INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

Machine Learning Approach to Analyse Ensemble Models and Neural Network Model for E-Commerce Application

P Kalaivani[1]*, S Selvi[2]

1 Faculty of Information Technology, St.Joseph's College of Engineering, Chennai, India. Tel.: 9884452304
2 Faculty of Computer Science and Engineering, R.M.K Engineering College, India. Tel.: 9600142824

Abstract

Objectives: The main objective of this study is to compare the performance evaluation of ensemble based methods and neural network learning on various combinations of unigram, bigram, and trigram feature vector along with feature selection (IG) and feature reduction (PCA) for sentiment classification of movie reviews. **Methods:** Bagging and Adaboost are the techniques used in ensemble learning to learn the sentiment classifier to get better classification accuracy, using SVM, NB as a core learner for different models of attribute vectors. The classification results of the ensemble approach are compared with neural network learning for classification of movie reviews. Among the ensemble methods, AdaBoost with base learner SVM outperforms in classifying attribute vectors for model m-iii. The backpropagation algorithm is used to improve classification accuracy in the neural network learning and IG and PCA are used in sentiment classification to reduce the feature length and training time. **Findings:**The classification results of ensemble based approach are compared with neural network learning. Between the two ensemble based methods, Adaboost + SVM outperform in classifying the sentiment of movie reviews for m-iii feature vector. IG and PCA are used in sentiment classification in order to reduce the feature length. Between the IG and PCA methods, IG performs better than PCA. Among IG+Adaboost+SVM and neural network learning methods, IG+Adaboost+SVM performs better than neural network learning. **Improvement:** In our application, we are using the ensemble based methods and neural network learning, these methods are compared and analyzed the performance for various levels of feature vectors. A classification algorithm may be designed to analyze the performance with other neural network methods.

Keywords: Machine learning; sentiment classification; bagging; AdaBoost; ensemble learning; back propagation neural network; feature selection; movie review

1 Introduction

Sentiment analysis is a continuous study of knowledge discovery . This is a process for determining people's thoughts, opinions and feelings about a given topic or item. "Sentiment analysis" is highly domain-dependent [1]. The performance will differ considerably from one field to other that makes a very exciting and difficult task. Machine learning methods have been analyzed for classification performance. Ensemble learning and neural network learning have been applied in various relevance domains for the Sentiment classification.

Sentiment analysis aims to categorize product reviews or opinions into good or bad to measure the comprehensive customer sentiment behind their brand. But, the problem is that most sentiment analysis work uses simple terms to express sentiment about product or service. "There are three classification levels in sentiment analysis, document level, sentence level and aspect level sentiment analysis. Document level sentiment analysis aims to classify an opinion document as expressing a positive or negative opinion or sentiment. It considers the whole document a basic information unit. Sentence level sentiment analysis aims to classify emotion expressed in each sentence. Aspect level sentiment analysis aims to classify the sentiment with respect to an aspect of entities." [2].

With the development e-commerce and social networking websites, the social media has huge amounts of data. Opinions shared on social media websites play a prominent part in the analysis of the business industry. As the amount of posts has been expanding at a quick rate, analyzing comments written on social media becomes complex and difficult for the customer [3].

In Ref. [4], Xia et al. suggested an attribute ensemble plus specimen selection, to learn a new classification function. PCA based ensemble selections were used in FE for adaptation. The experimental analysis was carried out in sentiment classification on multi domain dataset. The result showed that the SS-FE superior to individual FE and PCA-SS.

In Ref. [5], Namsrai et al. the authors uses the attribute selection method to select a set of original attributes from the arrhythmia dataset. Attribute or feature selection was used in sentiment classification to decrease the dimensionality of attribute vector and reduce the training data size. The voting methods were used to measure the score for each classification in the ensemble, both for the classification error and for the feature selection rate. Experimental analysis was performed for arrhythmia dataset to classify the model based on voting approach. SVM, NB, Bayesian, decision trees are the base classifier used in voting approach. Among SVM, NB, Bayesian and feature based ensemble method classifiers used in voting approach, it has shown that feature selection based ensemble method performs better than other methods.

In Ref. [6], Gang Wang et al. proposed the ensemble method as a machine learning technique in which multiple classifiers are trained to enhance the classification accuracy for the algorithm. To increase the classification accuracy authors examined "bagging" and "boosting" strategies in learning approaches to ensemble learning. The information gain method was employed to recognize and eliminate the duplicate dataset from the original dataset. The methods DT and NB were used as base learners for bagging ensemble learning technique. IGF-Bagging performed better than bagging and boosting which uses a decision tree and NB as a base learner.

In Ref. [7], Xia et al. presented a sentiment analysis process to identify the function of commonly evaluating as either good or bad review. The p art-of-speech (POS) based attribute selection and the feature relation based attribute selection was used in sentiment classification. An ensemble algorithm was implemented to incorporate various feature collection and classification algorithms. For each of the feature collection, authors investigated naive bayes (NB), maximum entropy (ME) and support vector machines (SVM) as a base learner. The evaluation was conducted on the various product dataset and film review dataset for the part-of-speech (POS) based ensemble. In part-of-speech based ensemble and word relation based ensemble, comparative study had taken based on three strategies. Strategy 1 is the results of individual classifiers, Strategy 2 is an ensemble along with classification algorithm and Strategy 3 is the grouping of attribute selection and supervised learning classification algorithm. Strategy3 outperformed in meta-classifier combination in sentiment classification.

In Ref. [8], Duncan, Brett, and Yanqing Zhang, discussed sentiment classification on twitter dataset. Bloggers rapidly spread negative comments or product assessments on digital worlds on commercial blogs. These negative remarks have often been hazardous to companies and could cause major damage. The neural network based measure was proposed to assist companies in quickly and effectively finding bad comments. Performance analysis had been performed for "feed forward pattern network". The result recommended that the "principle component analysis (PCA)" was not effective feature reduction technique.

In Ref. [9], Tan, Songbo, and Jin Zhang proposed study for Chinese dataset using MI, CHI, IG and DF for attribute selection. Authors evaluated KNN, SVM, centroid classifier, winnow classifier and NB classifier. The classification results for Chinese review dataset were compared. The evaluation shows that IG outperforms the best for sentimental term selection than other feature selection methods. Among the various machine learning approaches, SVM performed best in experimental methods.

Consequently, we need a system of an opinion mining which allows for the collection of comments. Companies can use such opinion mining system to determine how the person perceives their product, how the match up to competition. It is perfectly natural to often rely on the comments and the knowledge of others during the purchasing of items [10]. Individuals

and businesses are generally always interested in the opinions of others such as if anyone wishes to buy a new product, what others think about the product and what they think about it [11].

The key objective of the study would be to evaluate the performance with ensemble methods and neural network learning on various combinations of unigram, bigram and trigram feature vectors along with attribute choice and attribute reduction for opinion classification of movie reviews.

This paper consists of five sections. The introduction and literature surveys are is discussed in section 1. The methodology is covered in section 2. Section 3 discusses performance validation. Section 4 contains the study outcome and discussions. Section 5 deals with the conclusion of this study.

2 Methodology

Ensemble learning methods such as "bagging" and "AdaBoost" are used to enhance the accuracy in classifying the review dataset. The movie review dataset is collected and the opinions are categorized into good or bad class labels. Opinions are preprocessed and the attributes are extracted. The extracted attributes are grouped based on unigram, bigram, and trigram. After preprocessing, the IG and PCA are used to reduce the dimensionality of the movie review dataset. An analysis is done to compare the outcome of the ensemble approaches and the neural network method and provide the best combination result. The methodology for testing the different classification models has been described below.

1. Data preprocessing is performed to remove the irrelevant data.
2. Attribute vector space for the model, m-i is developed with unigram, the model, m-ii with unigram and bigram, and the model, m-iii with unigram, bigram, and trigram are developed.
3. IG on the models m-i, m-ii, m-iii to produce selected feature sets.
4. PCA on the models m-i, m-ii,m-iii to produce reduced feature sets.
5. The following classification models are developed with the models m-i,m-ii,m-iii

   (a) support vector machine (SVM) model
   (b) Naive Bayes model
   (c) Bagging with support vector machine(SVM) model
   (d) Bagging with Naive Bayes model
   (e) AdaBoost with support vector machine(SVM) model
   (f) AdaBoost with naive Bayes model
   (g) Neural Network (Backpropagation) model

6. The class label of the respective user opinion in the testing dataset is classified.
7. The performance measure accuracy and other measures are computed from the final table values.
8. Compare all the model's performance values.

2.1 Data preprocessing

In preprocessing steps, movie review datasets are preprocessed. The preprocessing includes tokenization, conversion of the upper case to lowercase, stop-word, stemming, and TF- IDF value is calculated for each feature in the review dataset. Tokenization is a process used to split the sentence into multiple words. Case transformation operation is used to convert the review upper case into a lower case. Stop words are words that no meaning to be used in a review dataset. The stemming process is used to remove suffix words from the given word in the dataset. The suffix characters are removed to reduce the word length in the dataset. The TF- IDF is a statistical measure which represents how important a feature for reviews. The number of times a feature occurs in a review is called "term frequency". The inverse document frequency is a statistical weight used for measuring the importance of a feature in a review data and it can be calculated as follows

$$TF = 1 + log(1 + log(frequency(w, D_w))) \tag{1}$$

$$IDF = log\left(\frac{1+D}{D_w}\right) \tag{2}$$

Where D is the number of opinion reviews, and D_W is the number of terms w present in reviews. TF -IDF can be calculated as follows

$$TF - IDF = TF * IDF \tag{3}$$

2.2 Attribute Selection and Attribute Reduction

In the general attribute selection and attribute reduction methods are used to reduce the dimension of feature vectors[12]. The selection of attributes is the procedure of selecting a sample of important attributes to be used in model development. Attribute reduction is finding a subspace that has less dimension that of original feature space. The popular feature selection and reduction methods, information gain (IG) and principle component analysis (PCA) are used as the dimension reduction of the feature.

2.2.1 Information Gain(IG)

IG is a method of selecting attributes. The preprocessed data of the review dataset are given as input for feature selection. "It is used to reduce bias towards multi-valued attributes by taking the number and size of branches into account when choosing an attribute. It is applied to attributes that can take on a large number of distinct values and might learn the training set too well. It is often used to select which of the attributes are the most frequently used in the dataset.[13]

2.2.2 Principal Component Analysis(PCA)

The PCA is commonly used as a statistical technique. In PCA technique, the original data space is mapped into the standardized transformation matrices. The covariance matrix for each word in a given review dataset is calculated and the correlation matrix is also calculated from the covariance matrices. The correlation matrix shows how the words in the dataset are strongly correlated. The value ranges from -1 to 1 in the correlation matrix. It is represented in a symmetric matrix. If the correlation value is -1, then the attribute is negative. If the correlation value is +1, then the attribute is positive. If the correlation value is 0, then the word is equally correlated[2]. If the variance covered is 0.1 then more words can be reduced and gives less accuracy. If variance covered is 0.95 then less words can be reduced and gives less accuracy. If variance is between 0.4 and 0.5, then average words are reduced and give more accuracy. The feature can be reduced based on variance used and sample features along with PCA weight.

2.3 Bagging

The bagging algorithm generates a series of models for learning a system in which every model provides a prediction with equal weight. The classifier is trained on a selection of review words taken from the training set and predicts the class label for each word in each classifier. Then each classifier returns its majority vote and predicts the class label to word for unknown sample X. Where M is the model to construct the classifier and P is the predicted output value for each classifier[14,15]. The bagging algorithm works as follows. Given Training reviews dataset R of $X=\{X_1, X_2... X_n\}$ and class label Y= {positive, negative} training words. For each iteration k (i.e. k=10), training words are sampled with replacement with the original set of review data. For each training dataset, the classifier learns and returns its majority votes classify unknown test sample attribute and gives the class label with most labels.

2.4 Adaptive boosting

AdaBoost algorithm work as follows. Given a training review dataset R of $X=\{X_1, X_2... X_n\}$ and Y= {positive, negative} training words. The AdaBoost assigns each training data to an equal weight of 1/Y, for each iteration k (i.e. k=10), the training words from movie reviews dataset R are sampled and derive a model M using a base classifier and calculate the error for each classifier M[16,17]. The inaccuracy value of a model M and the misclassification value of attribute X_j is calculated. If the review word is misclassified then error (M_i) is 1 otherwise it is 0. The error (M_i) is above 0.5 then reinitializes the weight and go to another distribution because the performance is poor when error rate exceeds 0.5. At each iteration, weights for the correctly classified training words are updated. The weights for each training data are standardized as a result of their addition of the weight is the same as the addition of the old weight. Boosting assigns a weight to each classifier's vote, based on how well the classifiers performed. The classifier error rate is lower, accuracy gets more. After calculating the weight for each classifier, sum the weight of each classifier that assigns class c to X. The class with the high score weight is returned as the class prediction.

2.5 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a best classification method for linear as well as non-linear sentiment dataset. The "support vector machine" is used to evaluate opinion data used for classification[18]. Let R be the movie review dataset with class labels positive and negative. Movie review dataset is viewed by a non-dimensional attribute vector $X=\{X_1, X_2.. X_n\}$ and Y= {positive, negative}, positive represents a positive attribute label and negative represents a negative attribute label. Based on a maximal

marginal hyperplane is calculated as follows

$$d\left(X'\right) = \sum_{i=1}^{l} Y_i a_i X_i X' + b_i \tag{4}$$

X_i is the set of training words in movie reviews, Y_i is the good or bad attribute label, X_i is the testing words, a_i, b_i are the parameter to determine automatically by an optimized support vector. If d(X') is positive, then X' needs to fall on or above the maximal marginal hyper plane and so the SVM predicts that X' as a positive (good) attribute label otherwise X' is a negative (bad) attribute label.

2.6 Naive Bayes (NB)

The navie bayes (NB)is one of supervised learning classification algorithms. Let R be the movie review dataset with class labels positive and negative. Each review dataset R is viewed by "n-dimensional" attribute vector, X={X_1, X_2... X_n}. The probability of each class label negative and positive, P ($C_{positive}$) and P ($C_{negative}$) can be computed based on the training words and conditional probabilities P (X|$C_{positive}$) and P (X|$C_{negative}$) and each attribute vector needs to be maximized P (X|C_i) P (C_i) [19].

2.7 Neural Network Learning

Backpropagation algorithm is a technique in neural network learning. It learns by iteratively processing reviews dataset of training words, comparing the network prediction for each word with a class label (positive/ negative). The words in the review dataset are entered and passed through the input neurons and are sent to a second layer known as hidden neurons. The input and output for the units in the hidden layer can be calculated from a weighted sum of input neurons along with weight at input level. The weighted output of the hidden unit is input to the output nodes, which provide network prediction of the given word in the dataset [20,21]. The user needs to determine the topology of the network, specifying a bias value, and specifying the learning value for the algorithm between 0.1 and 0.9.

3 Performance Validation

A confusion matrix includes information on classification actual and predicted classifications. Table 1 displays confusion matrix for a classifier of two classes. The matrix column shows the instance in the predicted class whereas the instance in an actual class is represented each row.

Table 1. Confusion Matrix

Actual		Predicted	
		Negative	Positive
	Negative	True Negative(TN)	False Positive (FP)
	Positive	False Negative (FN)	True positive (TN)

"Accuracy" is defined as proportions of total number of predictions were correct. "Precision" is defined as the ratio of estimated positive words, recall as the correctly identified ratio of positive words. "F- Measure" is defined as proportion of precision and recall of correctly identified instances. Performance measures can be calculated from the confusion matrix.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{8}$$

4 Experimental Result

In this Section, comparison studies are performed with two feature selection techniques (IG, and PCA), two classifiers (SVM, and NB), two ensemble methods, and Back propagation. These methods are discussed in the methodology section. The features are extracted from the movie review dataset. The attribute vector model is developed with a TF-IDF measure. The attributes are grouped as "unigram", "bigram", and "trigram". The impact of an attribute size is identified in three models and its features shown in Table 2. The model, m-i is described as attribute vector with unigram, m-ii is viewed as an attribute vector with bigram, and m-iii is viewed as an attribute vector with trigram.

The weka tool is used to identify the PCA of models m-i, m-ii, and m-iii. The stopping rule used is an eigenvalue, which is greater than value 1. The number of "principal components" is reduced to 165, 150, and 147 for the models, m-i, m-ii, and m-iii. Cumulative variance of 30% gained for the models, m-i, m-ii, and m-iii feature vector. The variance percentage is also reduced based on the stopping rule selected. The "information gain" value for the models, m-i, m-ii, and m-iii is identified by setting the threshold value of 0.002. Table 3 , displays an attribute reduced set of IG and PCA.

The SVM and NB are the base learners used in "ensemble learning" technique to classify training and testing dataset. The support vector machine that uses the kernel function, the chosen kernel type normalized poly kernel with a default value of kernel parameter. A backpropagation algorithm is used in neural network learning. The parameter for momentum is set to 0.2, learning value is assigned to 0.3 along with training time is assigned to 50.

4.1 Comparison of Classifiers

Based on the workflow, the performance of different classification algorithms are analyzed on various combinations of unigram, bigram, and trigram attribute vectors. In this analysis, accuracy is used to evaluate the various approach to the movie review dataset. All studies have been verified through tenfold cross-validation. Table 4 shows the experimental results when using the classifiers together with IG and PCA. The result in Table 4, shows that the combination of IG, SVM, and m-iii achieved 84.40% best accuracy.

Table 2. Properties of Input data

Properties	Model		
	m-i	m-ii	m-iii
Number of reviews	2000	2000	2000
Positive Review	1000	1000	1000
Negative Review	1000	1000	1000
Number of feature	1160	1138	1144
Type / Model	Unigram	Bigram	Trigram

Table 3. ThreeModel Description

Properties	Model		
	m-i	m-ii	m-iii
Number of feature	1160	1138	1144
Variance covered	0.37	0.37	0.37
Threshold value	0.002	0.002	0.002
Number of principal components	165	150	147
Number of feature in IG	168	153	150

The classification results of NB show that accuracy result is comparatively lesser than other classifier. NB is not an efficient algorithm on bigram. The reason for higher error rate performance is that all features are independent.The classification results obtained for feature selection IG, ensemble bagging and classifiers, SVM and NB combinations and feature reduction PCA, ensemble bagging, and classifiers, SVM and NB combination results are tabulated in Table 5. The results in Table 5, shows that the model, m-i, feature selection IG, ensemble bagging, and classifier, SVM combination outperformed than other models. The combination of PCA, bagging, and the model, m-ii gives less accuracy of 78.90%.The classification results of NB show that accuracy result is comparatively lesser than all other hybrid model.NB is not an efficient algorithm on bigram. The reason for less accuracy performance is that all features are independent.

Table 4. Comparison Result of Base Classifier

Model/ Performance	IG						PCA					
	SVM			NB			SVM			NB		
	m-i	m-ii	m-iii	m-i	m-ii	m-iii	m-i	m-ii	m-iii	m-i	m-ii	m-iii
Accuracy	84.10	83.65	84.40(↑)	79.60	81.50	82.00	84.05	83.90	83.70	81.50	79.50	80.05
Precision	83.20	84.10	84.70	81.10	80.50	80.80	83.80	83.80	83.50	82.50	80.80	80.05
Recall	84.80	83.30	84.10	78.70	82.05	82.70	84.30	83.90	83.80	80.80	79.10	79.70
F-Measure	83.24	82.50	84.30	79.80	81.26	81.73	84.00	83.80	83.60	81.64	79.90	79.80

Table 5. Comparison Result of Bagging

Model/ Performance	IG						PCA					
	Bagging + SVM			Bagging + NB			Bagging + SVM			Bagging + NB		
	m-i	m-ii	m-iii	m-i	m-ii	m-iii	m-i	m-ii	m-iii	m-i	m-ii	m-iii
Accuracy	84.35(↑)	83.15	83.90	80.75	81.60	82.25	82.45	81.85	83.70	80.95	78.90	79.90
Precision	84.20	83.10	84.10	82.10	80.80	81.40	81.10	81.30	83.50	81.70	79.50	80.30
Recall	84.50	83.40	83.76	79.94	82.10	82.80	83.30	82.90	83.80	80.40	78.50	79.60
F-Measure	84.30	83.24	83.92	81.00	81.40	82.09	82.18	82.20	83.60	81.04	78.90	79.90

Table 6. Comparison Result of AdaBoost

Model/ Performance	IG						PCA					
	AdaBoost + SVM			AdaBoost +NB			AdaBoost + SVM			Adaboost + NB		
	m-i	m-ii	m-iii	m-i	m-ii	m-iii	m-i	m-ii	m-iii	m-i	m-ii	m-iii
Accuracy	84.15	83.95	84.45(↑)	81.50	83.40	83.65	84.10	84.15	83.78	81.50	80.35	79.45
Precision	83.20	82.50	84.80	81.90	85.00	83.10	83.80	83.90	83.50	82.50	80.00	78.60
Recall	84.80	83.90	84.15	81.25	82.36	84.02	84.30	84.80	83.80	80.80	80.50	79.90
F-Measure	83.99	83.19	84.32	81.57	83.65	83.55	84.00	84.10	83.60	81.64	80.24	78.25

Table 7. Comparison Result of IG +AdaBoost + SVM and IG +Neural Network

Model/ Performance	IG + AdaBoost + SVM			IG + NN		
	m-i	m-ii	m-iii	m-i	m-ii	m-iii
Accuracy	84.15	83.95	84.45(↑)	80.95	80.35	82.15
Precision	83.20	82.50	84.70	83.40	84.30	79.70
Recall	84.80	83.90	84.10	79.50	78.12	83.80
F-Measure	83.99	83.19	84.30	81.40	81.00	81.69

The classification results attained for feature selection IG, ensemble AdaBoost and classifiers, SVM, and NB combinations and also feature reduction PCA, ensemble AdaBoost and classifiers, SVM and NB combinations are tabled in Table 6. The results presented in Table 6 show that model, m-ii, feature selection IG, ensemble AdaBoost, classifier, SVM combination outperformed other models. The combination of PCA, AdaBoost ensemble, and the model, m-iii gives less accuracy of 79.45%. The classification results of NB show that accuracy result is comparatively lesser than all other hybrid model. The NB is not an efficient algorithm on trigram.

, shows the classification result of IG along with AdaBoost and SVM, and the combination of IG, Neural Network. It can be observed from Table 7 , combination of AdaBoost, classifier SVM performs better than other model. The performance of feature selection IG, ensemble AdaBoost, and classifier SVM outperformed than IG, Neural Network. Compared with two methods, model, m-iii, AdaBoost, and SVM combination predicts high accuracy of 84.45%. the results show that the combination of IG, NN, model, m-ii provide a less accuracy value of 80.35%.

Among all models, model, m-iii of AdaBoost, SVM predicts reviews with high accuracy of 84.45%. This shows that this model predicts positive reviews that were correctly classified as positive reviews and negative reviews were correctly classified as negative reviews for trigram feature. Among all models, the combination of PCA, Bagging, model, m-iii gives less accuracy of 78.90%. This shows that this model predicts negative reviews that were incorrectly classified as positive reviews and vice versa

for trigram feature. Figure 1 , depicts the ROC curve for different classifiers with IG and PCA as feature selection and feature reduction. Figure 2 , depicts the ROC curve for Neural Network and AdaBoost. In Figure 1, and 2, X-axis is a false positive rate and Y-axis is a true positive rate. The combination of AdaBoost , SVM classifier is known to perform better than other methods.
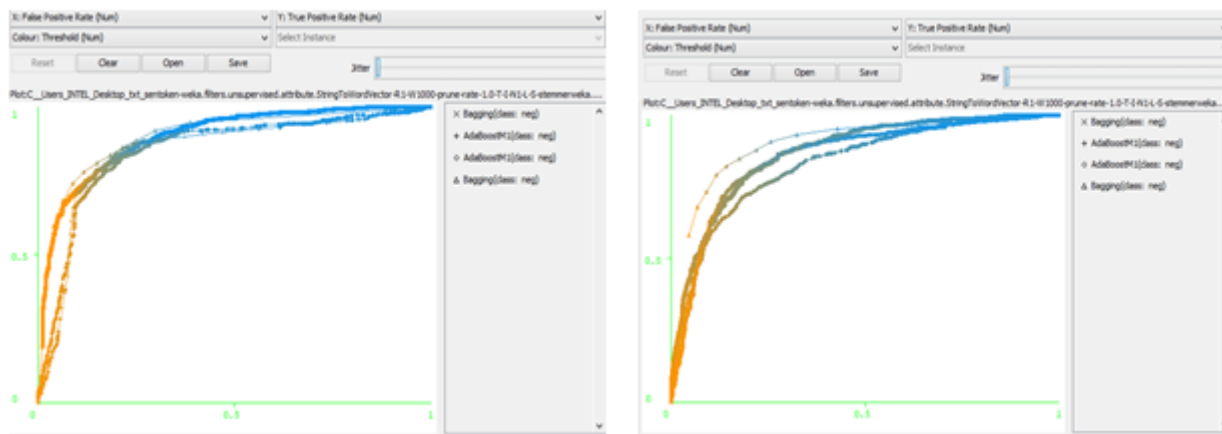

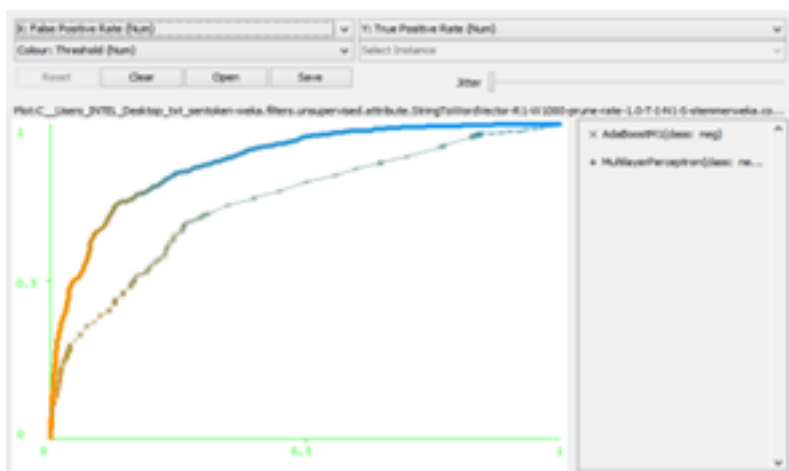**Fig 1.** ROC Curveof Feature Selection and reduction IG and PCA

Fig 2. ROC Curvefor Neural Network, and AdaBoost +SVM

5 Summary & Conclusion

In this study, movie reviews are collected and preprocessed. The features are extracted from the movie review dataset and the attribute vector model is developed with a TF-IDF measure. The attributes are grouped as "unigram", "bigram", and "trigram" and the impact of an attribute size is identified in three models. We empirically compared two feature selection techniques (IG, and PCA), two classifiers (SVM, and NB), two ensemble methods, and Back propagation. Between the IG and PCA methods, IG performs better than PCA. Between the two ensembles methods and Back propagation, Adaboost + SVM outperform in classifying the sentiment of movie reviews for m-iii feature vector.

References

1) Mtetwa N, Awukam AO, Yousefi M. Feature Extraction and Classification of Movie Reviews. In: and others, editor. 5th International Conference on Soft Computing and Machine Intelligence, ISCMI. 2018;p. 67–71. Available from: https://doi.org/10.1109/ISCMI.2018.8703235.

2) Gül S, Kabak Ö, Topcu I. A Multiple Criteria Credit Rating Approach Utilizing Social Media Data. *Data and Knowledge Engineering*. 2018. Available from: https://doi.org/10.1016/j.datak.2018.05.005.

3) Elamin N. Sentiment Analysis with Supervised Learning Techniques: A Survey. *Indian Journal of Science and Technology*. 2020;13(3):249–268. Available from: https://dx.doi.org/10.17485/ijst/2020/v13i03/148900.

4) Xia R, Zong C, Hu X, Cambria E. Feature ensemble plus sample selection: Domain adaptation for sentiment classification. *IJCAI Int Jt Conf Artif Intell*. 2015;p. 4229–4262.

5) Namsrai E, Munkhdalai T, Li M, Shin JH, Namsrai OE, Ryu KH. A Feature Selection-based Ensemble Method for Arrhythmia Classification. *Journal of Information Processing Systems*. 2013;9(1):31–40. Available from: https://dx.doi.org/10.3745/jips.2013.9.1.031.

6) Wang G, Ma J, Yang S. Igf-bagging: Information gain based feature selection for bagging. *Int J Innov Comput Inf Control*. 2011.

7) Xia R, Zong C, Li S. Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences*. 2011;181(6):1138–1152. Available from: https://dx.doi.org/10.1016/j.ins.2010.11.023.

8) Duncan B, Zhang Y. Neural Networks for Sentiment Analysis on Twitter. In: and others, editor. Proceedings of 2015 IEEE 14th International Conference on Cognitive Informatics and Cognitive Computing, ICCI*CC 2015. 2015. Available from: https://doi.org/10.1109/ICCI-CC.2015.7259397.

9) Tan S, Zhang J. An Empirical Study of Sentiment Analysis for Chinese Documents. *Expert Systems with Applications*. 2008;34(4):2622–2629. Available from: https://doi.org/10.1016/j.eswa.2007.05.028.

10) Dhande LL, Patnaik PGK. Analyzing Sentiment of Movie Review Data using Naive Bayes Neural Classifier. *Int J Emerg Trends Technol Comput Sci [Internet]*. 2014;3(4):313–333. Available from: https://pdfs.semanticscholar.org/a030/b72c0546a3c832627855d42bccc3f6819e75.pdf.

11) Bhatt R, Gupta P. Sentiment Analysis. *Indian Journal of Science and Technology*. 2019;12(41):1–6. Available from: https://dx.doi.org/10.17485/ijst/2019/v12i41/145556.

12) Kummer O, Savoy J. Feature selection in sentiment analysis. In: and others, editor. CORIA 2012: Actes de la neuvieme COnference en Recherche d'Information et Applications - Proceedings of the 9th Conference on Information Retrieval and Applications. 2012.

13) Hira ZM, Gillies DF. A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data. *Advances in Bioinformatics*. 2015;2015:1–13. Available from: https://dx.doi.org/10.1155/2015/198363.

14) Yap BW, Rani KA, Rahman HAA, Fong S, Khairudin Z, Abdullah NN, et al. An Application of Oversampling, Undersampling, Bagging and Boosting in Handling Imbalanced Datasets. 2014.

15) Wang G, Sun J, Ma J, Xu K, Gu J. Sentiment classification: The contribution of ensemble learning. *Decision Support Systems*. 2014;57(1):77–93. Available from: https://dx.doi.org/10.1016/j.dss.2013.08.002.

16) Sharma A, Dey S. A Comparative Study of Selection and Machine Learning Techniques for Sentiment Analysis. In: and others, editor. Proceeding of the 2012 ACM Research in Applied Computation Symposium, RACS 2012. ;p. 1–7. Available from: https://doi.org/10.1145/2401603.2401605.

17) Dedhia C, Ramteke J. Ensemble Model for Twitter Sentiment Analysis. In: and others, editor. Proceedings of the International Conference on Inventive Systems and Control, ICISC 2017. Institute of Electrical and Electronics Engineers Inc. 2017. Available from: https://doi.org/10.1109/ICISC.2017.8068711.

18) Singh P. Sentiment Analysis Using Tuned Ensemble Machine Learning Approach. In: and others, editor. Lecture Notes in Networks and Systems. Springer. 2018;p. 287–97. Available from: https://doi.org/10.1007/978-981-10-8360-0_27.

19) Singh V, Saxena P, Singh S, Rajendran S. Opinion Mining and Analysis of Movie Reviews. *Indian Journal of Science and Technology*. 2017;10(19):1–6. Available from: https://doi.org/10.17485/ijst/2017/v10i19/112756.

20) Moraes R, Valiati JF, Neto WPG. Document-Level Sentiment Classification: An Empirical Comparison between SVM and ANN. *Expert Systems with Applications*. 2013;40(2):621–633. Available from: https://doi.org/10.1016/j.eswa.2012.07.059.

21) Julie MD, Kannan B. Attribute reduction and missing value imputing with ANN: prediction of learning disabilities. *Neural Computing and Applications*. 2012;21(7):1757–1763. Available from: https://dx.doi.org/10.1007/s00521-011-0619-1.