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

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

Sentiment Classification using Feature Weights

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

Research Scholar, PG & Research Dept. Computer Science, Periyar E.V.R.College, Tiruchirappalli, India[1]

Assistant Professor, PG & Research Dept. Computer Science, Periyar E.V.R.College, Tiruchirappalli, India[2]

Abstract: Sentiment Analysis mainly refers to analyze feelings, emotions, or opinion of people expressed through social media, blogs and reviews. It extracts customer's reviews from the web and classifies the reviews using sentiment classification approach whether it is positive or negative. This paper proposes a new technique for sentiment classification to select most important features using different feature weights. Firstly, different data pre-processing techniques are applied on the labeled polarity movie reviews; Yelp restaurant and Amazon product reviews dataset. Secondly, Information Gain, Uncertainty and Gini Index methods are used to select most influential features. Finally, the *sentiment classification* task is done using Rapid Miner, an open source data mining tool. The performance of Support Vector Machine (SVM), is examined in combination with different feature selection schemes to obtain the results for Sentiment Analysis. The paper concludes with the investigation of experimental results show the effectiveness of the classifier with Information Gain.

Keywords: Gini Index, Information Gain, Sentiment Analysis, Support Vector Machine, Opinion, Uncertainty

I. INTRODUCTION

Sentiment Analysis is one of the most and recent research areas using Natural Language Processing (NLP) techniques. In recent days, the people are expressing their sentiments and opinions on products, movies, events, restaurants, individuals etc., in the form of blogs, tweets, face book messages, comments and reviews. People always prefer to hear others opinion before making decisions. There is a need to analyze the user's opinion, because it is very difficult to find out and it expresses whether it is positive or negative sentiments. For that situation Sentiment Analysis techniques are used to classify sentiments from text data in their appropriate class either positive or negative.

In this paper, the movie reviews, product and restaurant reviews are used for document and sentence level sentiment classification, because special challenges are associated with movie and product reviews. Movie review classification is different from other topic-based classification, because it based on domain specific and semantic words [9]. The proposed model mainly concerns with supervised learning techniques on a labelled movie reviews benchmark dataset created by Pang and Lee [10] and freely available on the Internet. Also, mobile phone reviews and restaurant reviews are used for sentence level sentiment classification created by Kotzias et al., [11].

Opinion in Sentiment Analysis classified at three types namely feature level, sentence level and document level [19]. The feature level classifications extract the important features from document and then classifies whether it is positive, negative or neutral opinions. Sentence level classification considers classification of reviews at individual sentence. Document level sentiment classification is used to classify the whole document contains as positive or negative reviews. Machine learning algorithms applied to classify and predict whether a document represents positive or negative sentiment. In general supervised classification algorithms has proved effectively and widely used in sentiment classification [12].

Specifically, the proposed model uses Support Vector Machine (SVM) classifier for classifying sentiments in sentence level and document level, finds the results and compares with the existing results. This paper is organized into five sections. In the first and second sections the introduction and previous related work is described. Section three describes the detailed methodology of the proposed model, and Section four discusses the experimental results of proposed model. Finally, Section five concludes the paper along with scope for future work.

II. RELATED WORK

Pang et al., reported 87% of accuracy rate of document level sentiment classification of the movie reviews using unigram feature and Support Vector Machine classifier [1]. Shotaro Matsumoto et al., proposed syntactic relations between words in sentences for document sentiment classification and used text mining techniques to extract frequent word sub-sequences and dependency sub-trees sentences in a document and use them as features of support vector machines [2]. They achieved 93.2% of accuracy using movie review dataset.

Isabella et al., used movie reviews for sentiment classification and evaluated a range of feature seletors to improve the performance of the classifers systematically [4]. Abinash Tripathy et al., applied NB and SVM Machine algorithms for classifying sentiments and obtained 89.5% of accuracy using NB classifier [5]. O'Keefe et al., proposed a new technique to select features using attribute weights and applied NB and SVM

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

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

classifiers [6]. The author obtained 87.15% of classification accuracy using only selected attributes. Qi et al., extracted most relevant feature subset using adjectives and classified the opinion in words as either positive or negative using Word Net package [8]. Siddhartha Ghosh et al., discussed the concept of polarity in sentiment analysis using movie reviews and obtained 70.50% of accuracy with Rapid Miner Tool [9]. Dimitrios Kotzias et al., proposed new approach to the problem of predicting labels for sentences given labels for reviews, using a convolutional neural network to infer sentence similarity [11]. Pang et al., achieved 82.90% of accuracy using SVM classification for unigrams [12].

Gautami Tripathi et al., investigated different feature selection methods to obtain the results for sentiment analysis using NB and Linear SVM classification algorithms and observed that Linear SVM has high accuracy using higher order n-grams [13]. Ahmad Kamal investigated feature-level summarization technique to visualize mined features, opinions and their polarity values using different supervised machine learning techniques for sentence level subjectivity and objectivity classification [14]. Saruladha et al., implemented Feature-Based Sparse Non- Negative Matrix Factorization method (FS-NMF) [15]. The author selected highest weighted features, create weighted term-sentence matrix and group the review sentences into feature relevant clusters and achieved higher accuracy.

P.Kalaivani et al., applied SVM, NB and KNN algorithm for sentiment classification using movie reviews [16]. They used 3-fold cross validation and obtained accuracy of 81.45% of by using SVM classifier. M. Rushdi et al., explored the Sentiment Analysis task and carried 3-fold and 10-fold cross validations in SVM for Pang Movie review corpus [17]. Mouthami et al., implemented a new algorithm called Sentiment Fuzzy Classification Algorithm to improve classification accuracy of Movie review dataset [18]. 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 clustering algorithm [20].

Li et al., proposed active learning approach that combines the active learning strategy and the label propagation algorithm to make the classification decision[21]. Recently many a number of research papers are published by presenting new ideas and innovative techniques to perform sentiment analysis [22][23][24][25].

In this study, the proposed model focuses to achieve better accuracy of sentiment classification of movie reviews and mobile phone reviews using Information Gain based feature weighting method.

III. PROPOSED MODEL

Design of Proposed Model

This section presents the design and methodology of sentiment classification using movie and Amazon reviews. In this study, binary sentiment classification technique is used to classify the user's reviews of documents into two classes either positive or negative. Fig. 1 shows the diagrammatic representation of proposed model.

The dataset is pre-processed and processed data is converted into Bag of words. 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 of 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.

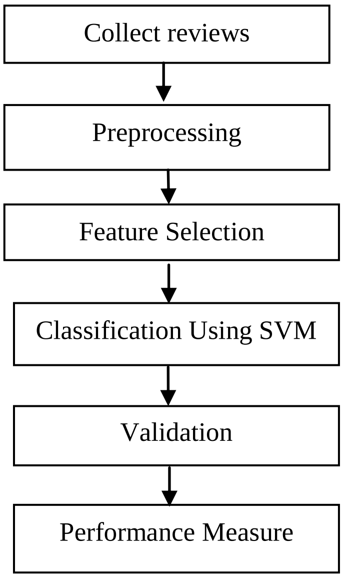

Fig. 1 Proposed Model

Feature Selection

Feature selection is the process of selecting relevant features. The proposed model uses three types of weighting scheme for feature selection. They are Information Gain, Uncertainty and Gini Index.

Information Gain

Information gain is usually a good measure for deciding the relevance of an attribute. It is used to decide which of the attributes are the most relevant. Also, calculates the weight of attributes with respect to the class label. It is one of the most powerful feature selection techniques and it is easy to compute and simple to interpret. Information Gain (IG) of a feature X and the class labels Y is calculated as

$$IG(X, Y) = H(X) – H(X/Y)$$

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

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

Uncertainty

The Uncertainty calculates the weight of attributes with respect to the label attribute by measuring the symmetrical uncertainty with respect to the class. The higher the weight of an attribute, the more relevant it is considered. Following is the equation for symmetric uncertainty.

$$\text{SU (X, Y)} = 2\left[\frac{IG(X\mid Y)}{H(X)+H(Y)}\right]$$

Where IG (X|Y) is the information gain of feature X, that is an independent attribute and Y is the class attribute. H(X) is the entropy of feature X and H(Y) is the entropy of feature Y. Information gain has a desired property, i.e. it is symmetric. The amount of information given by a feature Y about another feature X is effectively the same as that of the information given of feature X and the feature Y.

Gini Index

Gini index is supervised multivariate feature selection algorithm of the filter model to measure for quantifying a feature's ability to distinguish between classes. Given C classes, Gini Index of a feature *f* can be calculated as Gini Index can take the maximum value for a binary classification. It calculates the weight of attributes with respect to the label attribute by computing the Gini index of the class distribution. The higher the weight of an attribute, the more relevant it is considered. It can be calculated using following equation.

$$\text{Gini Index (f)} = 1 - \sum_{i=1}^{C}[p(i\mid f)]^2$$

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 origin. 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.

Validation

The proposed model uses split-validation. It has two sub processes: a training sub process and a testing sub process. The training sub process is used for learning or building a model. The trained model is then applied in the testing sub process. The performance of the model is also measured during the testing phase. The testing accuracy of the SVM classifier depends on the training object. Split validation is used to increase the performance of training data that increases the prediction of user 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 reviews that are correctly predicted whereas FP (False Positive) gives the value for number of positive 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 format is shown in below table.

TABLE I: CONFUSION MATRIX

Predicted Class		Actual Class	
		Positive	**Negative**
	Positive	TP	FP
	Negative	FN	TN

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 Studio software with its text processing extension, web mining and word net extension. Rapid Miner supports the design and documentation of overall data mining process and machine learning algorithms. This model is implemented using SVM classifier with different feature selection methods.

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

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

First, the data set is preprocessed and the bag of words are created using TF-IDF. Information Gain, Uncertainty and Gini Index based feature selection is used. SVM Classifier is applied on the reduced dataset and Split validation is applied and performance measures are evaluated.

Dataset Used

The proposed model uses two datasets. The first dataset contains reviews from Amazon and Yelp that can find at the UCI, the machine learning repository. This dataset was created for the paper "From Group to Individual Labels Using Deep Features," by Kotzias et al., for KDD 2015. The dataset contains 1000 labeled reviews equally divided into 500 positive and 500 negative sentences. Those were selected randomly for larger datasets of reviews and the goal was for no neutral sentences to be selected. The model uses 800 reviews (400 positive reviews and 400 negative reviews) for training the classifier and 200 reviews (100 positive and 100 negative) for testing the classifier.

The second dataset is 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 positive and negative (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 positive or negative. 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.

Data Preprocessing

The dataset consists of irrelevant and redundant information like punctuation marks, numbers, and special character. Several preprocessing steps are applied on the available dataset to optimize it for further experimentations. Tokenization is used to split the text into sequence of tokens using 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 3 characters and maximum length to 20 characters i.e. tokens with less than 2 characters and more than 15 characters were discarded. Stop words are removed. Stemming operator is used to stem English words using Porter stemming algorithm. 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 text to either lower case or upper case. In he proposed model all characters are converted into lower case letters. The results obtained for the various preprocessing stages are shown in table II.

TABLE II. VARIOUS PREPROCESSING LEVELS

Preprocessing	Restaurant Reviews	Mobile Phone Reviews	Movie Reviews
Initial Tokens	2018	2142	38911
Filtering Stop Words	1774	1785	38577
Filtering by Length	1752	1733	38234
Stemming	1455	1237	25004

The Term Frequency-Inverse Document Frequency (TF-IDF) scheme gives maximum accuracy for SVM. Therefore, the proposed model is implemented by using TF- IDF word vector creation method. The existing model is modified by applying Information Gain, Uncertainty and Gini Index weight feature selection method using SVM. In the proposed model different feature weights are applied to select features which are having highest values. For each method the confusion matrix is created and performance measures are obtained.

The performance measures are calculated using different feature weights using Restaurant reviews, Mobile Phone reviews and Movie reviews are shown in Table III, Table IV & Table V respectively. Also the graphical representations of Results are shown in Fig. 2, Fig. 3 and Fig. 4.

From Table III, the proposed model gives maximum accuracy of 83.50% using Gini Index feature selection and 83.00% of accuracy using Information Gain feature selection for Restaurant reviews at 0.2 threshold levels. From Table IV, the maximum accuracy 86.50% is obtained using Information Gain at 0.4 threshold level for Mobile phone reviews. From Table V, the movie reviews has maximum accuracy of 94% at 0.2 level using Information gain. By comparing those values, the proposed model gives better accuracy using Information Gain feature selection method.

TABLE III: RESULTS OF SVM CLASSIFIER WITH INFORMATION GAIN, UNCERTAINTY AND GINI INDEX BASED FEATURE SELECTION USING RESTAURANT REVIEWS

Thres hold	Number of Features	Information Gain			Uncertainty			Gini Index		
		Accuracy %	Precision %	Recall %	Accuracy %	Precision %	Recall %	Accuracy %	Precision %	Recall %
0.1	**146**	**83.00**	77.97	92.00	76.50	71.54	88.00	81.50	77.39	89.00
0.2	**291**	**83.00**	78.95	90.00	77.00	73.68	84.00	**83.50**	79.13	91.00
0.3	437	82.50	78.76	89.00	79.50	76.58	85.00	81.50	78.90	86.00

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

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

0.4	582	82.00	79.09	87.00	73.50	71.17	79.00	82.00	79.63	86.00
0.5	728	79.50	77.57	83.00	72.50	70.64	77.00	78.00	76.42	81.00
0.6	873	78.00	77.45	79.00	74.00	71.05	81.00	77.50	76.19	80.00
0.7	1018	77.50	75.23	82.00	75.50	72.97	81.00	77.00	75.00	81.00
0.8	1164	77.50	75.23	82.00	77.00	74.55	82.00	78.50	76.15	83.00
0.9	1310	77.00	75.00	81.00	78.00	75.00	84.00	76.50	74.13	81.00

TABLE IV: RESULTS OF SVM CLASSIFIER WITH INFORMATION GAIN, UNCERTAINTY AND GINI INDEX BASED FEATURE SELECTION USING MOBILE PHONE REVIEWS

Thres hold	Number of Features	Information Gain			Uncertainty			Gini Index		
		Accuracy %	Precision %	Recall %	Accuracy %	Precision %	Recall %	Accuracy %	Precision %	Recall %
0.1	124	75.50	68.89	93.00	72.00	66.92	87.00	76.50	69.92	93.00
0.2	247	84.00	79.31	92.00	74.50	68.99	89.00	82.50	77.31	92.00
0.3	371	83.50	80.18	89.00	81.00	76.27	90.00	83.00	79.46	89.00
0.4	**495**	**86.50**	85.44	88.00	80.50	77.98	85.00	85.00	83.65	87.00
0.5	619	83.50	84.54	82.00	81.50	80.00	84.00	83.50	83.17	84.00
0.6	742	84.50	84.85	84.00	81.00	80.39	82.00	82.00	81.37	83.00
0.7	866	80.00	80.61	79.00	81.00	79.81	83.00	80.50	80.61	83.00
0.8	990	80.00	80.00	80.00	80.50	79.05	83.00	79.50	79.21	80.00
0.9	1113	79.00	77.80	77.88	81.50	80.00	84.00	78.50	77.14	80.00

TABLE V: RESULTS OF SVM CLASSIFIER WITH INFORMATION GAIN, UNCERTAINTY AND GINI INDEX BASED FEATURE SELECTION USING MOVIE REVIEWS

Thres hold	Number of Features	Information Gain			Uncertainty			Gini Index		
		Accuracy %	Precision %	Recall %	Accuracy %	Precision %	Recall %	Accuracy %	Precision %	Recall %
0.1	2500	92.50	97.22	87.50	85.50	87.77	82.50	90.50	96.02	84.50
0.2	**5001**	**94.00**	96.81	91.00	84.00	87.36	79.50	91.75	96.13	87.00
0.3	7501	92.00	95.65	88.00	84.50	86.70	81.50	91.25	94.12	88.00
0.4	10002	93.00	95.74	90.00	82.25	83.42	80.50	92.25	94.71	89.50
0.5	12502	89.75	90.36	89.00	84.25	85.13	83.00	88.75	89.74	87.50
0.6	15002	88.75	89.34	88.00	77.75	80.33	73.50	86.75	87.31	86.00
0.7	17503	86.50	86.50	86.50	79.50	79.80	79.00	84.25	84.08	84.50
0.8	20003	84.25	83.09	86.00	79.00	77.62	81.50	82.50	81.86	83.50
0.9	22504	80.50	80.20	81.00	79.50	78.64	81.00	80.50	80.20	81.00

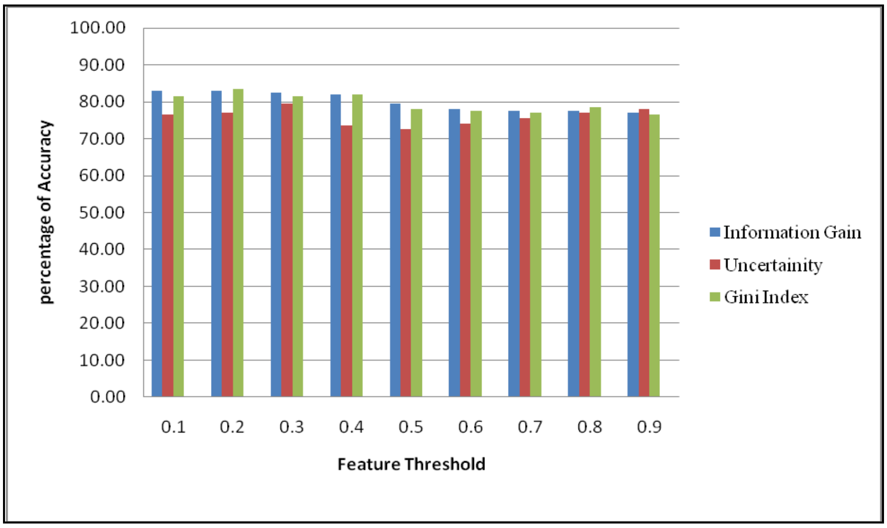

Fig. 2 Accuracy of various Feature Weighting Methods using Restaurant Reviews

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

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

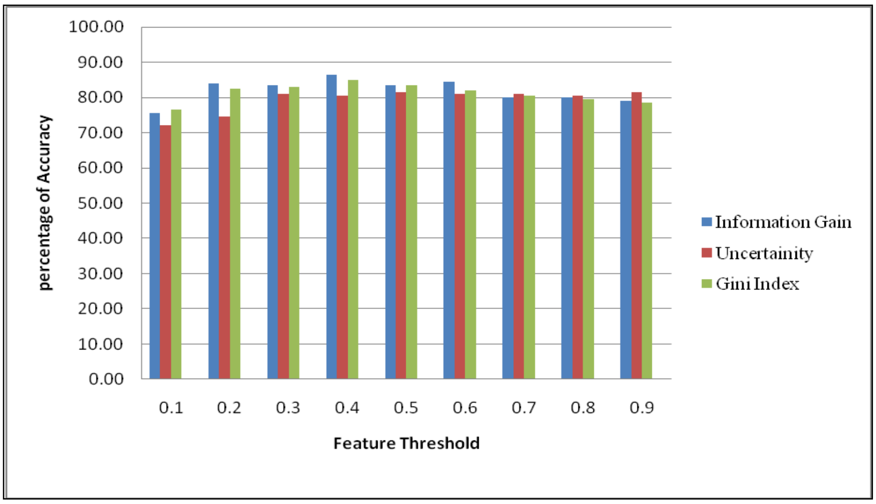

Fig. 3 Accuracy of various Feature Weighting Methods using Mobile Phone Reviews

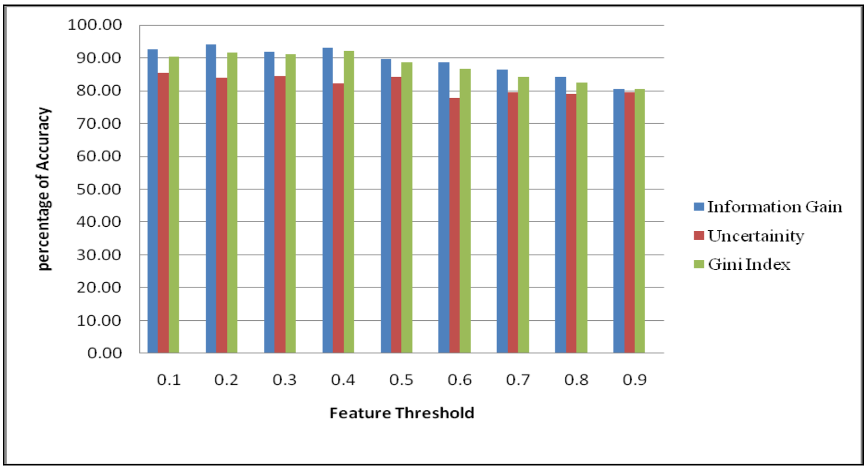

Fig. 4 Accuracy of various Feature Weighting Methods using Movie Reviews

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 Shotaro Matsumoto et al., [2], Gautami Tripathi et al.,[13] and O'Keefe et al., [6] models.

All three models used the same labelled movie review polarity dataset with 1000 positive and 1000 negative reviews. The following Tables VI & VII shows the comparison of obtained result with other methods and graphical comparison is shown in Fig. 5 and Fig. 6 using Movie Reviews and Restaurant Reviews.

TABLE VI: COMPARISON OF PROPOSED WORK WITH EXISTING LITERATURES USING MOVIE REVIEWS

Various Models	Accuracy %
Shotaro Matsumoto's Model	93.20
Gautami's Model	84.75
O'Keefe's Model	87.15
Proposed Model	**94.00**

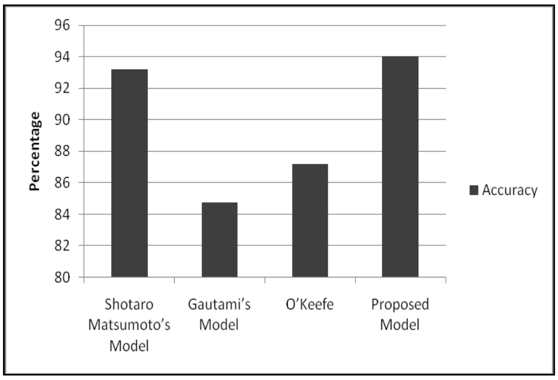

Fig. 5 Comparison between different models for Movie reviews

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

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

From the Table VI and Fig. 5, it is observed that the maximum accuracy is obtained by proposed model. Shotaro Matsumoto et al., [2] achieved 93.20% of accuracy using cross validation. Gautami Tripathi et al.,[13] obtained an accuracy of 84.75% for SVM using 5 fold cross validation for classification and O'Keefe et al., [6] got 87.15% of accuracy using SVM.

TABLE VII: COMPARISON OF PROPOSED WORK WITH EXISTING LITERATURE USING RESTAURANT REVIEWS

Various Models	Accuracy %
Kotzias's Model	78.16
Proposed Model	**83.00**

In this proposed model Split Validation gives maximum accuracy of 94.00% using Information Gain feature selection method for SVM classification using Movie reviews.

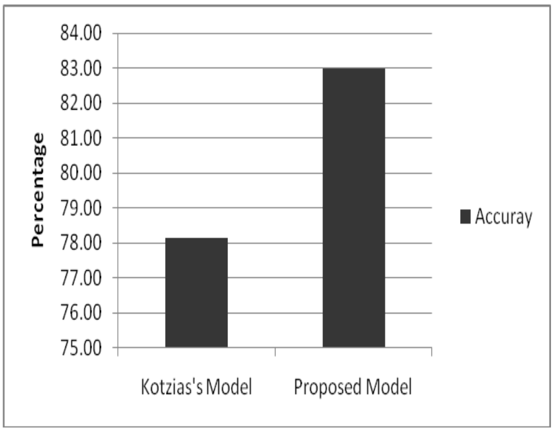

Fig. 8 Comparison between different models for Restaurant Reviews

From Table VII and Fig. 6, the proposed model gives higher accuracy than Kotzias's Model using Restaurant reviews.

V. CONCLUSION

The proposed model has been made to analyze sentiments for Movie, Mobile Phone and Restaurant reviews using SVM classification because it is observed that SVM classifier outperforms every other classifier in predicting the sentiment of a review. It presents an approach for Sentiment Analysis with various feature selection methods by using different datasets. Experimental results show that Information Gain feature weight selector achieves the best feature subset for classification and gives better accuracy of 94.00% for sentiment movie review data set and 86.50% for Mobile Phone reviews using SVM classifier.

In this paper, the proposed model is implemented for multiple domains using only unigrams. In future, this model can be extended by applying different classification algorithm by combining with different feature selectors for dimensionality reduction and classification.

REFERENCES

[1] Bo Pang, Lillian Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts", Proc. of 42nd ACL, pp. 271–278, 2004
[2] Shotaro Matsumoto, Hiroya Takamura, and Manabu Okumura , "Sentiment Classification Using Word Sub-sequences and Dependency Sub-trees", Proc. of 9th Pacific-Asia Conference, PAKDD 2005, Hanoi, Vietnam, May 18-20, pp. 301-311, 2005.
[3] Tony Mullen and Nigel Collier. Sentiment Analysis using Support Vector Machines with Diverse Information Sources. Proc. of 9th EMNLP, pp.412–418, 2004.
[4] 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), Vol. 3, Dec 2012-Jan 2013.
[5] 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.
[6] 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.
[7] 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.
[8] 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.
[9] 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 , Volume: 3 Issue: 8 , pp. 5167 – 5172, ISSN: 2321-8169, August 2015.
[10] Polarity dataset version 2.0, Sentiment Analysis Dataset. [Online]. Available: http://www.cs.cornell.edu/people/pabo/movie -review-data/review polarity.tar.gz
[11] Dimitrios Kotzias, Misha Denil, Nando De Freitas, Padhraic Smyth, "From Group to Individual Labels using Deep Features", KDD 2015.
[12] 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.
[13] Gautami Tripathi and Naganna S, "Feature Selection and Classification Approach for Sentiment Analysis", Machine Learning and Applications: An International Journal, Vol.2, No.2. June 2015.
[14] Ahmad Kamal, "Review Mining for Feature Based Opinion Summarization and Visualization", International Journal of Computer Applications 119(17):6-13, June 2015.
[15] Saruladha. K, Banupriya. D, Nargis Banu. J, "Opinion Summary Generation for Product Reviews", International Journal of Engineering Research & Technology, (ISSN: 2278-0181), Vol. 3, Issue 3, March 2014.
[16] 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, Vol. 4 No. 4, Aug-Sep 2013.
[17] M. Rushdi Saleh, "Experiments with SVM to Classify Opinions in Different Domains, Expert Systems with Applications", 2011.
[18] 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, pp. 271–276, 2013.
[19] Bing Liu, "Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers", 2012.

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

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

[20] Benito Alvares, Nishant Thakur, Siddhi Patil, "Sentiment Analysis using Opinion Mining", International Journal of Engineering Research & Technology , (ISSN: 2278-0181), Vol. 5, Issue 04, April 2016.
[21] 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, pp. 139-148, 2012.
[22] 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, New York, NY, USA, Article 5, pp. 1-8, 2012.
[23] Aditya Joshi, Balamurali A R and Rajath Mohanthy, "C-feel-it: A Sentiment Analyzer for Micro Blogs", Proceedings of ACL: Systems Demonstrations, pp. 127–132, 2011.
[24] ZhongwuZhai, Bing Liu, Hua Xu and Peifa Jia, "Clustering Product Features for Opinion Mining, WSDM"11, February 9–12, Hong Kong, China, 2011.
[25] ZhongwuZhai, Bing Liu, Lei Zhang, Hua Xu and Peifa Jia, "Identifying Evaluative Sentence in Online Discussions", Association for the Advancement of Artificial Intelligence, 2011.

BIOGRAPHY

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.

Dr.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.