SENTIMENT ANALYSIS USING REPRESENTATIVE TERMS A GROUPING APPROACH FOR BINARY CLASSIFICATION OF DOCUMENTS

[1]N.SRINIVASA GUPTA, [2]B.VALARMATHI, [3]SANJU JOSEPH

[1]Faculty of Manufacturing Engineering Division, School of Information Technology and Engineering, VIT University, Vellore, India.

[2]Faculty of Soft Computing Division, School of Information Technology and Engineering, VIT University, Vellore, India.

[3]Assistant Software Engineer, Tata Consultancy Services, Bangalore, India.

E-mail: [1]guptamalai@gmail.com, [2]valargovindan@gmail.com, [3]aramathsanju@gmail.com

ABSTRACT

Automatic classification of customer reviews as either positive or negative has been of great interest among the academic and business community in the recent times. In this paper, an attempt has been made to represent the text documents using just eight representative terms (RT) viz. good, very good, excellent, recommended, bad, very bad, disgusting, and never recommended. Thus a new way of representing text documents as a structured data matrix has been created. A consistent classification accuracy of near 80% and above was achieved for datasets of various sizes ranging from 403 to 25000. The precision (P), recall(R) and F-Measure were also very consistent and comparable to the previously reported results. A comparative analysis of classification performance has been carried out using machine learning algorithms like Naïve Bayes (NB), Bayesian logistic regression (BLR), multi layer perceptron (MLP) etc., revealed that the proposed way of representing the text documents results in consistently superior performance.

Keywords: *Sentiment Detection, Review Classification, Opinion Mining, Machine Learning Algorithms, Dimensional Reduction*

1. INTRODUCTION

Over the past twenty years, we have been experiencing the information explosion year after year. With the ever increasing number of internet users and also the ever increasing usage level of the existing users of internet, one can easily conclude that the information explosion is going to continue in the coming decades too. Due to this, we now have enormous amount of text documents from various users on variety of topics. In order to uncover the hidden knowledge in those text documents and to find out the connection among them, a field of study was created in the name of Computational Linguistics (CL).

During the early stages of practice, the experts in CL focused on finding ways to search and categorize the concepts found in research articles, books, legal documents, patent details and other records that were available in the digital formats. Recently, the focus has shifted to mine the textual information available with online news papers, wikis, blogs, websites, emails and the other databases. In general this task is known as Text mining (TM) or text analytics.

Linguistics-based TM uses the concepts of natural language processing (NLP), whereas the statistics-based TM relies on self-learning tools such as support vector machines (SVM), fuzzy-neural networks, Bayesian classification, and latent semantic analysis etc.

Opinion Mining (OM) or sentiment mining aims at detecting the sentiment expressed in a review by using statistical self-learning techniques or NLP. This is also known as sentiment detection. As per Bing Liu [2], the objective of sentiment detection is to classify the review documents as either positive or negative based on the sentiment expressed by the customers in their reviews. This kind of classification is carried out at the document level, without discovering about what people liked or did not like.

2. RELATED WORKS IN SENTIMENT ANALYSIS

The research in sentiment analysis can be broadly classified into two categories viz. 1. Data mining approach 2. Natural language processing approach

2.1 Related Works using the Data Mining Approach

Data mining approach accomplishes the job of classification by expressing the unstructured text documents as structured term-document matrix containing numerical scores. This approach borrows several techniques from computational linguistics and information retrieval to convert the unstructured data (text documents written in natural language) as structured data. After this, the machine learning algorithms are applied for the purpose of classification.

Turney, D [10] proposed a simple unsupervised learning algorithm for the classification of reviews into either recommended or not recommended category. He achieved 66% of accuracy for movie related reviews, 80% and 84% for banks and automobile related reviews.

Dave et al. [3] used machine learning algorithms like Naïve Bayes (NB), Maximum Entropy (ME) and Support Vector Machines (SVM) and achieved a classification accuracy of 88.9% using his unigrams, bigrams and trigrams model. Pang & Lee (2004) used NB and SVM for classifying the movie reviews and achieved a maximum of 87.2% accuracy. He used only unigrams in his experiment. Gamon (2004) used SVM for classifying customer feedback and achieved a classification accuracy of 77.5%.

Pang & Lee [8] used SVM, and regression tools for classifying the movie reviews and for assigning the sentiments on a three-point/four-point scale. He achieved 66.3% of accuracy in his experiment. Konig & Brill [6] used SVM and hybrid techniques for movie review classification and achieved approximately 91% of accuracy, but they achieved 91% accuracy for a 90% training data. They used unigrams, bi and trigrams in their experiment. Sequential minimal optimization algorithm was used for training the SVM classifier.

Valarmathi et al. [11] used Mahalanobis distance as a measure to classify review documents as either positive or negative. Only unigrams were considered in their experiment and they used SVD for reducing the dimensions. They achieved 96.6% of accuracy for 300 movie reviews consisting of 150 positive and 150 negative reviews. But they used 93% of the positive reviews to create the mahalanobis space (MS), using which the mahalanobis distance of the test documents was calculated.

2.2 Related Works using the NLP Approach

Natural language processing (NLP) approach to sentiment analysis deals with automatic extraction of meaning/sentiment from the natural language text using POS tagging, developing a lexicon and pattern analysis.

Yi et al [12] used NLP based sentiment analyzer for capturing the sentiment of the topic. Their research focused on assigning sentiments to each of the references corresponding to the given subject rather than assigning the sentiment for the entire document.

Nasukawa et al [7] used NLP for assigning sentiments at topic level using a pattern based approach. 255 camera reviews have been used for the purpose of evaluation. They achieved an accuracy of 94.5%.

Hiroshi et al [5] used NLP for assigning sentiments at topic level using a pattern based approach. 200 camera reviews have been used for the purpose of evaluation. They achieved an accuracy of 89-100%.

In this paper, the Representative Term – Document Matrix (RTDM) is created using the opinion phrase library created by reading the reviews manually. A PERL program has been used for capturing the opinion words and phrases from the review documents and to assign appropriate RT for them.

3. EXPERIMENTS

This section describes the experiments carried out using the large movie review benchmark dataset provided by Andrew L. Mass et al. [1] for evaluating the performance of machine learning algorithms when the input is given in the form of RTDM.

3.1 Experimental Procedure

First the documents were structured as RTDM using the PERL program developed for this purpose. Each review document is represented as a row in the RTDM with eight features representing it. The eight representative features are Good, Very good, Excellent, Recommended, Bad, Very bad, Disgusting and Never recommended. Based on the RT occurrence (number of times the corresponding category of phrase/word is found in a document), the RTDM is constructed. A sample of RTDM is shown in the Table 1, in which the rows represent the reviews and columns represent the features mentioned above.

The numbers in Table 1 represents the number of times a corresponding category of RT appeared in that review document. The rules to capture the RT from the review documents were written based on 200 positive reviews and 200 negative reviews from the large movie review dataset containing 25000 reviews. A sample set of rules used for capturing the relevant RT in the document are given below:

```
$_ =~ s/\s+joy watch\s+/ good /g;
$_ =~ s/\s+just one fabulous\s+/ good /g;
$_ =~ s/\s+last forever\s+/ very_good /g;
$_ =~ s/\s+looks good\s+/ very_good /g;
$_ =~ s/\s+best actor\s+/ excellent /g;
$_ =~ s/\s+not forget\s+/ excellent /g;
$_=~ s/\s+pleasently surprised\s+/ excellent /g;
$_ =~ s/\s+hooked me\s+/ recommended /g;
$_=~s/\s+incredibly brilliant\s+/ recommended /g;
$_ =~ s/\s+must see\s+/ recommended /g;
$_ =~ s/\s+oscar worthy\s+/ recommended /g;
$_ =~ s/\s+never shine\s+/ bad /g;
$_ =~ s/\s+not smart\s+/ bad /g;
$_ =~ s/\s+poorly planned\s+/ bad /g;
$_ =~ s/\s+no suspense\s+/ very_bad /g;
$_ =~ s/\s+not able grab\s+/ very_bad /g;
$_ =~ s/\s+nothing perfect\s+/ very_bad /g;
$_ =~ s/\s+never exemplify\s+/ disgusting /g;
$_ =~ s/\s+no fizzing\s+/ disgusting /g;
$_ =~ s/\s+poor performance\s+/ disgusting /g;
$_ =~ s/\s+most awful\s+/ never_recommended /g;
$_ =~ s/\s+nasty\s+/ never_recommended /g;
```

The rules were written considering all the individual words, phrases and negative patterns. POS tagger has not been used for the purpose of capturing the opinion words or phrases. The entire process of writing the rules was based on how the human mind would understand while reading a review. So the rules were written manually by reading 400 reviews consisting of 200 positive reviews and 200 negative reviews.

Table 1 A Sample RTDM

Document No.	RT							
	Good	Very Good	Excellent	Recommended	Bad	Very Bad	Disgusting	Never Recommended
1.	2	0	0	0	3	5	2	2
2.	1	3	2	0	3	4	4	1
3.	12	1	0	4	4	0	4	0
4.	9	7	4	0	13	5	1	2
5.	4	3	2	4	4	1	1	0
6.	7	2	1	2	6	1	2	0
7.	0	0	0	1	2	1	1	0

.

4 4. EXPERIMENTAL RESULTS

Table 2 to Table 5 show the performance measures like Precision (P), Recall (R), F-Measure and Accuracy (A) of various classifiers. The RTDM of movie reviews has been given as input to the classifiers. The performance of classifiers like Naïve Bayes (NB), Bayesian Logistic Regression (BLR), Multilayer perceptron (MLP), Sequential minimal optimization (SMO), Classification and Regression Tree (CART) are given in Table 2 to Table 5. All the results are based on a 10 fold cross validation test.

Table 2 Performance of Various Classifiers for 403 Reviews (201 Positive and 202 Negative)

Classifier	P	R	F-Measur	Accuracy
NB	0.89	0.89	0.89	0.89
BLR	0.895	0.89	0.89	0.89
MLP	0.89	0.87	0.88	0.88
SMO	0.9	0.88	0.89	0.89
CART	0.84	0.8	0.82	0.82

Table 3 Performance of Various Classifiers for 2000 Reviews (1000 Positive and 1000 Negative)

Classifier	P	R	F-Measure	Accuracy
NB	0.74	0.88	0.81	0.79
BLR	0.84	0.83	0.8	0.83
MLP	0.84	0.82	0.83	0.83
SMO	0.83	0.84	0.84	0.83
CART	0.8	0.77	0.78	0.79

Table 4 Performance of Various Classifiers for 11000 Reviews (5500 Positive and 5500 Negative)

Classifier	P	R	F-Measure	Accuracy
NB	0.72	0.85	0.78	0.76
BLR	0.8	0.82	0.81	0.81
MLP	0.81	0.8	0.8	0.80
SMO	0.81	0.81	0.8	0.80
CART	0.78	0.78	0.78	0.78

Table 5 Performance of Various Classifiers for LDS25000 (12500 Negative and 12500 Positive)

Classifier	P	R	F-Measure	Accuracy
NB	0.71	0.85	0.77	0.75
BLR	0.8	0.8	0.8	0.8
MLP	0.78	0.81	0.8	0.79
SMO	0.8	0.8	0.8	0.8
CART	0.79	0.77	0.77	0.78

4.1 Discussion

The NB classifier showed a consistently superior recall compared to all the other classifiers, whereas the accuracy of BLR, MLP and SMO has been consistent for all the sizes of dataset ranging from 403 to 25000. Accuracy level of near 90% for 403 reviews and a consistent 80% and above accuracy for all other sizes of dataset has been achieved by BLR, MLP and SMO using the eight representative dimensions. Generally it is reported that, the performance of the NB classifier is the worst among the classifiers and not dependable, but with the new format of expressing the documents as RTDM, even the NB classifier performed with a comparable accuracy with respect to the other classifiers.

5. CONCLUSION

In this research paper, an attempt has been made to combine the best of rule based classification and machine-learning approaches to achieve a better accuracy, precision, recall and F-measure. The important aspect of this research is the creation of RTDM to represent the text documents. In the usual approach using data mining techniques, after the identification of features of all the documents, singular value decomposition (SVD) technique is used to obtain a reduced dimension matrix. In this research, we have succeeded in creating a reduced dimension matrix with just eight representative dimensions, which results in a comparable classification accuracy and other performance measures like precision, recall and F-Measure. Though this approach requires creation of opinion phrase library specific to each product type or domain, the benefit is measurable in terms of better classification accuracy, precision, recall and F-measure.

REFRENCES:

[1] J.B. Andrew L. Mass., Raymond, E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng., and Christopher Potts., ”Learning word vectors for sentiment analysis”, *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, Oregon, USA, 2011, pp. 142-150.

[2] Bing Liu, Web Data Mining Exploring Hyperlinks, Contents, and Usage Data, Springer, 2008, pp. 411- 448.

[3] Dave, K., Lawrence, S., and Pennock, D.M., “Mining the peanut gallery: opinion extraction and semantic classification of product reviews”, *Proceedings of the 12th International WWW Conference*, Budapest, Hungary, 2003, pp. 519-528.

[4] Gamon, M., “Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis”, *Proceedings of the 20th International Conference on Computational Linguistics*, Geneva, Switzerland, 2004, pp. 841-847.

[5] Hiroshi, K., Tetsuya, N., and Hideo, W., “Deeper sentiment analysis using machine translation technology”, *Proceedings of the 20th International Conference on Computational Linguistics*, Geneva, Switzerland, 2004, pp. 494-500.

[6] Konig, A.C., and Brill, E., “Reducing the human overhead in text categorization”, *Proceedings of the 12th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Pennsylvania, USA, 2004, pp. 598-603.

[7] Nasukawa, T., and Yi, J., “Sentiment analysis: capturing favorability using natural language processing”, *Proceedings of the 2nd International Conference on Knowledge Capture*, Florida, USA, 2003, pp. 70-77.

[8] Pang, B., and Lee, L., “A sentiment education: sentiment analysis using subjectivity summarization based on minimum cuts”, *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, Barcelona, Spain, 2004, pp. 271-278.

[9] Pang, B., and Lee, L., “Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales”, *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, University of Michigan, USA, 2005, pp. 115-124.

[10] Turney, P. D., “Thumbs up or thumbs down? sentiment orientation applied to unsupervised classification of reviews”, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, USA, 2002, pp. 417-424.

[11] Valarmathi, B., and Palanisamy, V., “Opinion mining of customer reviews using mahalanobis-taguchi system”, *European Journal of Scientific Research*, Vol. 62, 2011, pp. 95-100.

[12] Yi, J., Nasukawa, T., Niblack, W., and Bunescu, R., “Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques”, *Proceedings of the 3rd IEEE International Conference on Data Mining*, Florida, USA, 2003, pp. 427-434.