Sentiment Reviews Classification using Hybrid Feature Selection

K. Bhuvaneswari¹ and R. Parimala²

Assistant Professor, Department of Computer Science, Government Arts College, Kulithalai, Tamilnadu, India¹ Assistant Professor, P.G & Research Dept of Computer Science, Periyar E.V.R. College, Trichy, Tamilnadu, India² Email: bhuvaneswarik27@gmail.com¹ and rajamohanparimala@gmail.com²

Abstract

In recent years there has been a steady increase in interest from brands, companies and researchers in Sentiment Analysis and its application to business analytics. It is the process of determining the emotional tone behind a series of words, used to gain an understanding of the attitudes, opinions and emotions expressed within an online mention. Sentiment analysis is a feature of text analysis and natural language processing (NLP) research that is increasingly growing in popularity as a multitude of use-cases emerges. Lexicon based and Machine learning is the two methods used for analysis the sentiments from the content. The proposed feature selection model Ssentiment Reviews Classification using Hybrid Feature Selection (SRCHFS) that extract synsets feature set coupled with Correlation feature selection method can improve the performance of sentiment classification. Nouns, verbs, adjectives and adverbs are organized into synsets, each representing one underlying lexical concept. A set of cognitive synsets is selected using WordNet based POS (Part Of Speech). Support Vector Machine (SVM) classifier is used for sentiment classification on a data set of Movie reviews, Multi Domain product reviews, Amazon Cell phone reviews and Yelp Restaurant reviews. The experimental outcome might result into better accuracy with the existing studies.

Keywords: Sentiment Analysis, Classification, Support Vector Machine, Feature Selection, WordNet.

1. Introduction

Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service and classify review as 'positive', 'negative' or 'neutral' polarity. The applications for sentiment analysis are endless. Its applications have spread to almost every possible domain to improve their business. Sentiment analysis can be categorised into Document Level, Sentence Level and Aspect Level. The document level sentiment analysis is for a specific set of documents, a sentiment analysis classifies each document into one of the two classes either positive or negative. The simplest method uses a bag-of-words model. The features of the document vectors generated from the bag-of-words approach represent the occurrences of each word in a document. In order to overcome dimensionality problem, this study proposes a SRCHFS model.

Dimension of feature can be reduced using any feature selection model. Reduced feature vector generated after the feature selection method is used by the classifier. Support Vector Machines (SVMs) are a relatively new method based on the principle of statistical learning theory to solve classification and regression problems. Tan *et al.*, concluded that SVM performs better than other classifier for sentiment classification [10]. The SRCHFS model uses a dictionary based lexicon method to generate a set of sentiment words. A feature set containing adverbs, adjectives and verbs are extracted using

WordNet dictionary. Redundant features are removed from the feature set using correlation based feature selection. The main objective of this study is to improve sentiment classifier accuracy with dimensionality reduction. The SRCHFS model is experimented with publically available multi domain product dataset, UCI machine repository Amazon dataset, Yelp restaurant dataset and polarity Movie review dataset.

The paper is organized as follows: Section 2 provides the details of related work in Sentiment Analysis. The methodology is mentioned in Section 3. Section 4 discusses the results of SRCHFS model. Section 5 concludes the paper.

2. Related Work

Pang *et al.*, performed document level sentiment classification using Naïve Bayes, Maximum Entropy and SVM techniques for movie reviews and obtained 82.90% accuracy for SVM using three fold cross validation for unigram features [2]. Hu and Liu proposed the new technique for generating Feature Based Summarization system (FBS) of customer reviews of products and also generated opinion based summary as either positive or negative opinion using review sentences and extracting semantic oriented adjective words [4]. Pang and Lee obtained 86.4% of accuracy of document level sentiment classification of the movie reviews using text categorization techniques and to extract subjective portions of the document by applying Naïve Bayes classifier [5]. Shotaro Matsumoto *et al.*, applied syntactic relations between words in sentences for document sentiment classification [6]. The model extracted frequent word sub-sequences and dependency sub trees sentences in a document and used them as features of support vector machines where the model have achieved high rate of accuracy using movie review dataset.

Chaovalit and Zhou compared supervised and unsupervised algorithm for classification and obtained 83.54% of accuracy for supervised method and 77% of accuracy for unsupervised method [7]. Li and Zong proposed fusion of two approaches, feature level and classifier level using multi domain sentiment classification and obtained 79% and 84.50% of accuracy using Books and DVD reviews respectively [9]. O'Keefe and Koprinska proposed a new technique to select features using attribute weights and applied NB and SVM classifiers [11]. The classifier results with 87.15% classification accuracy using only 29% of the selected attributes. Rushdi *et al.*, explored the Sentiment Analysis task and carried 3-fold and 10-fold cross validations in SVM for Pang Movie review corpus [12]. Isabella and Suresh used movie reviews for sentiment analysis and evaluated a range of feature selectors to improve the performance of the classifiers systematically [14].

Kalaivani and Shanmuganathan applied SVM, NB and KNN algorithm for sentiment classification using movie reviews. They used 3-fold cross validation and obtained accuracy of 81.45% of by using the SVM classifier [16]. Mouthami *et al.*, implemented a new algorithm called Sentiment Fuzzy Classification Algorithm to improve classification accuracy of Movie review dataset [13]. Anitha and Bhargavi implemented a document-level sentiment analysis to extract adverb and adjective features for improves the accuracy of classification by using SentiWordNet for calculating the score of a word [15]. The results show Naive Bayes and SVM classifiers are better than SentiWordNet approaches.

Gautami Tripathi and Naganna investigated different feature selection methods to obtain the results for sentiment analysis using NB and Linear SVM classification algorithms for unigrams, bigrams, trigrams and four grams [18]. In this work and in related references it was observed that Linear SVM has high accuracy. Siddhartha Ghosh *et al.*, discussed the concept of polarity values in sentiment analysis using movie reviews. The polarity movie review dataset from Bo Pang and Lillian Lee used for sentiment classification and Naïve Bayes classifier is applied and obtained the accuracy of 71% for 10 validations and 70.50% for 50 validations [19]. Kotzias *et al.*, proposed a new

approach to the problem of predicting labels for sentences given labels for reviews, using a convolution neural network to infer sentence similarity and got 78.16% [20]. Abinash Tripathy *et al.*, extracted all features and converted each feature to numerical vectors using Movie review dataset. The vector of features results 89.5% of accuracy using Naïve Bayes classifier [17].

Wu and Huang proposed new domain adaptation approach which maps sentiments from multiple source domains to target domain using sentiment graphs [21]. Sanju and Mirnalinee extracted unigram and bigram features from the reviews and selected best opinion features by computing relevance score and obtained 78% and 78.85% of accuracy using DVD and Books reviews respectively [22]. Benito Alvares *et al.*, used sentence level classification of reviews using POS tagging and feature pruning by extracting opinion words using opinion sentences and generate opinion summary using a clustering algorithm [23].

In this study, the SRCHFS model focuses to improve an accuracy of sentiment classification of different datasets using dictionary based WordNet with correlation feature selection by applying the SVM classifier approach.

3. Approaches and Techniques for SRCHFS model

Sentiment analysis refers to the task of identifying the opinion of the users from reviews. Sentiment analysis techniques can be classified into supervised machine learning and unsupervised lexicon based semantic orientation approaches [25]. Machine learning classification methods are applied to predict the polarity of sentiments based on training and test data sets. In contrast, the semantic-orientation approach does not require prior training data.

3.1. Lexicon Based Approach

The lexicon oriented semantic approaches to do classification based on positive and negative sentiment words and phrases and mining the data requires no prior training. Two types of techniques are used: corpus-based and dictionary-based. The corpus-based techniques find co-occurrence patterns of words to determine their sentiments using large corpora to set of opinion words. Researchers have proposed different strategies to determine sentiments. Turney calculated a phrase's semantic orientation to be the mutual information between the positive polarity words minus the mutual information between the negative polarity words [1]. Riloff Wiebe used a bootstrapping process to find out subjective expressions from objective expressions [3]. Dictionary-based techniques make use of synonyms, antonyms and hierarchies in WordNet determine word sentiments. It uses a predefined dictionary of words. The dictionary used may be WordNet or SentiWordNet or other. SentiWordNet (SWN) is a lexical resource for recent sentiment classification studies that has more sentiment-related features [24]. It assigns to each synset of WordNet three sentiment scores regarding positivity, negativity, and objectivity range from 0 to 1. The score is automatically allotted from the WordNet. It uses a semi supervised learning method and an iterative random walk algorithm. The corpus-based techniques useful on a large corpus to calculate the statistical information needed to decide the sentiment orientation for each word or phrase.

3.2 Preprocessing

The dataset consists of unrelated and redundant information. Several preprocessing steps are applied on the available datasets to optimize it for further experimentations. Tokenization is used to split the text into a sequence of tokens using unigrams. The stop words are removed and then length based filtration scheme is applied for reducing the generated token set and tokens with less than 3 characters and more than 15 characters are discarded.

International Journal of Database Theory and Application Vol.10, No.7 (2017)

3.3 Feature Extraction

Feature Extraction is the process of extracting relevant features. In the existing research on sentiment analysis considered as all speech words are features. The proposed SRCHFS model retrieves only three parts of words as features. The verbs, adverbs and adjectives play an important role in opinions. The WordNet dictionary is used to perform tagging and extracts all the Verbs (V), Adverbs (A), Adjectives (AJ) and their combinations Adverbs + Adjectives (AAJ), Adverbs + Verbs (AV), Adverbs + Adjectives + Verbs (AJV) and Adjectives + Verbs (AJV) as sentiment features using Movie, Restaurant, Cell phone and Product reviews dataset.

3.4 Correlation based Feature Selection

Feature selection is the process of selecting a subset of relevant features. Filter based feature selection methods apply a statistical measure to provide a ranking of all features by the score and the best features are selected from the dataset. The filter based correlation feature selection method is applied to select most important features from the extracted sentiment words. A correlation is a relationship between features or data attributes. Features may be correlated with one another or redundant. Different combinations of features Adverbs + Adjectives + Correlation (AAJC), Adverbs + Verbs + Correlation (AVC), Adverbs + Adjectives + Verbs + Correlation (AAJVC) and Adjectives + Verbs + Correlation (AJVC), are selected by applying correlation weight which are having highest values. The correlation of an each attribute is computed with respect to the label attribute using Eqn. (1).

Correlation =
$$\frac{\sum (X - \bar{X})(Y - \bar{Y})}{(n-1)S(X)S(Y)}$$
(1)

where X and Y are two attributes with mean values of \overline{X} and \overline{Y} , and standard deviations S(X) and S(Y).

3.5. Machine Learning Approach

Machine learning approaches include Support Vector Machine (SVM), Naïve Bayes, Maximum Entropy, Decision Trees, Random Forest, etc. Machine Learning algorithm has generated a model of the training data, used to classify new unlabeled documents automatically. In this study SVM classifier is used to classify sentiment review datasets.

3.5.1 SVM Classifier

SVM is a new paradigm of learning system. The resulting classifier called the maximal margin classifier. . SVM is a supervised learning method used to analyze the data and recognize data patterns that can be used for classification and regression analysis. The aim of the SVM classifier is that finding the optimal hyperplane that maximizes the margin between the decisions using two classes. The vectors that identify the hyperplanes are the support vectors which is used to find the label of the test data whose label is unknown, datasets..

The power of SVMs lies in their ability to transform data to a higher dimensional space and construct a linear binary classifier in the higher dimensional space. A linear hyperplane in the higher dimensional space transforms to a complex nonlinear decision region in the input feature space. Let x be a set of input feature vectors, and y be the class labels for the feature vectors, this can be represented as tuples

 $\{x_i, y_i\}$ where $i = \{1, ..., J\}$ and $y = \pm 1$. The points lying on the decision surface satisfy Eqn. (2).

$$w \bullet x + b = 0 \tag{2}$$

where w is the normal to the decision region, and b is the distance of the hyperplane from the origin. The generalization power of SVMs lies in its ability to work as a soft decision classifier.

$$\begin{split} x_i \bullet w + b \ge 1 - \xi_i \ for \ y_i &= +1 \\ x_i \bullet w + b \ge 1 + \xi_i \ for \ y_i &= -1 \\ \xi_i \ge 0 \quad \forall i, \end{split}$$

where ξ_i is a slack variable.

The decision function is of the form

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b$$

where α is the lagrangian multiplier.

3.6 Classifier Measure

A confusion matrix is a table that is often used to describe the performance of a classifier on a set of test data for which the true values are known. True positive (TP) is the number of positive sentiment detected when it is actually positive sentiment. True negative (TN) is the number of negative detected when it is actually negative. Classifiers have long been evaluated on their accuracy only.

| | Actual Class | | | | | | |
|-----------|--------------|----------------|----------------|--|--|--|--|
| Predicted | | Positive | Negative | | | | |
| Class | Positive | True Positive | False Positive | | | | |
| | | (TP) | (FP) | | | | |
| | Negative | False negative | True Negative | | | | |
| | | (FN) | (TN) | | | | |

Figure 1. Confusion Matrix

The accuracy is calculated by using Eqn. (3)

Accuracy =
$$\frac{TP+TN}{TP+TN+FP+FN}$$
 (3)

4. Methodology

The flow diagram of SRCHFS model is given in Figure 2. The first step of SRCHFS model is to collect review corpus. Review Corpus is represented in unstructured format. The unstructured format is transformed into structured format. Tokenization is the task of chopping document reviews up into pieces, called tokens, perhaps at the same time throwing away certain characters, such as punctuation, stop word, strip white spaces and remove numbers. The sentiment words are extracted using WordNet dictionary and Term Frequency - Inverse Document Frequency (TF-IDF) word vector is created. The correlation based feature selection is applied to the Document Term Matrix. The redundant features are removed from Document Term Matrix using Correlation based feature selection for different threshold. The reduced feature subset is applied to the SVM

classifier. The confusion matrix is tabulated in Figure 1 and is used to evaluate the performance of sentiment classification.



Figure 1. Flow Diagram of SRCHFS Model

The algorithm is given below.

Algorithm SRCHFC

- i) Read sentiment review corpus; perform preprocessing such as tokenization, removing stop words and filter tokens by length.
- ii) Apply WordNet dictionary to extract bag of words containing Adjectives, Adverbs, Verbs and their combinations.
- iii) Create TF-IDF word vector for extracted sentiment features.
- iv) Apply Correlation on document term matrix.
- v) Sort the features in descending order of correlation value.
- vi) Select top n% of features with highest correlation values.
- vii) Split reduced feature subset obtained from Step v into Training set and Test set
- viii) Train SVM classifier on Training set.
- ix) Evaluate the classifier measures on Test set.

5. Results and Discussions

5.1. Dataset Used

Dataset 1: This dataset was prepared by Pang and Lee in order to classify movie reviews [5]. The reviews are collected from Internet Movie Database (IMDB) review site available at http://www.cs. cornell.edu /people/pabo/movie-review-data.

Dataset 2: This is a Multi-Domain dataset created by Blitzer *et al.*, contains reviews for distinct product categories like Books, DVD, Electronics, Kitchen Appliances, camera [8]. The proposed model uses 1000 positive and 1000 negative reviews for each data set and each review consists of a plain text file. The model uses 1600 reviews (800 positive reviews and 800 negative reviews) for training the classifier and 400 reviews (200 positive and 200 negative) for testing the classifier.

Dataset 3: It contains reviews from Amazon and Yelp dataset from the UCI, the machine learning repository (http://archive.ics.uci.edu/ml). The Yelp dataset consists of Restaurant reviews and Amazon dataset contains Cell Phone reviews. This dataset was created for the paper "From Group to Individual Labels Using Deep Features," by Kotzias *et al.*, for KDD 2015 [20]. The dataset consists of 1000 labeled reviews equally divided into 500 positive and 500 negative reviews. Those were selected randomly for larger datasets of reviews and the goal was for no neutral sentences to be selected. The model uses 400 positive reviews and 400 negative reviews for training the classifier and 100 positive and 100 negative reviews for testing the classifier.

5.2. Experimental Setup

The SRCHFS model uses Rapid Miner Studio software with its text processing extension, web mining and WordNet extension. This model is implemented using SVM classifier by grouping various combinations of sentiment features using hybrid feature selection. First, the data set is preprocessed and Term Frequency –Inverse Document Frequency (TF-IDF) matrix is created. The WordNet dictionary is used to extract sentiment features like adjectives, adverbs, verbs and their combinations. Correlation based feature selection method is employed to select a combination of sentiment features using correlation value. SVM classifier applies to the reduced dataset and Split validation is applied and performance measures are evaluated.

5.3 Classification and Evaluation

The SRCHFS model is evaluated using three different datasets by applying the SVM classifier. The experiment shows that WordNet feature extraction gives better accuracy, using a combination of sentiment features AAJ, AV, AAJV and AJV than adverbs, adjectives and verbs alone. The SRCHFS model is implemented using WordNet to extract adjectives, adverbs, verbs and represented as sentiment features. Table 1 summarizes the performance of classification accuracy (Acc) of Dataset 1, Dataset 2 and Dataset3 using all features, combination of sentiment features and correlation value. The top most 30% of sentiment features are selected by using correlation value as the experiments obtained maximum accuracy for those features.

From the table, the SRCHFS model gives the better accuracy by using the combination of AAJVC for all datasets by applying the SVM classifier. The proposed SRCHFS model classifies the sentiment review as positives and negatives.

| Method | Data | set 1 | | Dataset 2 | | | | | Dataset 3 | | | |
|-----------------|-------|------------|-------|------------|--------|------------|-------|------------|------------|------------|------------|------------|
| Witchiou | Movie | | Book | | Camera | | DVD | | Cell Phone | | Restaurant | |
| | NF | Acc (%) | NF | Acc (%) | NF | Acc (%) | NF | Acc (%) | NF | Acc (%) | NF | Acc (%) |
| All features | 25004 | 78.25 | 14084 | 74.50 | 5881 | 84.75 | 15218 | 71.25 | 1237 | 79.50 | 1455 | 77.00 |
| AAJ | 7523 | 79.50 | 4820 | 69.00 | 2153 | 80.25 | 4627 | 72.00 | 501 | 75.50 | 568 | 77.00 |
| AV | 6292 | 76.50 | 4338 | 70.50 | 2454 | 80.50 | 4419 | 69.50 | 693 | 70.00 | 727 | 69.50 |
| AJV | 9792 | 75.75 | 6640 | 70.50 | 3225 | 81.50 | 6510 | 70.00 | 819 | 79.50 | 898 | 75.50 |
| AAJV | 10973 | 77.25 | 7311 | 72.00 | 3559 | 82.50 | 7165 | 69.50 | 906 | 79.00 | 981 | 75.50 |
| AAJC | 2557 | 92.00 | 1446 | 85.25 | 646 | 87.00 | 1388 | 86.00 | 401 | 77.00 | 454 | 81.50 |
| AVC | 1888 | 88.00 | 1301 | 87.00 | 736 | 86.75 | 1326 | 88.00 | 554 | 78.50 | 582 | 71.00 |
| AJVC | 2938 | 89.00 | 1992 | 88.50 | 968 | 90.75 | 1953 | 91.00 | 655 | 86.00 | 718 | 83.50 |

 Table 2. Classification Accuracy (Acc) of Three Datasets using Different

 Features

| AAJVC | 3292 | 92.25 | 2193 | 91.00 | 1068 | 92.00 | 2150 | 92.75 | 725 | 87.50 | 785 | 83.50 |
|---|------|-------|------|-------|------|-------|------|-------|-----|-------|-----|-------|
| *The hold faced values indicate better performance. NE Number of Features | | | | | | | | | | | | |

*The bold-faced values indicate better performance. NF – Number of Features

5.4 Comparative Analysis

The results are compared with other similar works on the same datasets; the results of SRCHFS model are promising. Table 2, 3 and 4 shows the results of proposed model with existing literatures of three datasets and graphical representation are shown in Figure 3, Figure 4 and Figure 5.

Table 2. Comparative Results among Different Literatures Obtained on Dataset 1

| Author & Literature | Classifier Used | Accuracy (%) | |
|---------------------------------|--------------------|--------------|--|
| Anitha & Bhargavi [15] | NB | 72.12 | |
| Pang & Lee [5] | NB | 86.40 | |
| Gautami & Naganna [18] | SVM | 84.75 | |
| O'Keefe & Koprinska [11] | SVM | 87.15 | |
| Kalaivani & Shanmuganathan [16] | SVM | 81.45 | |
| Bhuvaneswari & Parimala | SVM | 92.25 | |
| (SRCHFS) | | | |

Table 3. Comparative Results among Different Literatures Obtained onDataset 2

| Authon & Litonature | Classifier | Accuracy (%) | | | |
|-------------------------|------------|--------------|-------|--------|--|
| Author & Literature | Used | DVD | Book | Camera | |
| Li & Zong [9] | SVM | 84.50 | 79.00 | - | |
| Wu & Huang [21] | SVM | 79.13 | 78.29 | - | |
| Sanju & Mirnalinee [22] | SVM | 78.00 | 78.85 | - | |
| Bhuvaneswari & Parimala | SVM | 92.75 | 91.00 | 92.00 | |

Table 4. Comparative Results among Different Literatures Obtained onDataset 3

| | | Accuracy (%) | | | |
|-------------------------|------------------------|--------------|------------|--|--|
| Author & Literature | Classifier Used | Restaurant | Cell Phone | | |
| Kotzias et al., [20] | Logistic Regression | 78.16 | - | | |
| Bhuvaneswari & Parimala | SVM | 83.50 | 87.50 | | |

International Journal of Database Theory and Application Vol.10, No.7 (2017)



Figure 3. Comparison Results on Dataset 1



Figure 4. Comparison on Dataset 2



Figure 5. Comparison on Dataset 3

Table 2 and Figure 3 illustrates that the SRCHFS model gives maximum accuracy of 92.25% by applying a combination of adjectives, adverbs, verbs sentiment features with a correlation threshold (AAJVC) for Dataset 1. From Table 3 and Figure 4, it is observed that the model gives better accuracy of 91.00 % for Books reviews, 92.75% for DVD reviews and 92.00 % for camera reviews than existing models. Table 4 and Figure 5 prove that the SRCHFS model shows better accuracy of 83.50% for restaurant reviews and 87.50% of cell phone reviews.

6. Conclusion

The hybrid feature selection scheme for document level sentiment classification of various dataset is proposed. The experimental work makes two important contributions. One is the use of combination of sentiment words as AAJ, AV, AAJV and AJV to select

sentiment features explored for document level sentiment classification of a review. Another one is feature selection using combination of sentiment words with correlation value (AAJC, AVC, AAJVC and AJVC). The results show that AAJVC combination improves the accuracy of sentiment classification. This research assessed the use of the WordNet opinion lexicon in the task of sentiment classification of reviews. Future work should focus on improvise the accuracy of sentiment classification.

References

- [1] P. D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", Proceedings on 40th Annual Meetings of the Assoc .Computational Linguistics, ACL Press, (2002), pp. 417–424.
- [2] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques", Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, vol.10, (2002), pp. 79-86.
- [3] E. Riloff and . Wiebe, "Learning Extraction Patterns for Subjective Expressions, in Proceedings of Conference on Empirical Methods in Natural Language Processing", ACL Press, (2003), pp. 105–112.
- [4] M. Hu and B. Liu,"Mining and Summarizing Customer Reviews", Proceedings of the 10th ACM SIGKDD, International Conference on Knowledge Discovery and Data Mining, (2004).
- [5] B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis using Subjectivity Summarization Based on Minimum Cuts", Proceedings of 42nd ACL, (2004), pp. 271–278.
- [6] S. Matsumoto, H. Takamura and M. Okumura, "Sentiment Classification using Word Sub-sequences and Dependency Sub-trees", Proceedings of 9th Pacific-Asia Conference, PAKDD 2005, Hanoi, Vietnam, (2005), May 18-20, pp. 301-311.
- [7] P. Chaovalit and L. Zhou, "Movie Review Mining: A Comparison between Supervised and Unsupervised Classification Approaches", In System Sciences, HICSS'05, Proceedings of the 38th Annual Hawaii International Conference on IEEE, (2005), pp. 112c-112c.
- [8] J. Blitzer, M. Dredze and F. Pereira, "Biographies, Bollywood, Boom-Boxes and Blenders: Domain Adaptation for Sentiment Classification", 45th Annual Meeting of the Association of Computational Linguistics, (2007), pp. 440–447.
- [9] S. Li and C. Zong, "Multi-domain Sentiment Classification", Proceedings of Association for Computational Linguistics -08: HLT, Short Papers (Companion Volume), (2008), pp. 257–260.
- [10] S. Tan and J. Zhang, "An empirical study of sentiment analysis for chinese documents", Expert Systems with Applications (2008), pp. 2622–2629.
- [11] T. O'Keefe and I. Koprinska, "Feature Selection and Weighting in Sentiment Analysis", Proceedings of the 14th Australasian Document Computing Symposium, Sydney, Australia. (2009).
- [12] M. Rushdi Saleh, M.T. Martín Valdivia, A. Montejo-Ráez and L.A Urena-Lopez, "Experiments with SVM to Classify Opinions in Different Domains", Expert Systems with Applications, (2011), pp. 14799-14804.
- [13] K. Mouthami, K.N. Devi and V.M. Bhaskaran, "Sentiment Analysis and Classification based on Textual Reviews", Information Communication and Embedded Systems International Conference on IEEE, (2013), pp. 271–276.
- [14] J. Isabella and R.M. Suresh, "Analysis and Evaluation of Feature Selectors in Opinion Mining", Indian Journal of Computer Science and Engineering, vol. 3, no. 6, (2013), pp. 757-762.
- [15] B.M. Anitha and B.R. Bhargavi, "Opinion Classification Based on Verb, Adverb and Adjectives: Using Various Supervised Machine Learning Algorithms", Proceedings of International Conference on Multimedia Processing, Communication and Information Technology, MPCIT, (2013).
- [16] P Kalaivani and K.L. Shanmuganathan, "Sentiment Classification of Movie Reviews By Supervised Machine Learning Approaches", Indian Journal of Computer Science and Engineering, vol. 4, no. 4, (2013).
- [17] A. Tripathy, A. Agrawal and S. K. Rath, "Classification of Sentimental Reviews using Machine Learning Techniques", Proceedings 3rd International Conference on Recent Trends in Computing, (2015), pp. 821-829.
- [18] T. Gautami and S. Naganna, "Feature Selection and Classification Approach for Sentiment Analysis", Machine Learning and Applications:, An International Journal (MLAIJ), vol.2, no.2, (2015).
- [19] S. Ghosh, S. M. Thamke and U.R.S. Kalyani, "Sentiment Analysis using Rapid Miner for Polarity Dataset", International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169, vol. 3, no. 8, (2015), pp. 5167 – 5172.
- [20] D. Kotzias, M. Denil, N. De Freitas and P. Smyth, "From Group to Individual Labels using Deep Features, KDD 2015, (2015).
- [21] F. Wu and Y. Huang, "Sentiment Domain Adaptation with Multiple Sources, Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, (2016), pp. 301–310

- [22] P. Sanju and T.T. Mirnalinee, "Cross Domain Sentiment Classification By Extracting Best Opinion Features, Australian Journal of Basic and Applied Sciences, vol. 10, no. 2, (2016), pp. 192-198.
- [23] B. Alvares, N. Thakur and S. Patil, "Sentiment Analysis using Opinion Mining, International Journal of Engineering Research and Technology, vol. 5, no. 4, (2016).
- [24] R. Priyanka, Patil, Pratibha. S and Yalagi, "Sentiment Analysis of movie reviews using SentiWordNet Approach", International Journal of Engineering Trends and Technology, vol. 38, no. 5, (2016).
- [25] T. Hardeniya, A. Dilipkumar and Borikar, "Dictionary Based Approach to Sentiment Analysis A Review", International Journal of Advanced Engineering, Management and Science (IJAEMS), vol.2, no. 5, (2016).

Authors



K. Bhuvaneswari is a Research Scholar and currently working as Assistant Professor in Government Arts College, Kulithalai. She received her Master of Computer Applications (MCA) in 2000 and M.Phil Computer Science in 2005 from Bharathidasan University, Tiruchirappalli. Her area of research focuses on Web Mining.



R. Parimala graduated with M.Sc. Applied Science at the National Institute of Technology (formerly Regional Engineering College) Tiruchirappalli in 1990. She received her M.Phil Computer Science at Mother Teresa University, Kodaikanal in 1999. She started teaching in 1999 at National Institute of Technology and is currently working as Assistant Professor in Department of Computer Science, Periyar E.V.R. College (Autonomous), Tiruchirappalli. She completed her Ph.D. at National Institute of Technology, Tiruchirappalli. Her area of research interests includes Neural Networks, Data Mining and Optimization Techniques.

International Journal of Database Theory and Application Vol.10, No.7 (2017)