phrases, where *n* ranged from one to five. The importance of each term is measured based on the so-called TF-IDF (i.e., term frequency-inverse document frequency), while Singular Value Decomposition is also performed to reduce the dimensionality of the matrix (assuming a 0.95 cumulative variance threshold).

Having selected the features, the model building phase is triggered. At first, various classifiers are applied, including Decision Trees, k-Nearest Neighbor, Support Vector Machines, Logistic Regression, Naïve Bayes and Deep Learning. For each classifier, multiple combination of parameters are compared using a 10-fold stratified cross-validation, and the best performing parameter set is recorded. Using the same settings, we examine the performance of Decision Tree based ensemble models using the Bayesian Boosting, the AbaBoost and the Bagging operators. Lastly, voting and learning-based stacking ensemble models are evaluated considering multiple combinations of classifiers. In conclusion, ensemble models seem very powerful for predicting sentiment polarity; however, it is worth to highlight that simple models often yield similar accuracy and they are significantly much faster.

Keywords
Opinion Mining; Sentiment Analysis; Text Mining; Natural Language Processing; Deep Learning; Decision Trees; Naïve Bayes; Support Vector Machine; Logistic Regression; k-NN; Ensemble Models

References

[1] Rokach, L. (2010). "Ensemble-based classifiers". Artificial Intelligence Review. 33 (1-2): 1–39.
[2] Pang B. and Lee L. (2008), "Opinion Mining and Sentiment Analysis", Foundations and Trends in Information Retrieval: Vol. 2: No. 1–2, pp 1-135.