

Ensemble based Hybrid Machine Learning Approach for Sentiment Classification- A Review

Rabi Narayan Behera
Department of
Information Technology,
Institute of Engineering &
Management, Kolkata

Manan Roy
Department of
Information Technology,
Institute of Engineering &
Management, Kolkata

Sujata Dash
Department of
Computer Application,
North Orissa University,
Baripada, Orissa, India

ABSTRACT

Due to digital explosion, huge amount of data is generated from different sources which require critical analysis for decision making. In recent days one of the challenging issues like sentiment classification has drawn the attention of many researchers working in the area of opinion mining. The supervised machine learning technique is used for analyzing sentiments associated with unstructured text data. But, recently it has been observed from the findings that ensemble based learning algorithm achieves better understanding and acceptance of the solution in terms of diversity and accuracy. In this paper, an extensive study of ensemble based machine learning techniques in the domain of sentiment classification has been done to enhance the efficiency, by adopting multiple learning algorithms to obtain better predictive performance, that would be obtained from any of the constituent learning algorithms. Again, how the analysis will become stronger, some suggestions are proposed at the end of the discussion.

Keywords

Sentiment Analysis, Ensemble learning, Classification, Random Subspace, Bagging, Boosting

1. INTRODUCTION

Sentiment analysis (SA) is the computational study in which machine analyzes and classifies the human's sentiments, emotions, and opinions about any topic which are expressed in the form of either text or speech. It can be referred as the computational identification and categorization

of opinions expressed in a piece of text. Sentiment analysis corresponds to the method of extracting subjectivity and polarity from opinion text. Sentiment analysis deals with some challenging issues like a particular opinion word which can be considered to be positive in one domain may be considered as negative in another domain, for example long is considered as positive in "the battery life of the phone is long" whereas long is considered as negative in "the processing speed of the phone is long". Another challenge is that people do not always express opinions in the same way and can be contradictory in their statements which leads to poor understanding what others thought based on a short piece of text of lacking context, for example, "This movie is better than the last movie" is entirely dependent on what the person expressing the opinion thought for the previous movie.

SA is broadly categorized into lexicon based and machine-learning based approaches as shown in fig 1[1].

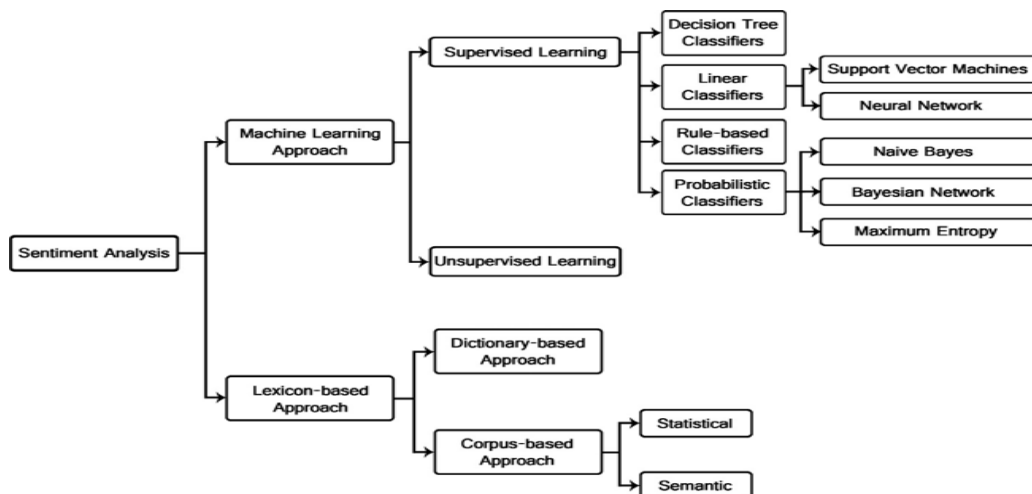


Fig 1: Sentiment Analysis Techniques

The Lexicon-based Approach depends on a sentiment lexicon, which is a collection of known and precompiled sentiment terms. The lexicon based approach is based on the assumption that the contextual sentiment orientation is the sum of the sentiment orientation of each word or phrase.

Machine Learning Approach is a field of artificial intelligence that trains the model from the existing data in order to forecast future behaviours, outcomes, and trends with the new test data.

It is classified into supervised and unsupervised learning- Supervised learning (SL) is the machine learning task of

inferring a function from labelled training data while unsupervised learning (USL) is a type of machine learning algorithm used to draw inferences from datasets consisting of unlabelled data. In supervised learning, there are defined rules, and the outcomes are known while in unsupervised learning, the algorithm follows certain rules to learn by itself and comes up with the result.

Many sentiment analysis tasks comprised of some popular classification algorithms namely Naive Bayes (NB), Support Vector Machines (SVM), K-Nearest Neighbours (KNN), Maximum Entropy (ME) etc, but it is not clear which of these perform more appropriately to achieve the desired accuracy in a generalized domain. Data classification is the categorization of data for its most effective and efficient use. Data classification is a two-step process:

1. Learning (Model Construction)
2. Classification (Model Usage)

In the first step, a classifier is built by describing a pre-determined set of data classes or concepts. This is the learning step (or training phase), where a classification algorithm builds the classifier by

learning from a training set made up of database tuples and their associated class labels. In the second step, the model is used for classification. A test set is used, made up of test tuples and their associated class labels. These tuples are randomly selected from the general data set. They are independent of the training tuples, meaning that they are not used to construct the classifier. However, it has been observed from significant amount of research in recent years that ensemble based learning outperforms the inefficiency of a single baseline classifier in classification prediction accuracy. A classifier ensemble is a group of classifiers whose individual decisions are merged in some manner to provide, as an output, a consensus decision.

2. LITERATURE SURVEY

Sentiment Classification techniques can be divided into machine learning approach, lexicon based approach and hybrid approach. The Machine Learning (ML) Approach applies some famous ML algorithms and uses linguistic features. The lexicon-based approach depends on finding the opinion lexicon which is used to analyze the text and the hybrid approach combines both the approaches.

A text classification problem can be viewed as a set of training records $X = \{x_1, x_2, x_3, \dots, x_n\}$ where each record is labelled to a class. The classification model is related to the features in the underlying record to one of the class labels.

Then for a given instance of unknown class, the model is used to predict a class label for it.

The text classification methods using ML approach can be further divided into supervised and unsupervised learning methods as described above. The supervised methods make use of a large number of labelled training datasets while the unsupervised methods are used when it is difficult to find the labelled training datasets.

A widely used supervised probabilistic classifier is Naive Bayes. It is a simple but an effective classification algorithm. The Naive Bayes algorithm is frequently used algorithm for document classification [2]. The basic idea is to estimate the probabilities of categories in a given test document by using the joint probabilities of words and categories. The naive part of such a model is the assumption of word independence. The

simplicity of this assumption makes the computation of Naive Bayes classifier far more efficient. Considering a training set of samples, each with the class labels T , there are k classes, C_1, C_2, \dots, C_k . With every sample consisting of an n dimensional vector, $X = \{x_1, x_2, \dots, x_n\}$, representing n measured values of the n attributes, A_1, A_2, \dots, A_n , respectively, the classifier will classify the given sample X such that it belongs to the class having the highest posterior probability. That is X is predicted to belong to the class C_i if and only if $P(C_i|X) > P(C_j|X)$ for $1 \leq j \leq m, j \neq i$. Thus it is

found that the class that maximizes $P(C_i|X)$. The maximized value of $P(C_i|X)$ for class C_i is called the maximum posterior hypothesis.

Hanhoon Khang et al. [3] has proposed an improved version of the Naive Bayes algorithm and a unigrams + bigrams was used as the feature, the gap between the positive accuracy and the negative accuracy was narrowed to 3.6% compared to when the original Naive Bayes was used, and that the 28.5% gap was able to be narrowed compared to when Support Vector Machine was used.

Support Vector Machine, a discriminative linear classifier is considered the best text classification method [4]. The main goal of SVM is to determine linear separators in the search space which can best separate the different classes. Based on the structural risk minimization principle from the computational learning theory, SVM seeks a decision surface to separate the training data points into two classes and makes decisions based on the support vectors that are selected as the only effective elements in the training set. Multiple variants of SVM have been developed in which multi-class SVM is used for sentiment classification. M. Rushdi Saleh et al. [5] has applied the new research area by using Support Vector Machines for testing different domains of data sets and using several weighting schemes. They have accomplished experiments with different features on three corpora. Two of them have already been used in

several works. The SINAI Corpus has been built from Amazon.com specifically in order to prove the feasibility of the SVM for different domains.

The K-nearest neighbour is a typical example based classifier that does not build an explicit, declarative representation of the category, but relies on the category labels attached to the training documents similar to the test document. Given a test document d , the system finds the k nearest neighbours among training documents. The similarity score of each nearest neighbour document to the test document is used as the weight of the classes of the neighbour document.

The Maximum Entropy classifier is a probabilistic classifier which belongs to the class of exponential models. The Maximum Entropy does not assume that the features are conditionally independent of each other. Maximum Entropy classifier is used when there is no scope to assume the conditional independence of the features. This is particularly true in text classification problems where features are usually words which are not independent.

ME classifier was used by Kaufmann [6] to detect parallel sentences between any language pairs with small amounts of training data. The other tools that were developed to automatically extract parallel data from non-parallel corpora use language specific techniques or require large amounts of training data. Their results showed that ME classifiers can produce useful results for almost any language pair. This can allow the creation of parallel corpora for many new languages.

Ensemble learning is selected over individual classifiers because of accuracy where a more reliable mapping can be obtained by combining the output of multiple classifiers and efficiency in which a complex problem can be decomposed into multiple sub problems so that it becomes easier to understand and solve [7].

Ensemble learning is the process by which multiple models, such as classifiers, are strategically generated and combined to solve a particular computational intelligence problem. Some researchers suggest that the performance of the ensembles depends on two properties, which are the individual success of the base classifiers of the ensemble and the independence of the base classifier's results from each other, while some other researcher suggests that the accuracy of individual models, diversity among the individual models, decision making strategy, and number of base classifiers used for constructing an ensemble are among the factors responsible for the success of an ensemble. An ensemble is itself a supervised learning algorithm because it can be trained and then used to make predictions. The trained ensemble, therefore, represents a single hypothesis. This hypothesis, however, is not necessarily contained within the hypothesis space of the models from which it is built. Thus, ensembles can be shown to have more flexibility in the functions they can represent. Ensembles combine multiple hypotheses to form a (hopefully) better hypothesis. In other words, an ensemble is a technique for combining many weak learners in an attempt to produce a strong learner. The term ensemble is usually reserved for methods that generate multiple hypotheses using the same base learner. Ensemble learning is primarily used to improve the classification, prediction, function approximation, etc., performance of a model, or reduce the likelihood of an unfortunate selection of a poor one. Other applications of ensemble learning include assigning a confidence to the decision made by the model, selecting optimal (or near optimal) features, data fusion, incremental learning, non-stationary learning and error-correcting.

Ensemble methods had been widely applied in many fields such as web ranking algorithm, classification and clustering, time series and regression problems, and water quality application, among others.

When constructing an ensemble, the ensemble size affects the accuracy of the ensemble. If there are a smaller number of individual classifiers, then the ensemble will not perform properly, whereas if there is a large number of individual classifier, the ensemble accuracy improves but will lead to increase of storage space and computational time. Due to the above problem, selective or pruning ensemble was proposed by many literatures that determine the optimal number of base classifiers for the construction of the ensemble system. It also reduces the complexity, storage requirements and enhances the performance of the system.

The ensemble framework can be classified into dependent and independent frameworks for building ensembles. In a dependent framework the output of a classifier is used in the construction of the next classifier. Thus it is possible to take advantage of knowledge generated in previous iterations to guide the learning in the next iterations. Alternatively, each classifier is built independently and their outputs are combined in some fashion.

There are two types of dependent learning namely,

Incremental Batch Learning- In this method the classification produced in one iteration is given as "prior knowledge" to the

learning algorithm in the following iteration. The learning algorithm uses the current training set together with the classification of the former classifier for building the next classifier. The classifier constructed at the last iteration is chosen as the final classifier.

Model-guided Instance Selection- In this dependent approach, the classifiers that were constructed in previous iterations are used for manipulating the training set for the following iteration. One can embed this process within the basic learning algorithm. These methods usually ignore all data instances on which their initial classifier is correct and only learn from misclassified instances.

While in independent framework, the original dataset is transformed into several datasets from which several classifiers are trained. The datasets created from the original training set may be disjointed (mutually exclusive) or overlapping. A combination method is then applied in order to output the final classification. Moreover, this methodology can be easily parallelized. These independent methods aim at either improving the predictive power of classifiers or decreasing the total execution time.

There are primarily three traditional ensemble methods namely bagging, boosting and random subspace. Bagging is an independent learning while boosting is a model-guided instance selection dependent learning.

Bootstrap aggregating or Bagging is a machine-learning ensemble meta algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression. It also reduces variance and helps to avoid overfitting. Although it is usually applied to decision tree classifier, it can be used with any type of base learners. It produces several different training sets of the same size with replacement and then builds a model for each one using the same machine learning scheme. Finally, it combines predictions by voting for a nominal target or averaging for a numeric target. In 1999, David Oritiz et al [8] evaluated on 23 data sets using both neural networks and decision trees as their classification algorithm. The data sets were obtained from the University of Wisconsin Machine Learning repository as well as the UCI data set repository (Murphy & Aha, 1994). Results clearly indicate that while Bagging is almost always more accurate than a single classifier, it is sometimes much less accurate than Boosting. On the other hand, Boosting can create ensembles that are less accurate than a single classifier, especially when using neural networks. Further, Boosting ensembles may overfit noisy data sets, thus decreasing its performance. Boosting is an algorithm based on an ensemble of similar to random forests. Instead of creating trees from different random subsets, boosted trees take the error from the previous tree and use it to improve the next one. It is an iterative algorithm. The idea is that create a model, and then take a look at the instance that are misclassified (The hard one to classify). Put extra weight on those instances to make a training set for producing the next model in the iteration. These encourage the new models to become an "expert" for instance that were misclassified. Finally, it uses voting for selecting the final prediction but weights models according to their performance.

Torralba, A et al [9] presented a multi-class boosting procedure (joint boosting) that reduces both the computational and sample complexity, by finding common features that can be shared across the classes. The detectors for each class are trained jointly, rather than independently. For a given performance level, the total number of features required is

observed to scale approximately logarithmically with the number of classes. In addition, they found that the features selected by independently trained classifiers are often specific to the class, whereas the features selected by the jointly trained classifiers are more generic features, such as lines and edges. If there is no structure to the features or if there is a limited amount of time to spend on a problem, one should definitely consider boosted trees. Furthermore, when random subsets of the dataset are drawn as random subsets of the features, then the method is known as Random Subspaces (RSM). In the RSM, the training set is also modified as in bagging. However, this modification is performed in the feature space (rather than example space). Let each training example X_j in the training sample set S be a p -dimensional vector $X_j = (x_{j1}, x_{j2}, \dots, x_{jp})$. In the RSM, p^* features are selected randomly from the training set S , where $p^* < p$. By this, the p^* dimensional random subspace is obtained from the original p -dimensional feature space. Therefore, the modified training set $S^* = \{X_1^*, X_2^*, \dots, X_n^*\}$ consists of p^* -dimensional training examples $X_j^* = (x_{j1}, x_{j2}, \dots, x_{jp^*})$ ($j=1, 2, \dots, n$). Afterwards, base-level classifiers are constructed from the random subspaces S_i^* (of the same size), $i=1, 2, \dots$, and they are combined by a voting scheme to obtain a final prediction. The RSM algorithm is a preferable option for classification problems where the number of features is much larger than the number of training objects.

Bagging is useful for weak and unstable classifiers where small changes in the training set result in large changes in predictions with a non-decreasing learning curve and critical training sample sizes. Boosting is beneficial only for weak, simple classifiers, with a non-decreasing learning curve, constructed on large training sample sizes [10]. Bagging has the advantage over Boosting in that it reduces variance and minimizes error [11]. The objective of boosting is to construct a composite classifier that performs well on the data, but a large number of iterations may create a very complex composite classifier, that is significantly less accurate than a single classifier. A possible way to avoid overfitting is by keeping the number of iterations as small as possible.

The random subspace method is advantageous for weak and unstable classifiers that have a decreasing learning curve and are constructed on small and critical training sample sizes viz. bagging, boosting etc, the selection of baseline classifier although is problem dependent, but on a generalized prospective if accuracy is highest priority then one must prefer a classifier model like Random Forest that consumes high learning time but has best accuracy. If processing power and memory is an issue, then the Naïve Bayes classifier should be selected due to its low memory & processing power requirements. If less training time is available but one have a powerful processing system and memory, then Max Entropy proves to be a worthy alternative and if one need to select a classifier that is average on all aspects then Boosted Trees might be the right choice.

Research by [12] in 2008, developed a new method for building dynamic ensemble from a collection of classifiers to predict sea water quality from spectrum channel data.

Genetic Algorithm (GA) search was used to optimize the ensemble. Experimental results were compared with SVM algorithm. Results showed that their method outperformed SVM, but the performance of the ensemble is critically affected by the quality of the population in the ensemble.

In 2010, the author in [13] developed a heterogeneous

ensemble and a framework for constructing different kinds of ensemble for classifying spam emails. Results indicate that the heterogeneous ensemble can increase diversity as well as performance when compared to individual classifiers and other ensemble models.

In 2011, the authors in [14] presented a novel ensemble classifier architectures and investigate their influence for offline cursive character recognition. Cursive characters are represented by feature sets that portray different aspects of character images for recognition purposes. The recognition accuracy can be improved by training ensemble of classifiers on the feature sets. Given the feature sets and the base classifiers, the authors have developed multiple ensemble classifier compositions under four architectures. The first three architectures are based on the use of multiple feature sets whereas the fourth architecture is based on the use of a unique feature set. Type-1 architecture is composed of homogeneous base classifiers and Type-2 architecture is constructed using heterogeneous base classifiers. Type-3 architecture is based on hierarchical fusion of decisions. In Type-4 architecture a unique feature set is learned by a set of homogeneous base classifiers with different learning parameters. The experimental results demonstrate that the recognition accuracy achieved using Type-4 ensemble classifier is better than the other recognition accuracies for offline cursive character recognition.

Further, in the following year in 2012, Ying Su et al [15] presented an ensemble method for sentiment classification of reviews. The diversity among the machine learning algorithms for sentiment classification such as different features, different weight measures and the modeling of negation, is investigated in three domains, which gives a space for improving the performance. Then the ensemble learning framework, stacking generalization is introduced based on different algorithms with different settings, and compared with the majority voting. According to the characteristic of reviews, the opinion summary of review is proposed in their work, composed of the first two and last two sentences of review. Their results showed that stacking has been proved to be effective in almost every domain, working better than majority voting, and that using the opinion summary can further improve the performance.

In 2013, Gang Wang et al [16] conducted a comparative assessment of the performance of three ensemble learning methods viz. bagging, boosting and random subspace based on five learners namely naive bayes, maximum entropy, decision tree, k-nearest neighbor and support vector machine. They randomly chose ten sentiment analysis datasets using 10-fold cross validation for their work. Based on a total of 1200 comparative group experiments, their results revealed that ensemble methods improve the performance of individual base learners for sentiment classification.

E. Fersini et al. [17] in the next year, developed a bayesian ensemble learning approach that takes advantages of multiple classifiers to predict the sentiment orientation of user-generated contents and improve the performance of polarity classification tasks. They used Review datasets viz. Sentence polarity dataset, Fine-grained Sentiment Dataset, Multi-Domain Sentiment Dataset and social dataset namely the gold standard dataset. They addressed the classifier selection problem by proposing a greedy approach that evaluates the contribution of each model with respect to the ensemble. Experimental results on gold standard dataset showed that their proposed approach outperforms both traditional classification and ensemble method.

In the same year in 2014, the researchers in [18] aimed to improve random forest accuracy by optimizing large number of decision trees within the forest by choosing only uncorrelated and good trees. They improve the accuracy of the enhanced random forest through maximization of individual trees strength and minimize the correlation between the trees in the forest. An increased difference from 1% to 6% was achieved with the four experimental datasets and M.Govindarajan [19] has implemented different ensemble based techniques like bagging and boosting with different types of baseline classifiers as Naive Bayes, Support Vector Machine and Genetic Algorithm on Movie Review dataset and established a claim that ensemble based techniques provides better accuracy compared to individual classifiers.

Matthias Hagen, Martin Potthast, Michel Buchner ,and Benno Stein [20] reproduced three approaches to classifying the sentiment expressed in a given tweet as either positive, neutral, or negative, and combine the three approaches to an ensemble averaging the individual classifiers' confidence scores for the three classes and deciding a sentiment polarity based on these averages. Experimental evaluation on SemEval dataset showed that the reimplementation to slightly outperform their respective originals. However, an error analysis shows that the ensemble classifier makes misclassifications, such as identifying a positive sentiment in a negative tweet or vice versa into neutral tweets which are far more safe.

In 2015, Yongjun Piao et al.[21] proposed an ensemble method for classification of high-dimensional data, with each classifier constructed from a different set of features determined by partitioning of redundant features. In their method, the redundancy of features is considered to divide the original feature space. Then, each generated feature subset is trained by a support vector machine, and the results of each classifier are combined by majority voting. The efficiency and effectiveness of our method are demonstrated through comparisons with other ensemble techniques, and the results show that our method outperforms other methods.

In the same year, Junyi Xu et al. [22] presented a ensemble learning method namely argumentation based multi-agent joint learning (AMAJL), which integrates ideas from multi agent argumentation, ensemble learning, and association rule mining. In AMAJL, argumentation technology is introduced as an ensemble strategy to integrate multiple base classifiers and generate a high performance ensemble classifier. They designed an argumentation framework named Arena as a communication platform for knowledge integration. Through argumentation based joint learning, high quality individual knowledge can be extracted, and thus a refined global knowledge base can be generated and used independently for classification. They performed numerous experiments on multiple public datasets using AMAJL and other benchmark methods. The results demonstrated that their method can effectively extract high quality knowledge for ensemble classifier and improve the performance of classification.

Also, Amine Bayoudhi et al. [23] tried to improve to improve the document classification findings in Arabic sentiment analysis by combining different types of features such as opinion and discourse features; and by proposing an ensemble-based classifier to investigate its contribution in Arabic sentiment classification. Results achieved showed an attainment of 85.06% in terms of macro-averaged F-measure, and showed that discourse features have moderately improved F-measure by approximately 3% or 4%.

Xueyi Wang [24] proposed a novel approach known as COB(core, outlier, and boundary) which quantitatively measures the accuracies of majority voting ensembles for binary classification. The author claimed that good ensemble methods require accurate and diverse individual classifiers. The data items were initially divided into three subsets, core, outlier and boundary based on the predictions of each individual classifier of ensembles and finally the accuracy of the ensemble method was recorded

for each of the subset and the results were combined together. In the work, the author used bagging, random forest and a randomized ensemble as three different ensembles and decision trees, k-nearest neighbours, and support vector machines were used as the machine learning algorithm for the models.

The performance of the proposed model was tested upon 32 datasets collected from the UCI repository. The purpose of the proposed model was to improve the performance of the data and predict better accuracy and the experiments showed that the COB model performed better than the binomial model .It was also observed to suggest that to achieve a high accuracy for an ensemble method, weak individual classifiers should be partly diverse instead of fully diverse, that is, be diverse on correctly predicted items but in agreement on some incorrectly predicted items.

Maher Ala'raj and Maysam Abbod [25] found that lending loans to borrowers is the main profit sources for banks and financial institutions. So, careful assessment and evaluations required while granting credit to borrowers. In 2015, they introduced a systematic credit scoring model based on both the homogenous and heterogeneous classifier ensembles using logistic regression (LR), artificial neural network (ANN) and support vector machines as classifiers as they stressed on using multiple classifiers over single ones to solve the credit scoring problems. Results showed that heterogeneous classifier ensembles give improved performance than homogenous and single classifiers.

Suresh Ramakrishnan et. al [26] introduced an alternative technique to improvise the performance rate of corporate default prediction problem by incorporating data mining techniques due to its ability to notice non-linear relationships and show a good performance in presence of noisy information. An ensemble based model was used to classify default and non-default Malaysian firms under Bursa Malaysia. Results revealed that using multi-stage classifiers with Adaboost increases the performance over single classifiers.

Gang Wang and Jian Ma [27] also integrated random subspace and boosting to predict credit risk problem. They called it as RS-boosting which was evaluated on two corporate credit datasets. Results revealed that their proposed model demonstrated better performance among logistic regression analysis (LRA), decision tree (DT), artificial neural network (ANN), bagging, boosting and random subspace.

3. CONCLUSION AND FUTURE WORK

From the above discussion, ensemble based learning techniques for sentiment classification primarily used to improve the classification, prediction, function approximation, and performance of a model, or reduce the likelihood of an unfortunate selection of a poor one. An AdaBoost based ensemble technique can be introduced to handle the multi-class and regression problems by manipulating the weight factor of the correctly and incorrectly classified instance to

achieve the desired accuracy.

4. REFERENCES

- [1] Walaa Medhat, Ahmed Hassan, Hoda Korashy (2014). Sentiment Analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, Elsevier, Volume 5, Issue 4, pp. 1093-1113
- [2] Melville, Wojciech Gryc (2009). Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification. *ACM*.
- [3] Hanhoon Kang, Seong Joon Yoo, Dongil Han (2012). Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications*.
- [4] Rui Xia, Chengqing Zong, Shoushan Li (2011). Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences*.
- [5] M. Rushdi Saleh, M. T. Martín Valdivia, Ráez L. A. Urena-Lopez, A. Montejo (2011). Experiments with SVM to classify opinions in different domains. *Expert Systems with Applications*.
- [6] Kaufmann JM. JMax Align: A Maximum Entropy Parallel Sentence Alignment Tool. In: *Proceedings of COLING'12: Demonstration Papers*, Mumbai; 2012. p. 277–88.
- [7] website: <http://publications.drdo.gov.in/ojs/index.php/dsj/article/view/1088/4752>
- [8] David Opitz, Richard Maclin (1999). Popular Ensemble Methods: An Empirical Study *Journal of Artificial Intelligence Research* pp.169-198.
- [9] Antonio Torralba, Kevin P. Murphy, William T. Freeman (2004). Sharing features: Efficient boosting procedures for multiclass object detection. MIT, Cambridge
- [10] Martin Sewell (2007). Ensemble Learning. *Research Note*. RN/11/02.
- [11] Anwar, H., Qamar, U., & Muzaffar Qureshi, A. W. (2014). Global Optimization Ensemble Model for Classification Methods. *The Scientific World Journal*, Hindawi Publishing Corporation.
- [12] Zeng, B., Luo, Z., & Wei, J. (2008). Sea Water Pollution Assessment Based on Ensemble of Classifiers. In *Natural Computation, ICNC'08*.
- [13] Fourth International Conference on Vol. 1, pp. 241-245. IEEE Wang, W. (2010). Heterogeneous Bayesian Ensembles for Classifying Spam Emails. *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8, IEEE.
- [14] A. Rahman and B. Verma (2011). Ensemble Classifier Composition: Impact on Feature Based Offline Cursive Character Recognition, *Proc. IEEE International Joint Conference on Neural Networks (IJCNN)*, San Jose, USA.
- [15] Ying Su, Yong Zhang, Donghong Ji, Yibing Wan, Hongmiao Wu (2012). Ensemble learning for Sentiment Classification. *Springer*.
- [16] Gang Wang, Jianshan Sun, Jian Ma, Kaiquan Xu, Jibao Gu (2013). Sentiment classification: The contribution of ensemble learning. *Elsevier*
- [17] E. Fersini, E. Messina, F.A. Pozzi (2014). Sentiment analysis: Bayesian Ensemble Learning. *Elsevier*.
- [18] Bharathidasan, S., & Jothi Venkataeswaran, C. (2014). Improving Classification Accuracy based on Random Forest Model with Uncorrelated High Performing Trees. *International Journal of Computer Applications*, vol. 101, issue 13, pp. 26-30.
- [19] M. Govindarajan (2014). Bagged Ensemble Classifiers for Sentiment Classification of Movie Reviews. *IJECS Volume 3, Issue 2*.
- [20] Matthias Hagen, Martin Potthast, Michel BÜchner, Benno Stein (2015). Webis: An Ensemble for Twitter Sentiment Detection. In the proceedings of the 9th International Workshop on Semantic Evaluation held on June 4-5 pp. 582–589. Denver, Colorado
- [21] Yongjun Piao, Minghao Piao, Cheng Hao Jin, Ho Sun Shon, Ji Moon Chung, Buhyun Hwang, Keun Ho Ryu (2015). A New Ensemble Method with Feature Space Partitioning for High-Dimensional Data Classification. *Mathematical Problems in Engineering*, Article ID 590678
- [22] Junyi Xu, Li Yao, Le Li (2015). Argumentation Based Joint Learning: A Novel Ensemble Learning Approach. *Plos One*.
- [23] Amine Bayoudhi, Lamia Hadrich Belguith, Hatem Ghorbel (2015). Sentiment Classification of Arabic Documents: Experiments with multi-type features and ensemble algorithms. In the proceedings of 29th Pacific Asia Conference on Language, Information and Computation held on October 30–November 1, pp. 196–205, Shanghai, China.
- [24] Xueyi Wang (2012). A new model for measuring the accuracies of majority voting ensembles. In proceedings of IEEE International Joint Conference on Neural Networks (IJCNN), Brisbane, QLD
- [25] Maher Ala'raj, Maysam Abbod (2015). A systematic credit scoring model based on heterogeneous classifier ensembles. In the proceedings of the IEEE 2015 International Symposium on Innovations in Intelligent Systems and Applications (INISTA), Madrid
- [26] Suresh Ramakrishnan, Maryam Mirzaei, Mahmoud Bekri (2014). Adaboost Ensemble Classifiers for Corporate Default Prediction. In the Proceeding of 1st International Conference of Recent Trends in Information and Communication Technologies, Johor, Malaysia
- [27] Gang Wang, Jian Ma (2011). Study of corporate credit risk prediction based on integrating boosting and random subspace. *Expert Systems with Applications*, volume-38, issue-11, pp. 13871-13878