ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 11, November 2017

System and a Method for Customer Service Survey Sentiment Classification and Neuro Analysis with Report Mining

Deekshith Raj Rathod[1]

Graduate Student, Department of Computer Science,California State University Sacramento, California, USA[1]

ABSTRACT: With the growing availability and popularity of opinion-rich resources such as online review sites,customer feedback surveys, and personal blogs,new opportunities and challenges arise as people can, and do, actively use information technologies to seek out and understand the opinions of others. The sudden eruption of activity in opinion mining and sentiment analysis, which deals with the computational treatment of opinion, sentiment, and subjectivity in text, has thus occurred at least in part as a direct response to the surge of interest in new systems that deal directly with opinions as a first-class object.

To help people digest and exploit the opinion information, I have selected Surveys from a School Facility department and performed varied analysis to classify polarity of text or comments from the survey through an exported excel data set. The system also consists of least significant and most significant options for user. It has Naïve BayesClassifier for predicting the trend of the customer satisfaction Rating from the review sentiment. The Implementation will help classify text or user feedback into three categories, Positive, negative and Neutral. The report will also specify details on the procedure and methodology which is used, for example, the different classification techniques and why a specific classification technique has been used and the contrasting results with a different classification library and present the results and prediction with analysis.

KEYWORDS: Data Mining, Sentiment Analysis, Term Frequency (TF), Inverse Document Frequency(IDF),Naïve Bayes Classification, Vectorizer.

I. INTRODUCTION

Sentiment Analysis - It is a classification of sentiment (Positive, Negative, Neutral) into Document-level, Sentence-level and feature or entity level, I have implemented Document-level analysis where it considers the whole document (Facilities Survey exported Excel from Survey Monkey).

The Data set used for Sentiment analysis is important to be known to understand its purpose as the analysis can help and provide important results for companies or entities to achieve a better context for reviewing or classifying the service/ product provided. The Implementation would include extracting relevant text and comments from the enormously data filled excel and create a training and testing data set and display results using relevant graphs.

This is an ever-improving field with new applications and enhancements achieved over the years and continuous improvement. There is a wide variety of applications of Sentiment analysis than just basic classification like Emotion detection, Opinion Mining.

The data set was chosen from a survey carried out via Survey Monkey and the surveys have been carried out from as early as 2006 to present day. The basic idea behind exploring this data set was to explore the comments which were provided were more of real world, raw data uncleaned and unprocessed which will allow to explore on training for anomalous textexample, spelling mistake, negative and positive text in a survey comment, negation issues like Not Happy, Not Satisfied and also sarcasm etc.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 11, November 2017

The main motivation behind exploring sentiment analysis with Facilities survey is that,
A lot of times a survey may seem to be imbalanced as end users may give ratings but that cannot be the sole metrics to evaluate. Generally, in a survey a customer would provide comments (Explicative) if the service provided is actually been very good or bad as per the customer's request. Now this will also have complications, after scanning and manually analyzing few comments, I found that,
Positive comments are generally easily classified, due to text having standard positive comments (words like Good, Great, Happy, excellent etc.).
Neutral comments are difficult to slot due to the fact that, few comments would say good service provided but there was slight delay, or the staff was friendly, but job was not done right, etc. There is a dilemma here to classify some comments as Neutral but some comments after overall text analysis as Negative. In the tests after training, I found that few comments were classified as Neutral and few as Negative.
Negative comments are slightly difficult to classify, there would be generic comments with negative words (like Bad, Concern, Problem etc.) but there is ambiguity when comments enlist negative service with negative and positive words. There is a dilemma to classify text with good and bad words from a comment. This is where there is slight inaccuracy in the project implementation where not all text from comments may be classified correctly and more research has to be done on the same.
To also get a better understanding of the results, I have also mapped the overall satisfaction ratings provided by all customers to compare the results of the overall service ratings and the survey sentiment analysis from the comments.

II. Related work

A large number of process' have been carried out in past in the domain of online review mining. There are a wide variety of research groups finding diverse ways to use text mining and sentiment analysis as the next generation model. Following are a few relevant existing works whichalign with the classification problem,

--Bo Pang and Lillian Leesurveyed techniques and approaches that promise to directly enable opinion-oriented information-seeking systems. Their focus on methods that seek to address the new challenges raised by sentiment-aware applications, as compared to those that are already present in more traditional fact-based analysis was good for defining current and future goals. They also included materialon summarization of evaluative text and on broader issues regarding privacy, manipulation, and economic impact that the development of opinion-oriented information-access services gives rise to.

--Jeffrey et al studied the classification of network traffic by exploiting the distinctive characteristics of applications when communicate over a network. The paper has two without regard to clustering algorithms, which are K-Means and DBSCAN, that have never been used for network traffic classification. then evaluation of the two algorithms and compared them to the previously used Auto Class algorithm and the results show that K-Means and DBSCAN work very well and much more quickly than Auto Class.

--Antonio et al states the problem of learning to classify the texts by exploiting information derived from both training and testing sets. To carry this, clustering is used as an extra step to text classification which is applied to both training as well as the testing sets. The experiments showed important improvements on classification result especially on small training sets.

--Pang et al applied machine learning, some ways of specifying the online movie reviews, collected from the Internet Movie Database (IMDb), to positive or negative, by obtaining the list of 14 affective key words (love, wonderful, best, great, superb, still, beautiful, bad, worst, stupid, boring, waste) which are then used in a way for specifying accuracy. The resultsproved that by using SVM they were better and accurate. Achieved significant improvement over the break even.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 11, November 2017

III. SYSTEM ARCHITECTURE

TEST TABLE

SR. NO	METHOD	ADVANTAGE	DRAWBACK	RESULT
1.	SBMNaïveBayesmethod	Positive andnegativeonly	Does not work on large system	Improvement over baseline
2.	Clustering	K means andDBSCANwork well	Time consuming for small systems	Accuracy Improved
3.	Text Classification	Best for small training sets	Does not work for large systems.	Substantial improvement.

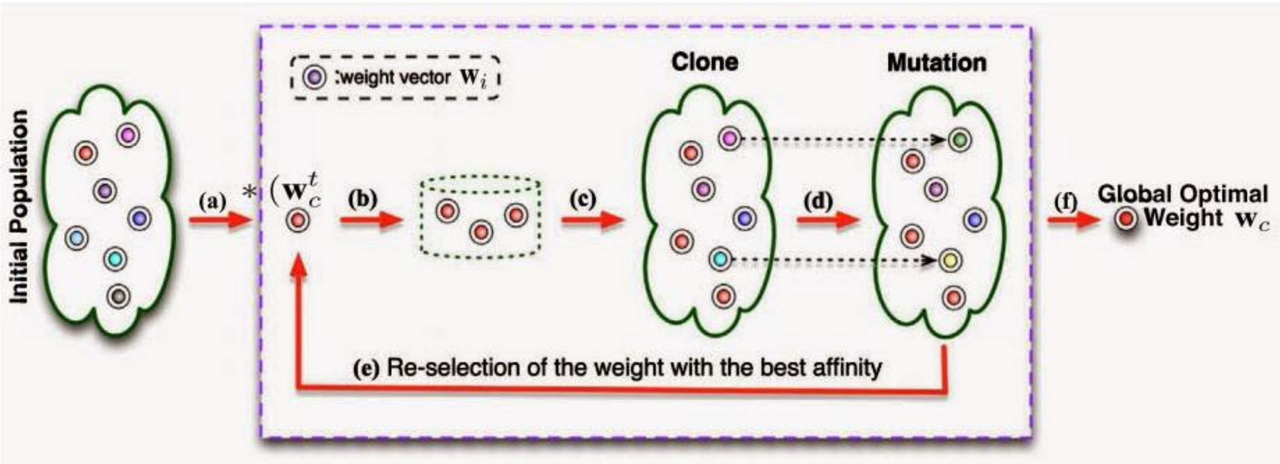
There are six steps in our experiments: sample data, pre-processing, transformation,feature selection,classification and then evaluation which is generating a predictive model.

Sentiment Analysis – Naives Bayes Classification

An objective of a sentiment analysis is to identify text in the form of comments, reviews and messages and classify as Positive or Negative.

There are multiple sentiment classification techniques, SVM, Max Entropy, SentiWordNet, WordNet etc., I have implemented the sentiment classification using Naives Bayes Classification. The reason for this selection is to explore another technique and also due to limited resources as with sentiment analysis with a large data set training time is crucial and with limited CPU and Memory, Naives Bayes is useful as it can be trained quickly. Though Naives Bayes is said to have usually being outperformed (in terms of accurate training and testing of texts) by other classifiers like Random Forest Tree. However even though the probability estimates of Naïve Bayes are of low quality, its classification decisions are quite good, the decision making is good and therefore the model has good accuracy. This can be seen with the final result graphs were the overall sentiment is classified similar to the overall customer satisfaction ratings.

Naïve Bayes Classifier implements bag of words model, where we check words enlisted in the excel/csv file and check which word appears in the positive sentiment word list and which words appear in Negative sentiment list. There are other variations of Naïve Bayes, but I will be using standard model for the demo.



ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 11, November 2017

IV. PROPOSED ALGORITHM

The **Training** Data, is created partially from existing comments and a lot more is manually added with different scenarios and opinions to include a wide variety of language complexity for the algorithm to train on to help achieve a better final test result. I have used Pandas to read CSV files and perform other processing required for excel file cleaning and collating column data.

NOTE: Since the survey contains lot of ambiguous text it has trained well for the existing text, however there are lot of comments which are difficult to classify due to its nature of having positive and negative words, irrelevant context or junk characters, spelling mistakes etc. The model has classified lot of such ambiguous text into Negative.

The Training CSV file consists of comments with classification of sentiment, 1 as positive and 0 as Negative. Now before training the survey excel needs to be cleaned and also unwanted Junk text needs to be removed, also we need to clean all stop words which are filtered out to avoid unnecessary processing and I have used the standard NLTK Stop words that includes a list of English stop words.

```
stopset = set(stopwords.words('english'))
```

After which I have transformed the Sentiment(0-Neg,1-Pos) and Comments in vectors and Vectorized using,

```
vectorizer = TfidfVectorizer(use_idf=True, lowercase=True, strip_accents='ascii', stop_words=stopset).
```

The basic idea is to use Count vector technique to find total number of occurrences of words in the list and we derive TFID by multiplying frequency of word in Train file with frequency of word in actual Test file.

Sample Code,

```
stopset = set(stopwords.words('english'))
vectorizer = TfidfVectorizer(use_idf=True, lowercase=True, strip_accents='ascii', stop_words=stopset)

y = df1.sentiment

x = vectorizer.fit_transform(df1.Comments)

x_train, x_test, y_train, y_test = train_test_split(x, y, random_state=42)

clf = naive_bayes.MultinomialNB()
clf.fit(x_train, y_train)
print("Model Accuracy -> {0}".format(roc_auc_score(y_test, clf.predict_proba(x_test)[:,1])))
print(roc_auc_score(y_test, clf.predict_proba(x_test)[:,1]))
srv = np.array(data)
surveyReviewVector = vectorizer.transform(srv)
```

After transforming comments and sentiments (0's and 1's) , train and testing set will be declared. I have used the Sci-kit learn library for training/learning using the Naives Bayes implementation done for text classification. Also used Multinomial NB which implements the Naïve Bayes algorithm for multinomial distributed data, the distribution is parameterized by vectors.
After training I have also tried to find accuracy and the model has trained with 83% accuracy.

using roc_auc_score(y_test, clf.predict_proba(x_test)[:,1]) gives 0.83333, which is a pretty good learning curve as compared to initial tests that were made with limited training data sets (Initial learning rate was close to 69% ie, with

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 11, November 2017

raw and uncleaned data).Final Result is an array of 1's (positive) and 0's(negative) which can be collated to summarize the overall sentiment. This is eventually shown in form of charts as results.

However, when provided a larger data set with higher unambiguous texts the accuracy dropped to a low 0.3952, as the data set was increased with larger words (bag of words) the accuracy dropped and stayed close to 35%. This shows that a higher training set is required with more columns or data to classify a particular type of text and predict with correct meaning. The ambiguity in texts has decreased the pattern recognition and sentiment analyzing accuracy.

Evaluation with another Model

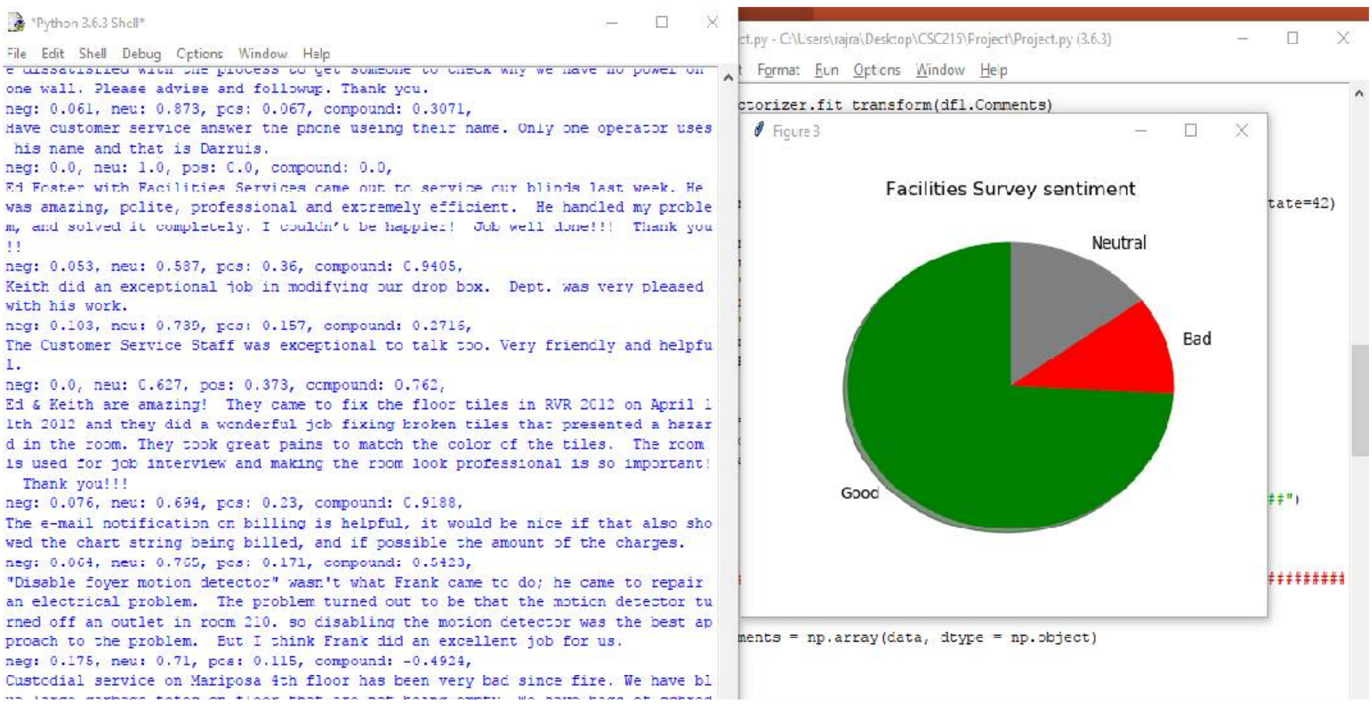
To compare existing implementation result and check for accuracy/consistency I have tried to compare with implementing another sentiment classification method using NLTK.Vader library. This library has a trained model on pre-defined texts and I have tried to train and test with the same Facilities Survey Data set.

The Result is fairly similar to the implemented Naïve Bayes model, this model however also lists Neutral comments (lot of ambiguous texts like positive and negative text in comments have been partially classified as Negative), also comments with no context to the survey will list under neutral category.

V. SIMULATION RESULTS

Overall, we see that the models have trained fairly well with training data of similar text and this can be verified by the overall satisfaction collation from the excel.

Following is the snapshot of output of texts being classified into different categories, along with the Overall sentiment using the NLTK.vader library.



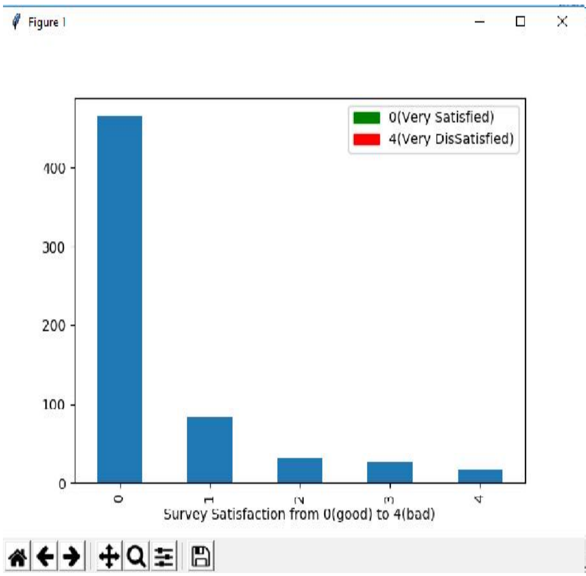
ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

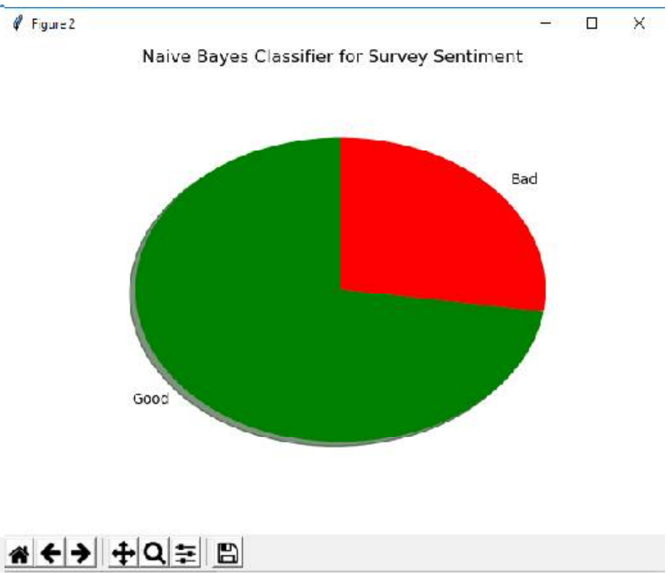
International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 11, November 2017





The Graph when compared with the sentiment analysis Pie charts of both implementations shows a similar pattern and we can see that the majority is good feedback with positive sentiment and Negative is a portion of the overall sentiment.

VI. CONCLUSION AND FUTURE WORK

The Implementation covers a basic implementation with a known model, the result does provide a good accuracy, however this was a small data set in comparison to large range of language context and inferences that can be derived from a text. There needs to be larger variety of texts for training as well to allow the testing to other unfamiliar data and provide good and accurate prediction.
There is also the inaccurate classification for sarcasm and understanding of other non-generic text is poor. Classification of Neutral text is unclear.
Also Implementation of further classification of emotions extraction is an interesting exploration which may result in better or more accurate classification as well.

Library/API used

1. Pandas
2. Python, NLTK
3. NLTK Corpus for StopWords
4. NLTK SentimentAnalyzer
5. SCI-KIT Learn TfidfVectorizer (Vectorize extracted text)
6. SCI-KIT naive_bayes
7. SCI-KIT Metrics to get Prediction Score
8. Numpy

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 11, November 2017

REFERENCES

1. M. K. Jiawei Han, "Data Mining:Concepts and Techniques". 500 Sansome Street, Suite 400, San Francisco, CA 94111: Diane Cerra, 2006.
2. B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002, pp. 79–86.
3. https://www.kaggle.com/crowdflower/twitter-user-gender-classification/feed
4. Bo Pang and Lillian Lee, Opinion Mining and Sentiment AnalysisFoundations and TrendsinInformation RetrievalVol. 2, Nos. 1–2 (2008) 1–135, 2008 B. Pang and L. Lee DOI: 10.1561/1500000001
5. L. Zhuang, F. Jing, X.-Y. Zhu, and L. Zhang, "Movie review mining and summarization,"inProceedings of the ACM SIGIR Conference on Information and Knowledge Management (CIKM), 2006.
6. D.-H. Park, J. Lee, and I. Han, "The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement," International Journal of Electronic Commerce, vol. 11, pp. 125–148, (ISSN 1086-4415), 2007.
7. K. Nigam and M. Hurst, "Towards a robust metric of polarity," in Computing Attitude and Affect in Text: Theories and Applications, number 20 in The Information Retrieval Series, (J. G. Shanahan, Y. Qu, and J. Wiebe, eds.), 2006.
8. V. Ng, S. Dasgupta, and S. M. N. Arifin, "Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews," in Proceedings of the COLING/ACL Main Conference Poster Sessions, pp. 611–618, Sydney, Australia: Association for Computational Linguistics, July 2006.
9. F. Mosteller and D. L. Wallace, Applied Bayesian and Classical Inference: The Case of the Federalist Papers. Springer-Verlag, 1984.
10. T. Mullen and N. Collier, "Sentiment analysis using support vector machines with diverse information sources," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 412–418, July 2004. (Poster paper).
11. T. Mullen and R. Malouf, "Taking sides: User classification for informal online political discourse," Internet Research, vol. 18, pp. 177–190, 2008.
12. T. Mullen and R. Malouf, "A preliminary investigation into sentiment analysis of informal political discourse," in AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW), pp. 159–162, 2006.
13. T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in Proceedings of the Conference on Knowledge Capture (K-CAP), 2003.
14. J. Liu, Y. Cao, C.-Y. Lin, Y. Huang, and M. Zhou, "Low-quality product review detection in opinion summarization," in Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), pp. 334–342, 2007. (Poster paper).
15. L. Lloyd, D. Kechagias, and S. Skiena, "Lydia: A system for large-scale news analysis," in Proceedings of String Processing and Information Retrieval (SPIRE), number 3772 in Lecture Notes in Computer Science, pp. 161–166, 2005.
16. A. Agarwal and P. Bhattacharyya, "Sentiment analysis: A new approach for effective use of linguistic knowledge and exploiting similarities in a set of documents to be classified," in Proceedings of the International Conference on Natural Language Processing (ICON), 2005.
17. C. O. Alm, D. Roth, and R. Sproat, "Emotions from text: Machine learning for text-based emotion prediction," in Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), 2005.
18. A. Andreevskaia and S. Bergler, "Mining WordNet for a fuzzy sentiment: Sentiment tag extraction from WordNet glosses," in Proceedings of the European Chapter of the Association for Computational Linguistics (EACL), 2006.
19. S. Argamon-Engelson, M. Koppel, and G. Avneri, "Style-based text categorization: What newspaper am I reading?" in Proceedings of the AAAI Workshop on Text Categorization, pp. 1–4, 1998.
20. R. Barzilay and K. McKeown, "Extracting paraphrases from a parallel corpus," in Proceedings of the Association for Computational Linguistics (ACL), pp. 50–57, 2001.
21. J. Berger, A. T. Sorensen, and S. J. Rasmussen, "Negative publicity: When is negative a positive?," Manuscript. PDF file's last modification date: October 16, 2007, URL: http://www.stanford.edu/∼asorense/papers/ Negative Publicity.pdf, 2007.