Analysis of Voting Techniques in Ensemble based Learning

Meenakshi A. Thalor[*]

[*] Department of Computer Engineering, AISSMS Institute of Information Technology, Pune, Maharashtra, India
thalor.meenakshi@gmail.com

Abstract- Now a day's data analytics and data science is at boom which works on big data and come up with some hidden patterns after data mining and analysis. Classification process on big data is a challenge as single classifier is unable to give an unbiased decision because of plasticity–stability problem. This paper gives the introduction about the ensemble based learning where multiple base classifiers participate in predicting the class of unlabeled data. Different voting techniques to combine the outcome of different base classifiers of ensemble are also explained. At the end a strict comparison is drawn to show the impact of voting technique in decision of ensemble.

Index Terms- Classification, Base Classifiers, Ensemble, Voting

I. ENSEMBLE BASED LEARNING

Sometimes the data on which one wants to perform learning is so huge that it's not possible to process or analyze such vast amount of data by one classifier. This huge amount of data can be analyze by partitioning data into smaller chunks and pass smaller chunks to multiple classifiers(ensemble) rather than one classifier as shown in fig. 1. Once the learning over different chunk of data is done or model is ready the unseen data can be passes to different classifiers for classification.

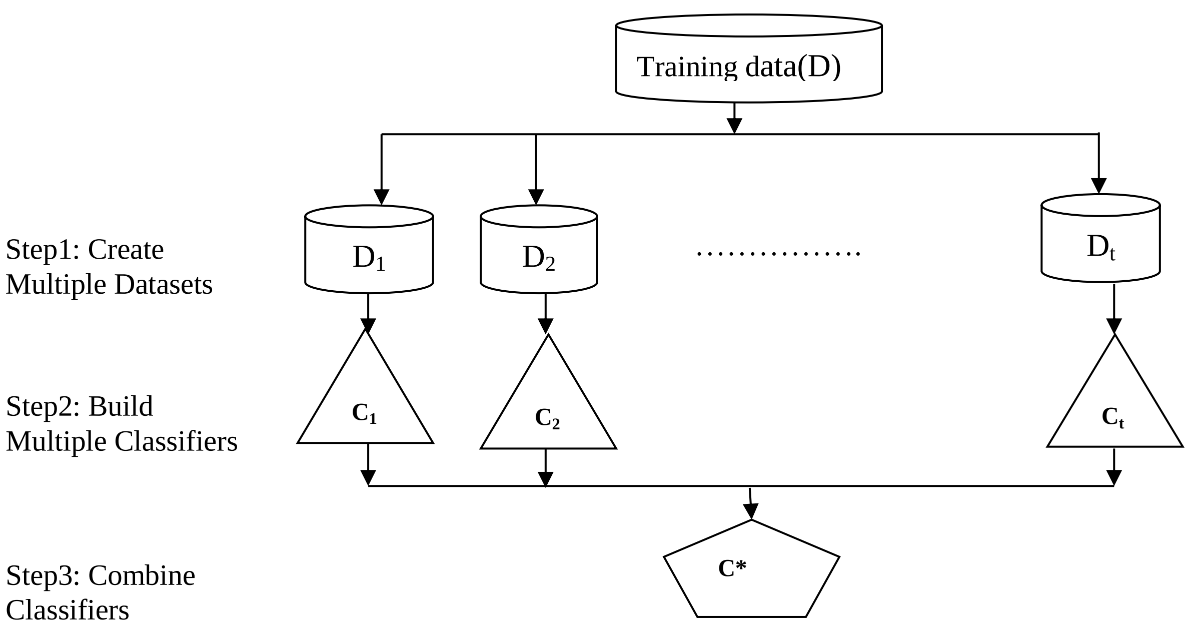

Fig. 1: Ensemble based Learning

Each classifier will predict the class of unseen data then by using different voting approaches one can combine the outcome and come to a final class conclusion. The Ensemble based classification [1,2] achieved by grouping the classifiers by some way then individual predictions of each classifier is combined to classify unseen data.

II. COMPONENTS OF ENSEMBLE

Ensemble is combination of set of diverse classifiers, which collectively do decision making and predict the label for unseen data. Generation of classifiers in ensemble can take place in different ways. Ensemble [3,4,5] can be generated in sequence, in parallel or in layered form. In order to do the classification using an ensemble approach, first one has to consider following elements of ensemble.

A. Training Set

For supervised learning, the basic requirement is a training dataset. A labeled training dataset consists of instances and their corresponding class variable. An instance is feature –value vector which can be represented in a variety of languages. Generally, set of instances are represented by X which consist of n attributes $X= \{x_1, ..., x_i , ..., x_n\}$ and Y to represent the class label or the target attribute.

B. Base Classifier

The base classifier is a classification algorithm that attains a training data and makes a model that specifies the association between the input variables and the class variables.

C. Diversity Generator

The main element which is responsible for creating the different classifiers is called diversity generator. In order to create diverse classifiers in ensemble various techniques are proposed in literature [6,7,8] as mentioned below:

a. *Using diverse training sets:*
   In order to create district classifier, diversity can be introduced in training set by partitioning it into N subsets to train individual classifier. Each member of ensemble will get a different part of training set and unique learning takes place.
b. *Using diverse feature subsets:*
   In order to create distinct classifier, diversity can be introduced by providing separate set of features to train individual classifier. Rather than dividing a training set, each member of ensemble will get a different part of feature/attribute set and unique learning takes place.
c. *Using diverse classifier models:*
   In order to create distinct classifier, diversity can be introduced by combining different types of individual base classifier. Here diversity in ensemble is introduced by keeping same training set and same set of features but using different base classifiers like naïve bayes, support vector machine and decision trees etc.
d. *Using diverse combination schemes:*
   In order to create district classifier, diversity can be introduced by using different types of combination rules. The way you combine the prediction of each base classifier is also introduce diversity in ensemble even though training data, feature set and base classifiers are same.

D. Combiner

The purpose of combiner is to join the prediction outcome of the various classifiers. The outcome of an ensemble depends on the selection of appropriate rule to unite the decision of each classifier. Voting rule is applied at the final step of ensemble system. In literature [9-14], various voting rules are presented which are explained in next section.

III. VOTING TECHNIQUES IN ENSEMBLE

Let x_t is testing instance, each base classifier(h_j) makes an prediction of class label for testing instance x_t which is given by as where Y_i=1….C where C represent no. of class labels.

A. Geometric Average Rule

Geometric Average Rule gives the prediction outcome of the testing instant x_t to that class which maximizes the product of as specified in eq. 1.

$$x_t \rightarrow Y_i \ satisfy \quad \max_{Y_i} \prod_{j=1}^{L} P_j(Y_i|x_t) \tag{1}$$

B. Arithmetic Average Rule

The Arithmetic Average rule is defined as discovering the maximal value of the arithmetic average of $P_j(Y_i|x_t)$ as specified in eq. 2.

$$x_t \rightarrow Y_i \ satisfy \quad \max_{Y_i} \frac{1}{L} \sum_{j=1}^{L} P_j(Y_i|x_t) \tag{2}$$

C. Median Rule

The Median Value rule will select the target class label whose median value is maximum using eq. 3.

$$x_t \rightarrow Y_i \ satisfy \quad \max_{Y_i} \left\{ median(P_j(Y_i|x_t)) \right\} \tag{3}$$

D. Majority Rule

In ensemble, each individual classifier provides the prediction of class label for testing sample and then majority voting rule select the final prediction as the one that takes the majority of votes from each individual classifier in the ensemble as specified in eq. 4. When there is tie among the class labels then a random class label is selected among all tie members.

$$x_t \rightarrow Y_i \ satisfy \quad \max_{Y_i} \sum_{j=1}^{L} \Delta_j(Y_i|x_t) \tag{4}$$

Where

$$\Delta_j(Y_i|x_t) = \begin{Bmatrix} 1; if\ h_j(x_t) = Y_i \\ 0; otherwise \end{Bmatrix}$$

E. Max Voting Rule

Max rule selects the target predicted class label depending on information provided by the maximal value of $P_j(Y_i|x_t)$ across all potential class labels as specified in eq. 5.

$$x_t \rightarrow Y_i \; satisfy \quad \max_{Y_i}\left\{\max_j\left(P_j(Y_i|x_t)\right)\right\} \tag{5}$$

F. Min Voting Rule

Min rule selects the target predicted class label depending on the maximal of the minimal values of $P_j(Y_i|x_t)$ across all potential class labels as specified in eq. 6.

$$x_t \rightarrow Y_i \; satisfy \quad \max_{Y_i}\left\{\min_j\left(P_j(Y_i|x_t)\right)\right\} \tag{6}$$

G. BC Rule

The Borda Count rule uses positional-scoring procedure where individual $Pj(Y_i|x_t)$ provides rank to each class label as specified in eq. 7. Based on the prediction given by each classifier, each classifier will rank all the candidate class labels. For a multi class problem, each class label gets 0 rank least vote received, 1 rank point for each next least vote, etc., up to C-1 points for majority vote received. Finally, the class label that receives largest votes will win the election.

$$x_t \rightarrow Y_i \; satisfy \quad \max_{Y_i} \sum_{j=1}^{L} \Omega_j(Y_i|x_t) \tag{7}$$

Where if classifier h_j ranked x_t in the kth position for class label Y_i , and C is the number of possible classes in multiclass problem.

H. Weighted Majority Voting Rule

A weight coefficient is consigned to each classifier in ensemble and after each prediction the weight of individual classifier is updated based on its performance. Here selects the target predicted class label depending on the weight of each individual classifier as specified in eq. 8.

$$x_t \rightarrow Y_i \; satisfy \quad \max_{Y_i} \sum_{j=1}^{L} w_j . \Delta_j(Y_i|x_t) \tag{8}$$

Where w_j is a weight coefficient for classifier h_j: w_j>=0 and $\sum_{j=1}^{L} w_j = 1$

IV. RESULTS AND ANALYSIS

The main aim of this paper is to determine which voting methods performed best in ensemble. Fig.2 depicts the performance of different voting methods where using diverse base classifier diversity is introduced in ensemble and accuracy is evaluated using different ensemble size. The “Pima Indian diabetes dataset” downloaded from “UCI Machine Learning Repository” is considered here for experimentation purpose. On small ensemble size, the product, min and max voting rules perform well but as ensemble size grow number of correctly classified instances decrease. On large ensemble size Avg., Weighted majority and borda count perform better as they consider the overall voting and less demanding than the confidence method. Lastly, the majority-vote performance is low as it rejects a large number of samples as they are not majority candidate which results the actual errors are much lower than the recognition performance suggests.

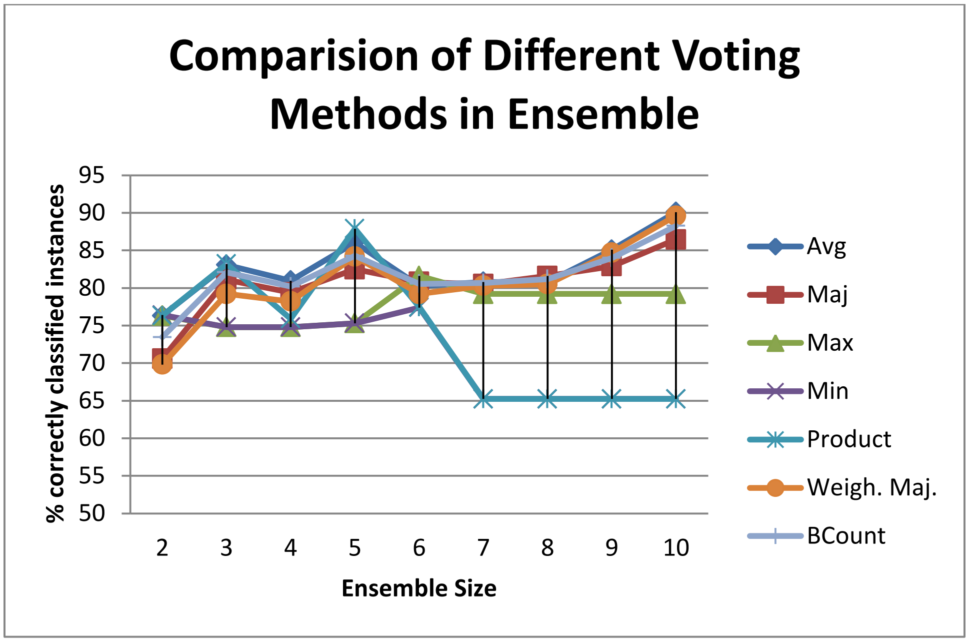

Fig. 2: Results of different voting methods

V. CONCLUSION

This paper gives a detail description of ensemble learning, basic components of ensemble, approaches to combine the decision of each classifier in ensemble and comparison of voting methods with respect to ensemble size. Based on review, conclusion is that the average voting and weighted majority voting are effective approaches as compare to others. The product and sum rule performed best on the smaller sized ensembles, while the Borda count gave good recognition results on larger ensembles. The weighted majority voting significantly improves the prediction accuracy on big data as consider the weight of each classifier.

REFERENCES

[1] L. Rokach, "*Ensemble-based Classifiers*," Artificial Intelligence Review, vol. 33, pp. 1-39, 2010.

[2] Robi Polikar, "*Bootstrap-Inspired Techniques in Computational Intelligence*", IEEE Signal Processing Magazine, 2007

[3] Meenakshi A. Thalor, Dr. S. T. Patil, "*Review of Ensemble Based Classification Algorithms for Nonstationary and Imbalanced Data*", IOSR Journal of Computer Engineering, Impact Factor 1.213, e-ISSN: 2278-0661, Vol. 16, pp. 103-107, February 2014.

[4] Meenakshi A.Thalor ,Dr. S.T.Patil, "*Comparison of Ensemble Based Classification Algorithms*", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4 Issue 4, Impact Factor 2.080, pp. 1126-1132 ,ISSN: 2277 128X, April 2014

[5] Meenakshi A.Thalor and S.T. Patil, "*Learning on High Frequency Stock Market Data Using Misclassified Instances in Ensemble*" International Journal of Advanced Computer Science and Applications (IJACSA), May 2016, Vol. 7, Issue 5, pp. 283-288.

[6] Meenakshi A. Thalor and S. T. Patil, "*Incremental Learning on Non-stationary Data Stream using Ensemble Approach*", International Journal of Electrical and Computer Engineering, CiteScore-0.37, SJR-0.164, August 2016,Vol. 6, No. 4, pp. 1811-1817.

[7] Meenakshi A. Thalor and S. T. Patil, "*Learning Framework for Non-stationary and Imbalanced Data Stream*", International Journal of Engineering and Technology (IJET), CiteScore-0.42, SJR-0.17, October – November 2016, Vol. 8, Issue 5, pp. 1942-1945.

[8] Meenakshi A. Thalor, Dr. S. T. Patil, "*Ensemble for Non Stationary Data Stream: Performance Improvement over Learn++.NSE*", IEEE International conference on Information Processing, December 2015.

[9] A.Salini, U. Jeyapriya A, "*Majority Vote Based Ensemble Classifier for Predicting Students Academic Performance*," International Journal of Pure and Applied Mathematics, Vol. 118, no. 24, 2018.

[10] K.Madasamy, Dr. M. Ramaswami, "*Parallel Heterogeneous Voting Ensemble for Effective Classification of Imbalanced Data*", International Journal of Innovative Research in Science,Engineering and Technology, Vol. 7, Issue 8, August 2018.

[11] N. Littlestone and M. Warmuth, "*Weighted majority algorithm*," Information and Computation, vol. 108, pp. 212–261, 1994.

[12] R. Battiti and A.M. Colla, "*Democracy in neural nets: Voting schemes for classification*", Neural Networks, vol. 7, no. 4, pp. 691–707, 1994.

[13] R.E. Schapire, Y. Freund, P. Bartlett, and W. S. Lee, "*Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods*," Annals of Statistics, vol. 26, no. 5, pp. 1651–1686, 1998.

[14] E. Bauer and R. Kohavi, "*An empiricial comparison of voting classification algorithms: bagging, boosting and variants*," Machine Learning,vol. 36, pp. 105–142, 1999.