A Modified Metaheuristic Algorithm for Opinion Mining

K.Saraswathi
Assistant Professor
Dept. Of Computer Technology- Ug
Kongu Engineering College
Perundurai, Erode, Tamilnadu, India.

A.Tamilarasi. Phd.
Professor & Head
Dept. Of Computer Applications
Kongu Engineering College
Perundurai, Erode, Tamilnadu, India

ABSTRACT

Opinion mining is a recent discipline combining Information Retrieval and Computational Linguistics which is concerned with the opinion a document expresses and not just with the topic in the document. Online forums, newsgroups, blogs, and specialized sites provide voluminous information feeds from where opinions can be retrieved. Opinion's polarity is established through application of machine learning techniques for classification of textual reviews as either a positive or negative class. In this paper, it is proposed to extract the feature set from reviews using Inverse document frequency and the reviews are classified as positive or negative using Bagging algorithms. The proposed method is evaluated using a subset of Internet Movie Database (IMBd).

General Terms

Data Mining, Classification Accuracy.

Keywords

Opinion mining, Sentiment analysis, Movie reviews, Naive Bayes, CART, Bagging.

1. INTRODUCTION

With the development of Web, user participation is on the rise in many websites. Many online forums, websites, blogs encourage users to post reviews on information in which they are interested. These reviews become useful to information promulgators and readers. For example, the government can perceive the influence of recent policies/events on common people based on the online review of political news/announcements, and the information from the reviews help formulate remedial action. On the other hand, through product reviews such as phones, laptops, movies, books, manufacturers can gather feedbacks from customers to improve their products and service further.

Online shopping is widely popular as it is convenient, reliable, and quick. Customers want to compare similar products and go through reviews before making purchasing decisions based on images and product descriptions given by the seller. Online retailers aim to give consumers a complete shopping experience based on price, manufacturer, and various other attributes of the products. Review of products by the customer is encouraged by the retailers as customer's product reviews are generally more honest, unbiased and comprehensive. Also, reviews written by other customers reveal usage experience and the perspectives of existing customers with similar needs. A study by comScore and Kelsey group [1] revealed that online customer reviews impacted prospective buyers significantly. But as availability of customer reviews increases, it is also impossible for a single user to read all reviews to make informed decisions. People can objectively evaluate a product by seeing others opinions, which will influence decisions on whether to buy a product or not. Opinion is biased when only a few reviews are read. Thus automatic review mining and summarization is a current hot research topic.

The availability of customer reviews has led to valuable research related to opinion mining, summarizing customer reviews [2, 3, 4, 5, 6, 7]. There was considerable work on sentiment analysis of sentences in reviews, as also of review sentiment orientation as a whole [8]. Most existing work on review mining and summarization focuses on product reviews. This paper focuses on another domain – movie review. Differing from product reviews, movie reviews have unique characteristics. When a person reviews a movie, the comments are not only on movie elements but also on people related to the movie. Whereas in product reviews, few will care for issues like who designed/manufactured a product. Hence, commented features in movie review are much richer than in a product review. This results in movie review mining being more challenging than product review mining.

In this paper, classification of opinion of online movie review is investigated based on opinion words and corpus words which are frequently used in the documents under review. The corpus is ranked using Singular value decomposition and prepares data for opinion mining. This paper is organized into the following sections: Section II briefly reviews some of the related works available in literature, Section III describes the materials and methods and classification algorithms, section IV describes the results obtained and discusses the same.

2. RELATED WORKS

Current works on opinion mining mainly focus on product reviews. Popescu and Etzioni proposed the OPINE system, that used relaxation labeling to find words' semantic orientation [9]. Gamon, et al., introduced Pulse system [10]; a bootstrapping process trained a sentiment classifier. Features were extracted by labeling sentence clusters based on key terms.As pioneers, Hu and Liu suggested a method which used word attributes, including occurrence frequency, part-of-speech and synset in WordNet [11]. First, product features are extracted and then were combined with nearest opinion words, from a generated and semantic orientation labeled list containing adjectives alone. Finally, a summary is produced through selecting and re-organizing sentences according to extracted features. Liu et al expanded an opinion word list through the addition of nouns for dealing with reviews in a special format [12]. Some of the sentiment analysis related to movie reviews is as follows.

In earlier sentiment analysis works, the determination is simply to decide a review is thumbs up or thumbs down. Pang, et al., [13] discussed the problem of the rating-inference, an authors' estimation must be determined regarding a multi-point scale (for exemplar; one to five "stars"). On the standard multi-class text categorization, the task represents an attractive twist as there are many various degrees of similarity among the class labels (for exemplar;

rather than to “one star”, “three stars” is intuitively near to “four stars”). In this task, initially the human performance is evaluated. Then, on the basis of a metric labeling formulation of the problem, a meta-algorithm is implemented, which modifies the output of a provided n-ary classifier to guarantee that similar labels are received by similar items in an explicit attempt. On both multi-class and regression versions of SVMs while implementing a new similarity measure suitable to the problem; the meta-algorithm proposed reveals essential enhancements.

Whitelaw et al., [14] proposed a novel method for sentiment classification. In much of the task-independent semantic taxonomy, based on Appraisal Theory, an appraisal group is characterized as a collection of attribute values. A lexicon of appraising adjectives and their modifiers are developed by employing semi-automated techniques. Implementing the features based on these taxonomies mutually with the standard “bag-of-words” features, the movie reviews are classified that revealed a result of 90.2% of state-of-the-art accuracy. The observation reveals that few appraisal types are more essential than the others for sentiment classification. Only by a taxonomic analysis of appraisal type it is able to insight the above mentioned type. Therefore the proposed method is capable to enhance the results of traditional word-based methods.

Kennedy et al., [15] introduced two techniques to resolve the sentiment that is expressed by a movie review. A review’s semantic orientation can be neutral, positive or negative. While classifying the reviews, the effect of valence shifters was investigated. Three types of valence shifters were investigated, they are: 1) intensifiers, 2) negations and 3) diminishers. To reverse the semantic polarity of a particular term, the negations are implemented whereas diminishers and intensifiers are implemented to decrease and increase respectively, the degree to that a term is negative or positive. The reviews are classified based on the number of positive and negative terms they contain by the first technique. To recognize the positive and negative terms along with the intensifiers, negation and diminishers terms is performed by implementing General Inquirer. The association scores with a small group of positive and negative terms are used to compute corpus-based semantic orientation values of terms. Along with contextual valence shifters, the extending of the term-counting technique enhances the classifications accuracy. A Machine Learning algorithm, Support Vector Machines is employed by the second method. Firstly, unigram features are used and then with the addition of bigrams that contains another word and a valence shifter is done. The valence shifter bigrams slightly increases the accuracy classification that results in very high accuracy. The high accuracy is contributed by the features that are the words in the lists of positive and negative terms. Instead of using one method, combining two methods yields proven better results.

Mullen et al., [16] proposed to bring jointly different sources of potentially pertinent information considering many favorability measures for adjectives and phrases and the available knowledge of the topic of the text utilizes support vector machines (SVMs) and presents a methodology to sentiment analysis. Implementing the features, the models presented are additionally united with unigram models that have been revealed to be efficient in the past and versions of the unigram models that are lemmatized. The hybrid SVMs that joins unigram-style feature-based SVMs are illustrated including on the basis of real-valued favorability measures that attains superior performance and gives the most excellent results as demonstrated from experiments on movie review data from Epinions.com.

3. MATERIALS AND METHODS

3.1 Dataset

The data set of movie reviews by Pang and Lee (2004) [17] containing 2,000 movie reviews: 1,000 positive and 1,000 negative is used for evaluating the classification algorithms. An earlier version of this data set, having 700 positive and 700 negative reviews, was used in Pang et al. (2002) [18]. The reviews were collected from an Internet Movie Database (IMDb) archive rec.arts.movies.reviews. Their positive or negative classification is automatically extracted from ratings, as specified by the reviewer. Only reviews where the movie rating is indicated by the stars or some numerical system are included in the dataset. For this study, a subset of 150 positive and 150 negative opinions is used.

3.2 Inverse Document Frequency (IDF)

To classify the documents, features are extracted. List of stop words (commonly occurring words) and stemming words (words with similar context) are prepared. The terms document frequency (df) is computed which is the number of document that contain the term. The terms occurring rarely are more informative than terms occurring frequently. Thus, it is required to assign weights such that rare words have higher weights than the frequent terms. This is captured by document frequency of term t (df$_t$). The inverse document frequency (idf$_t$) represents the scaling factor. The importance of the term t is scaled down if it appears frequently in documents. The idf$_t$ is defined as follows:

$$IDF\left(a\right) = \log \frac{1+\left|x\right|}{x_a}$$

x_a is the set of documents containing term a. Following is an example illustrating the relation between the df$_t$ and idf$_t$ for a total of million documents.

Term	df$_t$	idf$_t$
brilliant	100	4
good	1000	3
under	10000	2
with	100000	1
the, movie	1000000	0

3.3 Naïve Bayes

Given a set of objects belonging to a known class and having a known vector of variables, the goal is construction of a rule that will enable assigning future objects to a class, if only vectors of variables are given describing future objects. Such problems of supervised classification are ubiquitous, and methods for construction of such rules were developed. Naïve Bayes classifier is one of the most commonly used classification method. It is easy to build not requiring complicated iterative parameter estimation schemes so that it can be easily applied to large data sets. As it is easy to interpret, users not familiar with classifier technology easily understand why it is makes the classification it does. Lastly, the Naive Bayes model is appealing due to its simplicity, elegance, and robustness. One of the oldest classification algorithms, yet it is effective in its simplest form. It is used in areas like text classification and spam filtering. Many modifications were introduced, by statistical, data mining, machine learning, and pattern recognition communities, to make it highly flexible [19].

When the inputs are represented by its feature vector X and the classes are represented by C, Naïve Bayes predicts class as follows:

$$P\left(X\left|C\right.\right)=\Pi_{i=1}^{n}P\left(X_i\left|C\right.\right)$$

3.4 Classification and Regression Trees (CART)

The CART decision tree is a binary recursive partitioning procedure used to processing continuous and nominal attributes both as target and predictor [20]. Data is handled in raw form with no binning being required. Trees are grown to maximum size without a stopping rule and then pruned (usually split by split) to the root through cost-complexity pruning. The next split which is pruned contributes least to overall tree performance on training data (more than one split may be removed at a time). This method produces trees which are invariant under predictor attributes order preserving transformation. The CART mechanism aims to produce a sequence of nested pruned trees. The “right sized” or “honest” tree is identified through evaluating every tree’s predictive performance in a pruning sequence. CART offers no training data based internal performance measures for tree selection as such measures are suspect. Instead, tree performance is measured on independent test data or through cross validation and tree selection starts only after test-data-led evaluation. CART splitting rules are in the form

> *An instance goes left if* CONDITION, *and goes right otherwise,*

where CONDITION is expressed as “attribute Xi <= C” for continuous attributes. For nominal attributes, CONDITION is stated as membership of a values list.

3.5 Bagging

Bagging is used to improve stability and predictive power of classification and regression trees [20]. However its use is not restricted to improving tree-based predictions, but is a general technique applicable in a variety of settings to improve predictions.

To understand how and why bagging works and determine the situations where one can expect improvement through bagging, the problem of predicting the value of a numerical response variable, Yx, that will result from, or occur with, a given set of inputs, x, should be considered. Suppose that φ(x) is the prediction resulting from using a particular process like CART, or OLS regression with a prescribed method for model selection (using Mallows’ Cp to select a model from all linear models having only first- and second order terms from input variables). Allowing μφ denote E(φ(x)), where expectation is with regard to distribution underlying the learning sample (as viewed as a random variable, φ(x) is a function of learning sample viewed as a high-dimensional random variable) and not x (considered to be fixed), we have the following equations.

$$\left(E\left[y_x-\mu_\varphi\right]^2\right)=\left(E\left[y_x-\mu_\varphi\right]+\left[\mu_\varphi-\varphi\left(x\right)\right]^2\right)$$
$$=\left(E\left[y_x-\mu_\varphi\right]^2\right)+2E\left(y_x-\mu_\varphi\right)E\left(\mu_\varphi-\varphi\left(x\right)\right)+E\left[\mu_\varphi-\varphi\left(x\right)\right]^2$$
$$=\left(E\left[y_x-\mu_\varphi\right]^2+E\left[\mu_\varphi-\varphi\left(x\right)\right]^2\right)$$
$$=\left(E\left[y_x-\mu_\varphi\right]^2\right)+\mathrm{var}\,ience\left(\varphi\left(x\right)\right)$$
$$\geq\left(E\left[y_x-\mu_\varphi\right]^2\right)$$

The independence of future response, Yx, and predictor based on learning sample, φ(x), is used.) As in nontrivial situations, predictor variance φ(x) is positive (as not all random samples can yield the sample value for the prediction), so that the inequality above is strict, this leads us to the result that if μφ = E(φ(x)) is a predictor, it would have a reduced mean squared prediction error than does φ(x).

Breiman gives some idea of how bagging improves predictions, and of performance variability when considering different data sets [21]. When classification examples were examined, bagging reduced CART’s misclassification rates by 6% to 77%. Though this improvement is small, it could be due to the fact that there is only limited room for improvement. Thus it could be because both unbagged and bagged classifiers have misclassification rates close to the Bayes rate. It could be that bagging actually achieved limited amount of improvement possible. When compared with 22 other classification methods used in the Statlog Project (see Michie et al., 1994) on four publicly available data sets from the Statlog Project, Breiman shows that bagged trees were the best overall (though boosted classifiers were not considered) [21]. With regression examples considered by Breiman, bagging reduced CART’s mean squared error by 21% to 46%.

4. RESULTS

Experiments are conducted for sentiment classification using online movie review data. 300 instances (150 positive and 150 negative) were used for evaluation. Following Tables and Figures give the classification accuracy, precision and recall for the various classifiers used for classifying the opinion into positive or negative.

Following Tables 1 & 2 and Figures 1 to 3 give the classification accuracy, precision and recall for the various classifiers used for classifying the opinion into positive or negative. RMSE is an error metric which measures the difference between the predicted values and the actual values. RMSE is calculated as follows:

$$RMSE=\sqrt{\frac{1}{N}\sum_{i}^{N}\left(y_i-x_i\right)^2}$$

Where y_i is the predicted value and x_i is the actual value

Table 1. Classification Accuracy and RMSE for various classifiers used

Technique used	**Classification Accuracy**	**RMSE**
Naive Bayes	66.67%	0.5368
CART	65.33%	0.4743
Bagging	**80.67%**	**0.4397**

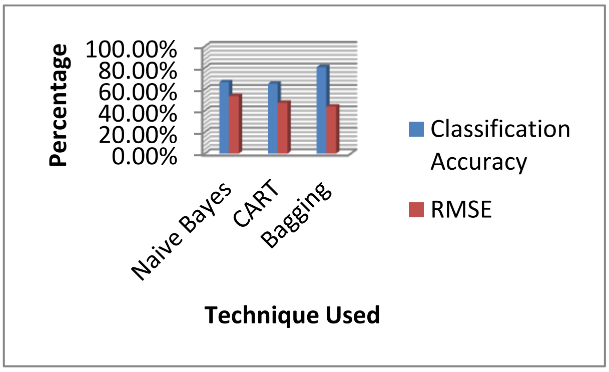

Figure 1: Classification Accuracy and RMSE for various classifiers used

It is seen from Figure 1, that the classification accuracy achieved by Bagging is much better than that of Naïve Bayes and CART. Bagging achieves 14 to 15.34% better classification accuracy than the other classifiers.

The precision, recall and f Measure values are given by:

$$\Pr ecision = \frac{True \text{ positives}}{True \text{ positives} + false \text{ positives}}$$

$$\text{Re}\, call = \frac{True \text{ positives}}{True \text{ positives} + false \text{ negatives}}$$

$$fMeasure = 2 * \frac{precision * recall}{precision + recall}$$

Table 2. Precision and Recall values

Technique used	Precision	Recall	F Measure
Naive Bayes	0.667	0.667	0.667
CART	0.661	0.653	0.649
Bagging	0.807	0.807	0.807

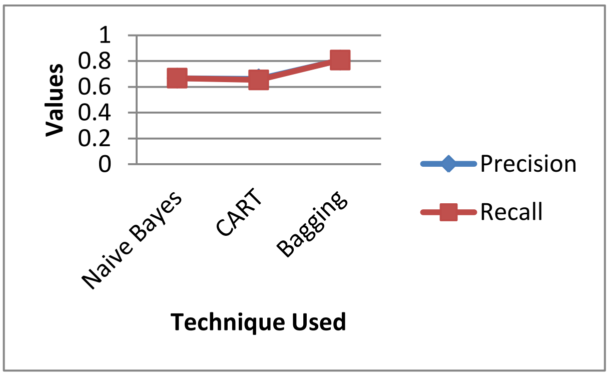

Figure 2: Precision and Recall

It is observed from Figure 2 that the precision and recall of Bagging is much higher than the Naïve Bayes and CART. As the recall is high, most relevant results are returned.

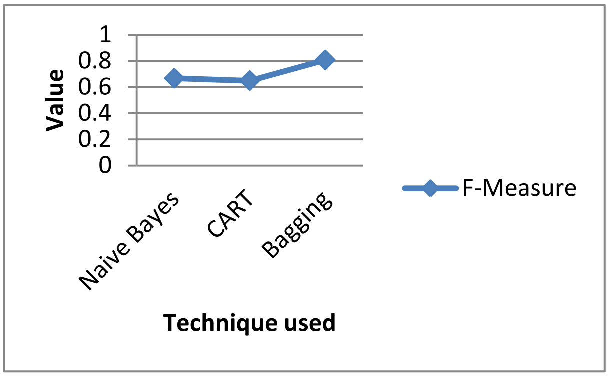

Figure 3: f Measure

5. CONCLUSION

In this paper, it was proposed to investigate the efficiency of Bagging to predict opinions as positive or negative for online movie reviews from IMBD dataset. 300 instances (150 positive and 150 negative) were used for evaluation. The classification accuracy of Bagging was compared against Naïve Bayes and CART. Results demonstrate the efficiency of Bagging. Bagging achieves 14 to 15.34% better classification accuracy than the other classifiers.

6. REFERENCES

[1] Comscore and Kelsey, http://www.shop.org/c/journal_articles/view_article_content?groupId=1&articleId702&version=1.0.

[2] Popescu and O. Etzioni, Extracting product features and opinions from reviews, Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing., (2005), pp. 339-346.

[3] M. Hu and B. Liu, Mining and Summarizing Customer Reviews, Proceedings of the 10th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD-2004), 8 (2004), pp. 168–174.

[4] B. He, C. Macdonald, J. He, and I. Ounis, An Effective Statistical Approach to Blog Post Opinion Retrieval, CIKM., 10 (2008), pp. 1063-1069.

[5] M. Hu and B. Liu, Mining Opinion Features in Customer Reviews, Proceedings of the 19th National Conference on Artificial Intelligence., 7 (2004), pp. 755-760.

[6] S. Kim, P. Pantel, T. Chklovski, and M. Pennacchiotti, Automatically Assessing Review Helpfulness, EMNLP., 7 (2006), pp. 423-430.

[7] B. Liu, M. Hu, and J. Cheng, Opinion Observer: Analyzing and Comparing Opinions,WWW., 5 (2005), pp. 342-351.

[8] B. Liu, Sentiment Analysis and Subjectivity, to appear in Handbook of Natural Language Processing, Second Edition, 2010.

[9] Ana-Maria Popescu and Oren Etzioni. Extracting product features and opinions from reviews. In Proceedings of EMNLP 2005, pp.339-346.

[10] Michael Gamon, Anthony Aue, Simon Corston-Oliver and Eric Ringger. 2005. Pulse: Mining customer opinions from free text. In Proceedings of IDA 2005, pp.121-132.

[11] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In Proceedings of ACM-KDD 2004, p.168-177.

[12] Bing Liu, Minqing Hu and Junsheng Cheng. Opinion Observer: Analyzing and comparing opinions on the web. In Proceedings of WWW 2005, pp.342-351.

[13] Bo Pang and Lillian Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In Proceedings of ACL 2005, pp.115-124.

[14] Whitelaw, Casey, Navendu Garg, and Shlomo Argamon. 2005. Using appraisal groups for sentiment analysis. In Proceedings of the 14th ACM International Conference on Information and Knowledge Management (CIKM-2005), pages 625–631.

[15] Kennedy, Alistair and Diana Inkpen. 2006. Sentiment classification of movie reviews using contextual valence shifters. Computational Intelligence, 22(2):110–125.

[16] Tony Mullen and Nigel Collier. Sentiment analysis using support vector machines with diverse information sources. In Proceedings of EMNLP 2004, pp.412-418.

[17] Bo Pang and Lillian Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of ACL 2004, pp.271-278.

[18] Bo Pang, Lillian Lee and Shivakumar Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of EMNLP 2002, pp.79-86.

[19] Domingos P, Pazzani M (1997) On the optimality of the simple Bayesian classifier under zero-one loss. Mach Learn 29:103–130

[20] Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Wadsworth, Belmont

[21] Breiman, L.: Bagging predictors. Machine Learning 24, 123–140 (1996).