ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

Correlation Based Feature Selection for Movie Review Sentiment Classification

K. Bhuvaneswari[1], Dr. R. Parimala[2]

Assistant Professor, Department of Computer Science, Government Arts College, Kulithalai, Tamilnadu, India[1]

Assistant Professor, Department of Computer Science, Periyar E.V.R. College, Trichy, Tamilnadu, India[2]

Abstract: Sentiment Analysis is one of the recent research areas in Data Mining concepts and Natural Language Processing techniques. It retrieves users or customer reviews from the web and classify the reviews using sentiment analysis approach. This paper proposes a method for sentiment classification using correlation based feature selection. First, different levels of data pre-processing techniques applied on the labeled polarity movie review dataset results in structured documents with Bag of Words. Second, correlation attribute method is used for feature selection to identify most important features. Finally, the two popular classifiers namely Naive Bayes(NB) and Support Vector Machine(SVM) are implemented and evaluated various performance measures of sentiment analysis. The proposed model concludes with the better results of accuracy using SVM classifier.

Keywords: Sentiment Analysis, Opinion Mining, Correlation, Naive Bayes, Support Vector Machine.

I. INTRODUCTION

In recent days, the people are expressing their sentiments, opinions, reviews, feedback on the web using different forums, blogs, social websites twitter, facebook, etc. A number of websites allow Internet users to submit movie reviews and scores. The recent trends in Internet that encourages users to contribute their opinion and suggestion created a huge Corpus of valuable information in the web. There is a need to analyze the user's opinion either positive or negative. Movie review classification plays an important role in providing classification decision on movies. In recent days, every week number of movies is released and large number of websites provides opinions or comments about movies. People before going for movie they want to know comments on that movie. By reading all reviews it is very difficult to find out it expresses whether it is positive or negative sentiments. For that situations sentiment analysis techniques are used to classify sentiments from text data in their appropriate class either positive or negative.

In this paper, the movie review dataset is used for document level sentiment classification because special challenges are associated with movie reviews. Movie review classification is different from other topic-based classification because it based on domain specific and semantic words [14]. The proposed model mainly concerns with supervised learning techniques on a labeled movie reviews benchmark dataset created by Pang and Lee [16] and freely available on the Internet. Opinion in Sentiment analysis classified at three levels namely sentence level, document level and aspect or feature level [30]. Document level sentiment classification is used to classify the whole document contains as positive or negative reviews. Sentence level sentiment classification considers classification of reviews at individual sentence. The feature level classifications first extract the important features from document and then classifies whether it is either positive or negative opinions. Machine learning algorithms applied to classify and predict whether a document represents positive or negative sentiment. Machine learning algorithm categorized as supervised and unsupervised machine learning algorithms. In general supervised classification algorithms [7] has proved effectively and widely used in sentiment classification.

Specifically, the proposed model uses Naïve Bayes (NB) and Support Vector Machine (SVM) classifiers for classifying sentiments, finds the results, and compares with the existing results. This paper is organized as follows: Section II presents the related work, Section III describes the detailed methodology of the proposed model, and Section IV discusses the experimental results of proposed model. Finally, Section V concludes the paper along with scope for future work.

II. RELATED WORK

Isabella et al., [1] used movie reviews for sentiment analysis and evaluated a range of feature seletors to improve the performance of the classifers systematically. O'Keefe et al., [2] proposed a new technique to select features using attribute weights and applied NB and SVM classifiers. The author obtained 87.15% classification accuracy using only 29% of the selected attributes. Pang et al.,[3] used labelled sentences in the document as subjective or objective and applied machine learning classifier to the subjective group which prevents polarity classification from considering useless and misleading data. They have explored extraction of methods on the basis of minimum cut formulation. Abinash Tripathy et

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

al., [4] applied NB and SVM Machine algorithms for classifying sentiments and observed SVM classifier was performed well. Mullen et al., [5] implemented a hybrid SVM approach by making use of potentially relevant information and achieved better result than existing models. Gautami Tripathi et al., [6] 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. In this work and in related references it was observed that Linear SVM has high accuracy also they proposed model for sentiment analysis using higher order n-grams. Pang et al., [7] performed document level sentiment classification using Naïve Bayes, Maximum Entropy and SVM techniques with unigrams and bigrams features and also achieved 82.90% accuracy using three fold cross validation for unigrams. Ding et al., [8] implemented Opinion Observer System that handled implicit opinions for feature indicator and considered semantic orientation of an opinion word based on reviews and combine multiple opinion words in the same sentence. Murthy G et al., [9] proposed a method to study sentiments using comparative sentences sentiments and deals context based sentiments by using web information.

Hu et al., [12] used frequent item sets to extract the most relevant features subset of nearby adjectives. In addition, the author implemented the classification of opinion words as either positive or negative using WordNet package. Dr. Siddhartha Ghosh et al., [13] discussed the concept of polarity in sentiment analysis in the dataset. The polarity movie review dataset from Bo Pang and Lillian Lee used for sentiment classification. Naïve Bayes classifier applied and calculated the accuracy for 71% for 10 validations and 70.50% for 50 validations with the Rapid Miner Tool. Prabowo et al., [15] proposed hybrid classification algorithm using rule based classifier with high accuracy using movie reviews, product reviews and MySpace comments by combining a rule-based classifier and supervised learning algorithm.

M. Rushdi et al., [17] explored the sentiment analysis task and carried 3-fold and 10-fold cross validations in SVM for Pang Movie review corpus. Mouthami et al., [22] implemented a new algorithm called Sentiment Fuzzy Classification Algorithm to improve classification accuracy of Movie review dataset. Li et al., [23] proposed active learning approach that combines the active learning strategy and the label propagation algorithm to make the classification decision. Anitha et al., [24] investigated Fine-grained relational topic weighted approach to find the opinion features, non-noun features, and implicit features of the topic and document phrases. Benito Alvares et al., [25] discussed sentence level classification of reviews using POS tagging and feature pruning by extracting opinion words and found opinion sentence orientation. Finally, opinion summary was generated using clustering algorithm.

Swati N. Manke et al., [26] proposed the Opinion feature extraction technique to identify domain dependent opinion features and comments. P.Kalaivani et al., [27] applied SVM, NB and KNN algorithm for sentiment classification using movie reviews. They used 3-fold cross validation and obtained better accuracy of 81.45% by using SVM classifier. Saruladha et al., [28] proposed the Feature-Based Sparse Non- Negative Matrix Factorization method (FS-NMF). The author selected highest weighted features and created weighted term-sentence matrix to group the review sentences into feature relevant clusters. Ahmad Kamal [29] implemented the design of feature-level summarization scheme to visualize mined features, opinions and their polarity values using supervised machine learning techniques for subjectivity and objectivity classification of review sentences. A number of research papers had published presenting innovative techniques and new ideas to perform sentiment analysis [18][19][20][21].

In this study, the proposed model focuses to achieve better accuracy of sentiment classification of movie review dataset using correlation based feature selection method. The NB and SVM classifiers studied for movie review classification.

III. PROPOSED MODEL

DESIGN OF PROPOSED MODEL

This section presents the design and methodology of sentiment classification in movie review polarity data set. In this study, binary sentiment classification technique used to classify the movie reviews or documents into two classes either positive or negative. Figure 1 shows the diagrammatic representation of proposed model.

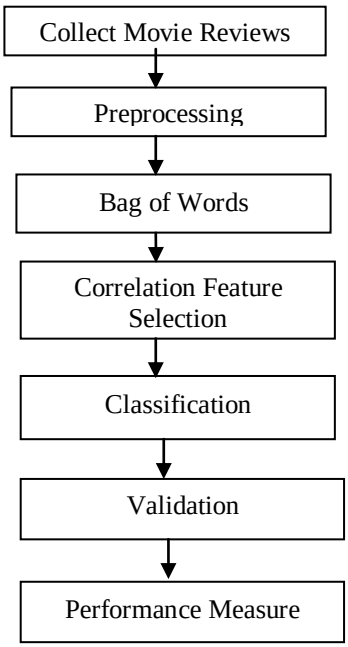

Fig. 1. Proposed Model

The collected movie review data is preprocessed and processed data is converted into Bag of words. The Term Frequency (TF) method measures how frequently a term occurs in a document and the Term Occurrences (TO)

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

define absolute number of occurrences of a term. In Binary Occurrences (BO) the term occurrence is either 0 or 1. Term Frequency - Inverse Document Frequency (TF-IDF) is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. Typically, the TF-IDF weight is composed by two terms: the first term computes the normalized Term Frequency (TF), the second term is the Inverse Document Frequency (IDF), computed as the logarithm of the number of the documents in the corpus divided by the number of documents where the specific term appears.

Feature Selection using Correlation

Feature selection is the process of selecting relevant features. Features may be correlated with one another or redundant. A correlation is a number between -1 and +1 that measures the degree of association between two features. Tightly correlated features are selected. A variable number of features above a preset threshold value of the correlation are selected. NB and SVM conducted on the reduced dataset.

Naïve Bayes Classifier

A NB classifier is a simple probabilistic classifier based on applying Bayes theorem with an assumption of independence among predictors. In simple terms, a NB classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. NB classifier considers a document as a bag of words and assumes that the probability of a word in the document is independent of its position in the document and the presence of other word. For a given text document'd' and for a class 'c' (positive, negative), the conditional probability for each class given a review is P (c|d). According to Bayes theorem this quantity can be computed using the following equation:

$$P(c|d) = \frac{P(d|c) * P(c)}{P(d)}$$

The advantage of the Naive Bayes classifier is that it only requires less computational effort than other classifier.

Support Vector Machine Classifier

SVM are based on the concept of decision planes that defines decision boundaries. The aim of the SVM classifier is that finding the hyperplane that maximizes the margin between the two classes. The vectors that define the hyperplane are the support vectors. In this study, SVM model represents each review in vectorized form as a data point in the space. This method is used to analyze the complete vectorized data and find a hyperplane to train a model. The set of textual data vectors are said to be optimally separated by hyperplane only when it is separated without error and the distance between closest points of each class and hyperplane is maximum. With the hyperplane, the test reviews are predicted to a class based on which side of the hyperplane they fall on. Researchers have achieved better results in SVM classifier. The testing accuracy of the SVM classifier depends on the training object. Cross-validation is used to increase the performance of training data that increases the prediction of movie reviews.

Performance Measure

Confusion Matrix is created to tabulate the performance of any classifier. This matrix shows the relation between correctly and wrongly predicted reviews. In the confusion matrix, TP (True Positive) represents the number of positive movie reviews that are correctly predicted whereas FP (False positive) gives the value for number of positive movie reviews that are predicted as negative by the classifier. Similarly, TN (True Negative) is number of negative reviews correctly predicted and FN (False Negative) is number of negative reviews predicted as positive by the classifier. The confusion matrix formation is shown in Fig. 2.

	Actual Class		
Predicted Class		**Positive**	**Negative**
	Positive	True Positive(TP)	False Positive(FP)
	Negative	False Negative(FN)	True Negative(TN)

Fig.2. Confusion Matrix

This confusion matrix is used to calculate different Performance evaluation parameter like precision, recall and accuracy.

Precision gives the exactness of the classifier. It is the ratio of correct positive observations.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall also known as true positive rate. It measures the completeness of the classifier. Also it is the ratio of correctly predicted positive events.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Accuracy is one of the most common performance evaluation parameter and it is calculated as the ratio of number of correctly predicted reviews to the number of total number of reviews present in the corpus. The formula for calculating accuracy is given as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

IV. RESULTS AND DISCUSSIONS

Experimental Setup

The proposed model uses Rapid Miner 5.3.015 software with its text processing extension. Rapid Miner supports the design and documentation of overall data mining process and machine learning algorithms. This model is implemented using the NB and SVM classifier. First, the data set is preprocessed and the bag of words are created using TF, TO, BO and TF-IDF. Correlation based feature

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

selection is used. NB and SVM Classifier is applied on the reduced dataset. 10-fold Cross validation is applied and performance measures are evaluated.

Dataset Used

The proposed model uses a dataset of classified movie reviews prepared by Pang and Lee (2004). The dataset consists of 2000 user created movie reviews on Internet Movie database available at http://www.cs. cornell.edu /people/pabo/movie-review-data. The reviews are equally partitioned into a positive set and a negative set (1000+1000). Each review consists of a plain text file and a class label representing the overall opinion. The class attribute has only two values pos or neg.

Data Preprocessing

The movie reviews dataset consist of irrelevant and redundant information. Several preprocessing steps are applied on the available dataset to optimize it for further experimentations. Tokenization is used to split the text of a document into sequence of tokens of unigrams. The splitting points are defined using all non letter characters. Then length based filtration scheme was applied for reducing the generated token set. The parameters used to filter out the tokens are the minimum length and maximum length. In the proposed model the minimum length was set to 2 characters and maximum length to 20 characters i.e. tokens with less than 2 characters and more than 20 characters were discarded.

Stop words are removed. Stemming operator is used to stem English words using Porter stemming algorithm applying an iterative, rule based replacement of word suffixes intending to reduce the length of the words until the minimum length is reached. The stemming technique increases the efficiency and effectiveness of the information retrieval and text mining processes. Finally Transform Case operator is used to transform all characters in a document to either lower case or upper case. In the proposed model all characters are converted into lower case letters.

Initially the wordlist generated for the dataset consist of 38911 tokens. The results obtained for the various preprocessing stage are shown in Fig. 3.

Preprocessing	Number of Tokens
Initial Tokens	38911
Filtering By Length	38883
Filtering Stop Words	38530
Stemming	25211

Fig.3. Various Preprocessing Levels

Bag of Word is created using TF, TO, BO and TF-IDF methods. The values of various evaluation parameters are obtained using NB and SVM. Table 1 shows the performance measures of NB and SVM classifier. Also the graphical representations of results are given in Fig. 4 and Fig. 5.

Table 1: Results obtained using NB and SVM Classifier

Vector Creation	Naive Bayes			SVM		
	Accuracy %	Precision %	Recall%	Accuracy%	Precision %	Recall%
TF-IDF	67.35	66.99	69.00	80.00	80.31	79.60
Term Frequency	70.25	71.01	68.60	79.95	81.85	77.10
Term Occurrences	66.85	66.57	68.00	72.30	91.79	48.90
Binary Occurrences	67.95	67.87	68.20	73.45	93.50	50.40

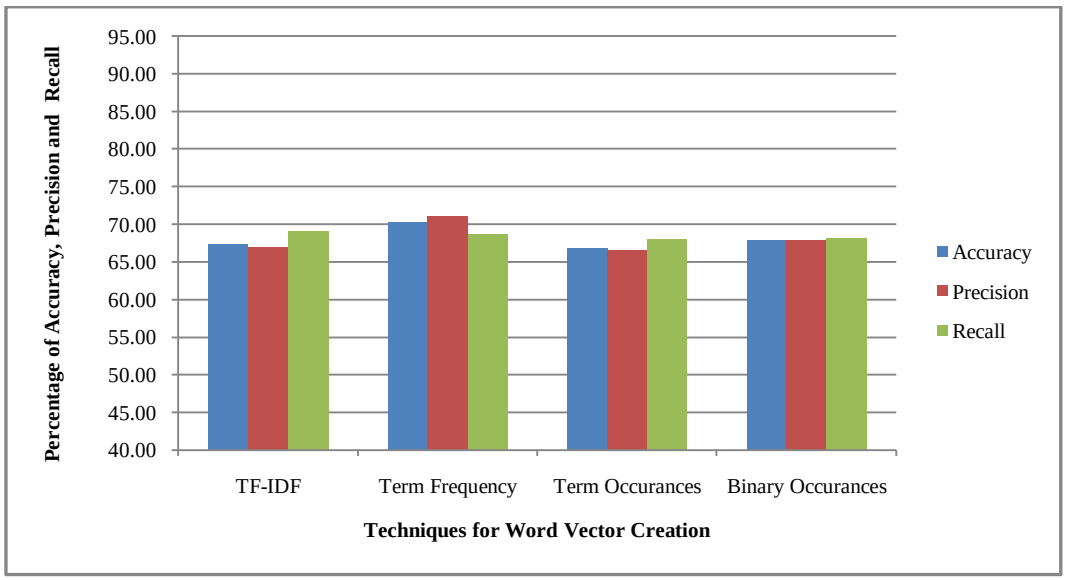

Fig. 4 Naïve Bayes

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

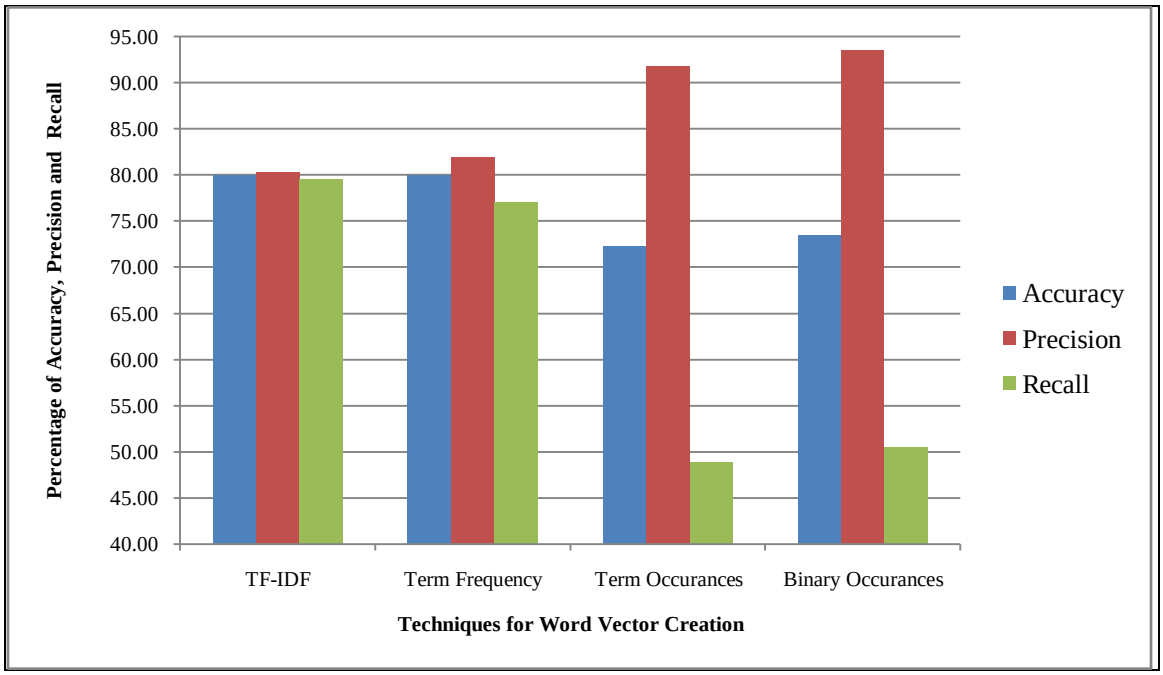

Fig5. SVM

The experiment shows that Term Frequency-Inverse Document Frequency (TF-IDF) scheme gives maximum accuracy for SVM. Term Frequency gives maximum accuracy for NB classifier and it is closely nearer to TF-IDF. So, the proposed model is implemented by using TF-IDF word vector creation method. The existing model is modified by applying correlation weight feature selection scheme using NB and SVM. In the proposed model different correlation weights are applied to select features which are having highest values. For each algorithm the confusion matrix is created and performance measures are obtained. The Confusion Matrix for NB algorithm using highest Correlation Weight (0.2) is shown in Table 2 and the Confusion Matrix for SVM algorithm using highest Correlation Weight (0.5) is shown in Table 3.
The performance measures are calculated using different correlation weights and shown in Table 4.
Also the graphical representations of Results are shown in Fig. 6 and 7.

Table 2: Confusion Matrix obtained by Correlation Weight (0.2) using NB Classifier

Predicted Values	Correct Labels		
		Positive	Negative
	Positive	948	41
	Negative	52	959

Table 3: Confusion Matrix obtained by Correlation Weight (0.5) using SVM Classifier

Predicted Values	Correct Labels		
		Positive	Negative
	Positive	973	32
	Negative	27	968

Table 4: Results of NB and SVM Classifier with Correlation Based Feature Selection

Correlation Threshold	Number of Features	NB			SVM		
		Accuracy %	Precision %	Recall %	Accuracy %	Precision %	Recall %
0.1	2521	92.95	93.34	92.50	93.65	91.15	96.80
0.2	**5042**	**95.35**	**95.88**	**94.80**	96.70	95.74	97.80
0.3	7563	95.05	95.80	94.30	97.00	96.39	97.70
0.4	10084	94.90	95.40	94.40	96.85	96.65	97.10
0.5	**12606**	94.90	95.40	94.40	**97.05**	**96.84**	**97.30**
0.6	15127	94.80	95.48	94.10	96.65	97.01	96.30
0.7	17648	94.00	94.56	93.40	96.15	96.79	95.50
0.8	20169	84.90	85.01	84.80	92.20	92.37	92.10
0.9	22690	75.85	75.94	76.00	86.90	87.81	85.80

DOI 10.17148/IJARCCE.2016.5765

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

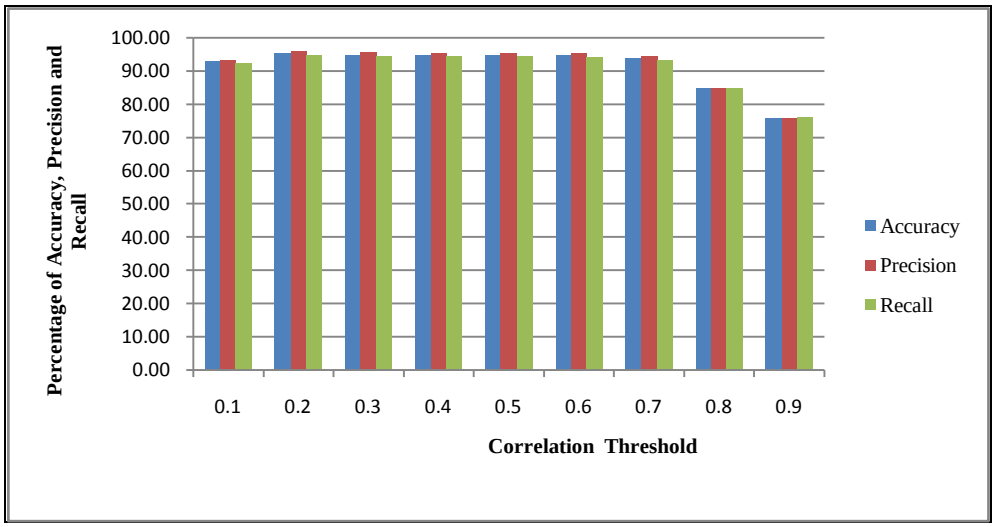

Fig. 6. Performance Measure of. NB classifier with Feature Selection

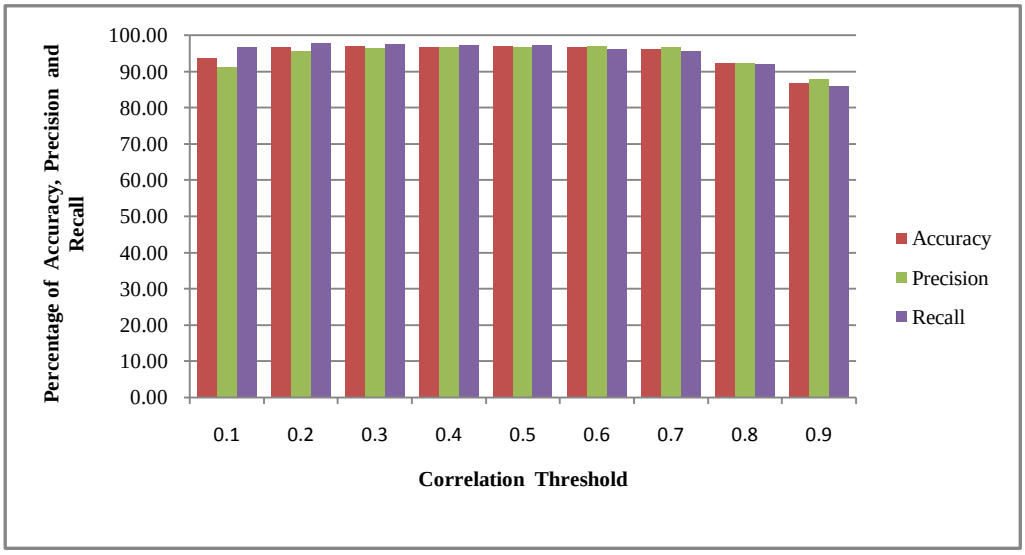

Fig. 7. Performance Measure of. SVM classifier with Feature Selection

From Table 4, Fig. 6 and Fig. 7, the proposed model gives 0.2 correlation weight has maximum accuracy of 95.35% using NB; correlation weight 0.5 has maximum accuracy of 97.05% using SVM. By comparing these values the proposed model achieved maximum accuracy of 97.05% using SVM classifier.

COMPARATIVE ANALYSIS

This section compares the output obtained using the proposed model with the output obtained in existing approaches.

The proposed model is compared with Gautami Tripathi et al.,[6] and Abinash Tripathy et al., [4] models.

Both the models used the same labeled movie review polarity dataset with 1000 positive and 1000 negative reviews.

The following Table 5 shows the comparison of obtained result with other methods and graphical comparison is shown in Fig. 8 and Fig. 9.

Table: 5 Comparison of Proposed Work with Existing Literatures

Various Models	SVM			NB		
	Accuracy %	Precision %	Recall %	Accuracy %	Precision %	Recall %
Gautami's Model	84.75	82.63	88.00	67.50	70.50	66.51
Abinash's Model	94.06	89.00	86.00	89.53	87.00	77.00
Proposed Model	**97.05**	**96.84**	**97.30**	**95.35**	**95.88**	**94.80**

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

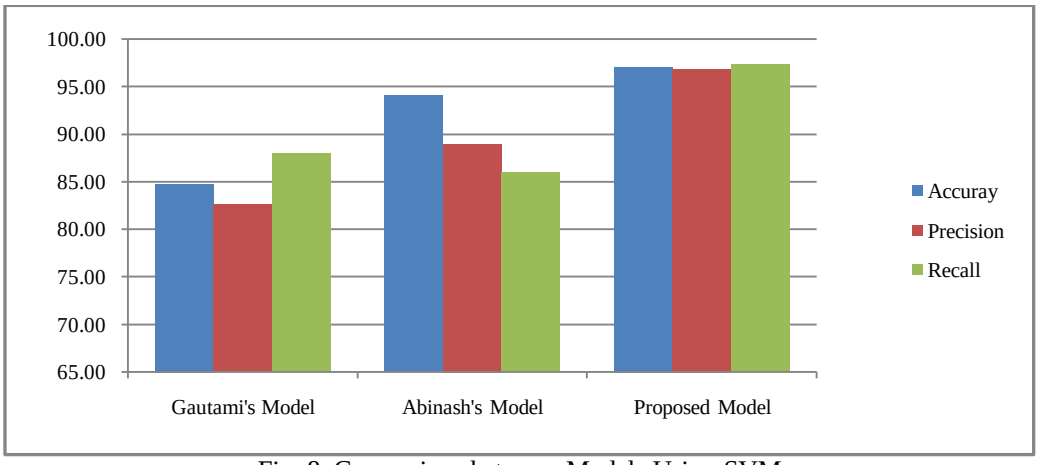

Fig .8. Comparison between Models Using SVM

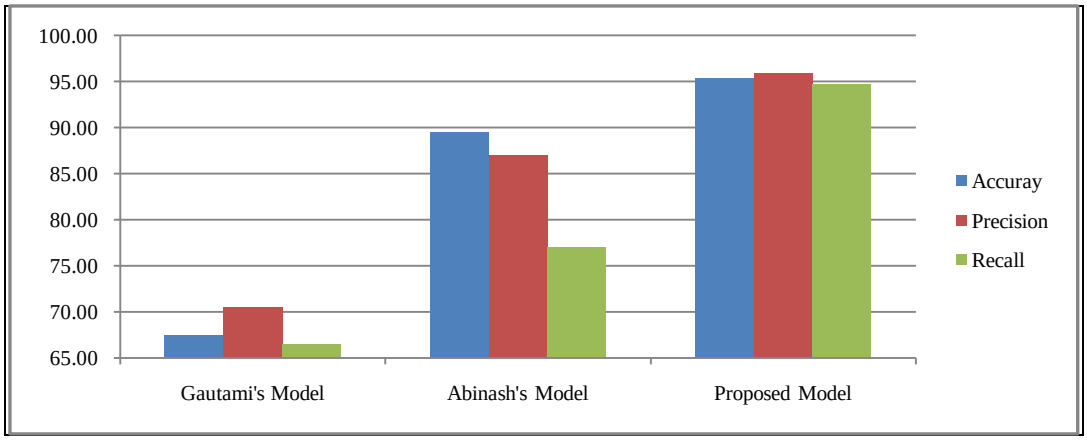

Fig .9. Comparison between Models Using Naive Bayes

From the Table 5 and Fig. 8 and 9, it is observed that the maximum accuracy is obtained using proposed model by comparing accuracy obtained in existing models. Abinash et al., [4] obtained an accuracy of 94.06% for SVM and 87% for NB classification using 10 fold cross validation. Gautami Tripathi et al., [6] achieved an accuracy of 84.75% for SVM and 67.50 for NB classification using 5 fold cross validation for classification. In this proposed model 10 fold cross validation gives maximum accuracy of 97.05% using TF-IDF vector scheme, Correlation weight and SVM classification.

V. CONCLUSION

In this study, an attempt has been made to classify sentiment analysis for movie reviews using machine learning techniques. Two different algorithms, namely, NB and SVM are implemented. These two algorithms have also been implemented earlier by different researchers and results of existing work have been compared. It is observed that SVM classifier outperforms every other classifier in predicting the sentiment of a review. The proposed model presents an approach for sentiment analysis by comparing the two classification methods in combination with correlation weight feature selection schemes. Experimental results show that correlation weight selector achieves the best feature subset for classification and gives better accuracy of 97.05% for sentiment movie review data set using SVM classifier.

In this paper, the proposed model is implemented by NB and SVM classifier for single domain using only unigrams. In future, this model can be extended by applying different classification algorithm by combining with different feature selectors and multi domain data set.

REFERENCES

[1] J. Isabella & Dr. R.M.Suresh, " Analysis and Evaluation of Feature Selectors in Opinion Mining", Indian Journal of Computer Science and Engineering, (ISSN : 0976-5166), Dec 2012-Jan 2013, Vol. 3 No.

[2] O'Keefe. T & Koprinska. I, "Feature Selection and Weighting in Sentiment Analysis", in Proceedings of the 14th Australasian Document Computing Symposium, Dec 2009, Sydney, Australia.

[3] B. Pang & L. Lee, "A Sentimental Education: Sentiment Analysis using Subjectivity Summarization Based on Minimum Cuts", In Proc. of the ACL, 2004, pp 271–278.

[4] Abinash Tripathy, Ankit Agrawal, Santanu Kumar Rath, "Classification of Sentimental Reviews Using Machine Learning techniques", in Proceedings 3rd International Conference on Recent Trends in Computing 2015.

[5] Mullen T, Collier N, "Sentiment Analysis Using Support Vector Machines with Diverse Information Sources", in Proceedings of EMNLP-2004, pp. 412–418.

[6] Gautami Tripathi and Naganna S, " Feature Selection and Classification Approach for Sentiment Analysis", Machine

Learning and Applications: An International Journal , June 2015, Vol.2, No.2.
[7] 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, 79-86.
[8] Xiaowen Ding, "A Holistic Lexicon-Based Approach to Opinion Mining", WSDM'08, February 11-12, 2008, Palo Alto, California, USA.
[9] Murthy G. and Bing Liu, "Mining Opinions in Comparative Sentences, Proceedings of the 22nd International Conference on Computational Linguistics , Manchester, August 2008, 241-248.
[10] Jeevanandam Jotheeswaran, "Feature Reduction using Principal Component Analysis for Opinion Mining", International Journal of Compute Science and Technology, May 2012, Volume 3, Issue 5, pp. 118 – 121.
[11] Andrew L. Maas, "Learning Word Vectors for Sentiment Analysis", Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 142-150.
[12] Qi Zhang, Yuanbin Wu, Tao Li, Mitsunori Ogihara, Joseph Johnson, Xuanjing Huang,"Mining Product Reviews Based on Shallow Dependency Parsing", SIGIR '09, Proceedings of the 32nd international ACM SIGIR Conference on Research and Development in Information Retrieval, 2009.
[13] Dr.Siddhartha Ghosh, Sujata M.Thamke, U.R.S Kalyani, " Sentiment Analysis using Rapid Miner for Polarity Dataset", International Journal on Recent and Innovation Trends in Computing and Communication , August 2015, Volume: 3 Issue: 8 , pp. 5167 – 5172, ISSN: 2321-8169.
[14] Pimwadee Chaovalit, Lina Zhou,"Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches" IEEE Proceedings of the 38th Hawaii International Conference on System Sciences, 2005.
[15] Weifeng Pan, Shan Li, "Tag Ontology Automatic Building for Semantic Searching of Services: a Case Study on Mashup Services," Journal of Computers, 2012, Vol 7, No 12, pp. 2979-2986.
[16] Polarity dataset version 2.0, Sentiment Aanaysis Dataset. [Online]. Available: http://www.cs.cornell.edu/people/pabo/Movie-review-data/review polarity.tar.gz
[17] M. Rushdi Saleh, "Experiments with SVM to Classify Opinions in Different Domains, Expert Systems with Applications", 2011.
[18] Mudinas, Dell Zhang and Marh Levene, "Combining Lexicon and Learning Based Approaches for Concept-Level Sentiment Analysis", Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, ACM, 2012, New York, NY, USA, Article 5, pp. 1-8.
[19] Aditya Joshi, Balamurali A R and Rajath Mohanthy, "C-feel-it: A Sentiment Analyzer for Micro Blogs", Proceedings of ACL: Systems Demonstrations, 2011,11, pp. 127–132.
[20] ZhongwuZhai, Bing Liu, Hua Xu and Peifa Jia, "Clustering Product Features for Opinion Mining, WSDM"11, February 9–12, 2011, Hong Kong, China.
[21] ZhongwuZhai, Bing Liu, Lei Zhang, Hua Xu and Peifa Jia, "Identifying Evaluative Sentence in Online Discussions", Association for the Advancement of Artificial Intelligence, 2011.
[22] K. Mouthami, K. N. Devi, and V. M. Bhaskaran, "Sentiment Analysis and Classification Based on Textual Reviews", Information Communication and Embeddedt Systems ,International Conference on IEEE, 2013, pp 271–276.
[23] Li. S, Ju. S, Zhou. G and Li. X, "Active Learning for Imbalanced Sentiment Classification", In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, ACL, 2012, pp. 139-148.
[24] S.Anitha, B.Sujatha, "Identifying Features from Opinion Mining Using Fine-Grained Relational Topic Weighted Approach", Transactions on Engineering and Sciences, (ISSN: 2347-1964), February 2015, Vol.3, Issue 2.
[25] Benito Alvares, Nishant Thakur, Siddhi Patil, "Sentiment Analysis using Opinion Mining", International Journal of Engineering Research & Technology , (ISSN: 2278-0181), April 2016, Vol. 5, Issue 04.
[26] Swati N. Manke, Nitin Shivale, "A Review on: Opinion Mining and Sentiment Analysis based on Natural Language Processing", International Journal of Computer Applications (0975 – 8887), January 2015, Volume 109, No. 4.
[27] P.Kalaivani, Dr. K.L.Shunmuganathan, "Sentiment Classificationb Of Movie Reviews By Supervised Machine Learning Approaches ", Indian Journal of Computer Science and Engineering, ISSN : 0976-5166 , Aug-Sep 2013, Vol. 4 No. 4.
[28] Saruladha. K, Banupriya. D, Nargis Banu. J, "Opinion Summary Generation for Product Reviews", International Journal of Engineering Research & Technology, (ISSN: 2278-0181), March 2014, Vol. 3, Issue 3.
[29] Ahmad Kamal, "Review Mining for Feature Based Opinion Summarization and Visualization", International Journal of Computer Applications 119(17):6-13, June 2015.
[30] Bing Liu, "Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers", 2012.