"The Worst Spam-Blender I've Ever Bought!": Leveraging Natural Language Processing On Sentiment Analysis of Product Reviews

Paul Elliott
University of Oregon
paule@cs.uoregon.edu

Kanika Sood
University of Oregon
kanikas@cs.uoregon.edu

I. PROPOSAL

With the rise of online commercial marketplaces such as Amazon.com, user-generated product reviews have become increasingly more influential sources of information. User reviews affect more than the purchasing decisions of other consumers; such content also informs decisions in product marketing, sales, and so on. In particular, if consumer sentiments– positive or negative views, attitudes, emotions, appraisals about an entity–could be extracted from a corpus of reviews, such information could be leveraged toward improving future sales.

The difficulty with performing sentiment analysis on user-generated reviews is somewhat self-explanatory– human-generated text is nontrivial to parse by machines. Grammar and syntactic issues aside, consumers post reviews from a wide range of contexts. How will a machine know the difference between the glowing reviews of the Acme Spam-Blender by Farmer John, a Caucasian middle-aged farmer in western Kansas, versus that of Pooja Verma, a urban high school student in Delhi? Such intricacies require a specialized process by which sentiment analysis is achieved.

Our project involves leveraging NLTK[3], a natural language processing toolkit, for the purpose of sentiment analysis on user-generated product reviews. Our goal is to determine whether we can accurately predict the user ratings and/or sentiments from a corpus of reviews using a variety of machine learning algorithms. Using NLTK, we will extract a set of features from the reviews and use those features in the learning tasks. Secondly, we are interested in investigating which algorithm performs best given the nature of the data.

The dataset that we are using for this project is a corpus of Amazon.com user reviews from four different product domains:

1) books,
2) DVDs,
3) electronics, and
4) kitchen appliances.

This dataset was made available by John Blitzer et al of the University of Pennsylvania[2], who performed some of the seminal work in this area.

While we still need to investigate the feature space, possibilities include:

1) N-grams,
2) parts of speech,
3) sentiment lexicons (grouping of words in emotion and content categories), and
4) synomym-sets (sets of synonyms for words in the review).

The algorithms that we will employ include Naive Bayes, decision trees, and logistic regression. We will leverage WEKA[1] data mining software, which contains implementations of these algorithms.

We will measure this sentiment analysis task by the accuracy of the predictions made by our algorithms. This could take multiple forms: for example, we should determine how accurately our algorithms predicts the number of stars in a given review, as well as how well they predict favorable versus critical categories (defined by Amazon as 4 to 5 stars, and 1 to 3 stars respectively). We could also extend this analysis by measuring our accuracy on polarized reviews (1 or 5 stars), as well as how well our algorithms could predict the most helpful favorable or critical reviews.

REFERENCES

[1] M. L. G. at the University of Waikato. Weka 3: Data mining software in java, 2012.
[2] J. Blitzer, M. Dredze, and F. Pereira. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Annual Meeting-Association For Computational Linguistics*, volume 45, page 440, 2007.
[3] N. Project. Natural language toolkit, 2012.