International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-5, June 2019

Improving Prediction Capability of Ensemble of Classifiers through Weighted Average Probabilities

Princy Christy. A, N. Rama

Abstract: Ensemble of classifiers has been proved to significantly impact the performance outcome by improving the capability of performance classification in education scenario. An approach, Optimised Voting Ensemble (OVE) has been proposed to increase the prediction accuracy in determining the final class of students based on their continuous performance. The base learners chosen for this approach is decision tree, multi layer perceptron and stocastic gradient descent classifiers that are well suited for a multi-class problem. The approach tends to improve the classification and prediction accuracy by optimization of hyperparameters. The hyperparameters of base learners are tuned such that they improve their capability to classify and predict the final grade. The optimization is carried out through grid search. A weighted average probability method is used to combine the tuned base learners to form an ensemble. This tends to combine the strength of all the base learners and improve the prediction accuracy by assigning weight to strong classifiers.
Keywords: Optimization, student performance, voting ensemble, hyperparameter, prediction

I. INTRODUCTION

Combination of various classifiers to improve the performance accuracy for a given dataset is known as an ensemble. Ensemble of classifiers have been widely studied and employed for classification and prediction task. Ensemble uses the collective capability of base learners and tends to improve classification an prediction accuracy based on various implicit and explicit features of the base learners and the technique used to create an ensemble itself

As said ensembles improve performance and prediction accuracy compared to individual base learners or learning algorithms. Capability of ensembles can be further optimized by carefully identifying and choosing their parameters. This is called hyperparameter tuning. Various hyperperameter tuning techniques are available. They are grid search, Bayesian optimization, evolutionary optimization, random search, gradient based and population based techniques.

This study uses the simple and easy grid search optimization technique to identify suitable hyperparameters for each base learner since it has been found that hyperparameter tuning improved the efficiency of individual classifiers [1].

There are various methods of creating ensembles like bagging, boosting, voting and stacking. It has been decided to use voting scheme to create an ensemble. There are many applications and researches that use voting ensemble technique.

Manuscript published on 30 June 2019.
***** Correspondence Author (s)
Princy Christy. A, Research Scholar, Post Graduate and Research Department of Computer Science, Presidency College, Chennai, Tamil Nadu, India.
N. Rama, Principal, Dr. M.G.R. Government Arts and Science College for Women, Villupuram, 605602, Tamil Nadu, India

There are two voting schemes which are hard voting and soft voting. This research uses soft voting for the proposed approach, optimized voting ensemble (OVE) which creates a voting ensemble from the optimized base learners. The voting scheme assumes a weighted average probability. This article contains the following four sections. The second section reviews the related study made in the areas of this proposed scheme. The third section puts forth the idea of creating a voting ensemble using multi-classifiers for a multi-class approach and the architecture. The last section discusses on the observations made during development and implementation of the proposed approach.

II. RELATED WORK

Implementation of voting strategy to create ensemble has been explored by many researchers. Multi-objective majority weighted voting ensemble has been used in sentiment analysis [2]. Bashir, et.al., [3] demonstrated the efficiency of vote based ensemble in diagnosis of breast cancer. The study had used a weighted voting scheme based on the classification accuracy of the base learner. Ade and Deshmukh [4] have developed an incremental ensemble of classifiers through majority voting. In their research have identifies that the majority voting scheme has brought about increased classification accuracy compared to other voting schemes. A study to predict the performance of students had been carried out by building ensemble of classifiers [5]. The ensemble was constructed using four individual classifiers with a voting scheme which used majority voting and was proved to have increased the prediction capability compared to the individual classifiers. In a proposed work performed on prediction of students academic performance it has been proved that employing ensemble of classifiers improved performance accuracy [6]. This conclusion had been arrived after comparing the result with that of the individual classifiers and the ensembles developed using bagging, boosting and voting. The results clearly indicate the improvement in the prediction accuracy.

Leveseq et. al. [7] proposed a novel to improve performance of ensemble using Bayesian optimization technique for optimizing the hyperparameters of the base classifiers. The study has shown that there is significant increase in the predictive capability of the ensemble after tuning the hyperparameters of the base learners. The proposed approach by Stapel et al. [8] indicates the use of ensemble of classifiers built by soft voting. They had inferred that their approach yielded a robust performance prediction. Three classifiers were combined in a research to predict the performance of students by creating an ensemble using majority voting [9]. Satyanaraya et al [10] have an ensemble to predict performance of students.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

The study also had induced a filtering technique into the ensemble to improve the accuracy further. Many researches were carried out to increase the performance prediction by employing ensembles [11-13]. A majority weighted voting scheme was proposed by Kim et al. [13].

III. PROPOSED APPROACH

The proposed approach is to build a voting ensemble with classifiers subjected to hyperparameter tuning to optimize its prediction accuracy. The architecture of the proposed approach is given in Fig.1.

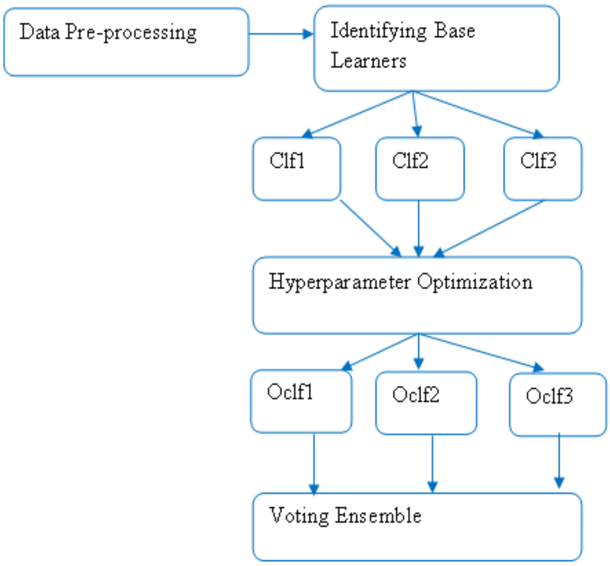

Fig. 1: Architecture of OVE

Initially the dataset was checked for any missing or duplicate values. Feature engineering was also performed. The proposed approach is to identify base learners depending on the dataset. Optimize the performance of these base learners through hyperparameter tuning and create a voting ensemble with soft voting, weights assigned to each of the base learners. To begin with base learners were identified based on their capacity to classify the education dataset that had multi-class labels. The following section discusses on the development and implementation of the OVE architecture.

A) Model Building

a) Identifying Base Learners

Decision tree (DT) has been identified since it learns simple decision rules inferred from the data features. It then creates a model that predicts the target variable. It uses common measures of impurity. In a classification scenario considering values 0,1,.... c-1 for node *n*, denoting a region R_n with N_n then the proportion of class k observations in node *n* is

$$p_{nc} = \frac{1}{N_n} \sum_{x_i \in R_n} I(y_i = c) \qquad \text{(Eq: 1)}$$

The common measures of impurity at *n* is given by the impurity function (Eq: 2) where X_n is the training data in node *n*

$$H(X_n) = \sum_k p_{nc}(1 - p_{nc}) \qquad (Eq{:}\,2)$$

Stochastic Gradient Descent (SGD) classifier and estimator employ linear models. The partial derivatives of the set of parameters with respect to the inputs (gradients) of the loss are estimated for each sample at a time and the model is updated along the way with a decreasing learning rate. It fits a linear Support Vector Machine (SVM) which is controlled with the loss parameter. SGD uses a regularizer which is a penalty added to the loss function using the absolute norm (L1) that shrinks the model parameters towards the zero vector. There are other penalty methods like the squared euclidean norm (L2) or elastic net a combination of both L1 and L2. These helps to allow for learning sparse models and achieve online feature selection if the parameter update crosses the value 0.0. The epoch parameter can be set accordingly for this estimator. Multilayer Perceptron (MLP) is a feed-forward supervised learning algorithm. The network is a function of one or more inputs or predictors that reduces prediction error of one or more target variables. This can handle both categorical and continuous values as well. The activation function which is a mapping of summed weighted input to the output of neuron governs the threshold at which the neuron is activated and the strength of the output signal as well. The identity activation of the hidden layer is a no-op activation used to implement linear bottleneck. The solver for weight optimization used is lbfgs, an optimizer in the family of quasi-Newton methods. Their performance accuracy and standard deviation was identified to check if they can be used to form a voting ensemble.

b) Hyperparameter optimization

The identified ensembles were subjected to hyperparameter tuning to optimize the performance of the base learners. This significantly increases their classification accuracy. It has been identified in the research that hyperparameter optimization of the chosen classifiers has considerably decreased the standard error rate of the classifiers [1]. This has further lead to the increase in classification accuracy. The study also indicates that the prediction accuracy of each of the classifiers DT and SGD has shown a increase of 2% and 12% respectively. While the prediction accuracy of MLP increased from 24% to 82%. Grid searchCV is done the following way. It performs a Cartesian product on the sets of heperparameter values provided and evaluates the performance on a internal cross-validation on the training set. The set of values that produced the highest accuracy score were retrieved to be used in the base learners.

c) Creating a voting ensemble

The next step after identifying the base learners was to create a voting ensemble. The voting ensemble was created which combines the base learners through average predicted probability. That is in average predicted probability the final class label is got from the class label with highest average estimate. Weights are explicitly assigned to the classifiers to weigh the occurrences of class estimates before averaging. This takes into consideration the confidence of each voter.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

Each class probabilities is presented by the classifier and the ensemble calculates the average of these and chooses the appropriate summed up weighted probabilities.

IV. RESULTS AND DISCUSSION

Implementation of OVE leads to the findings that would be discussed in this section. The classifiers were combined to create a voting ensemble with soft voting scheme. Weights were assigned to the voting scheme depending on their classification accuracy. Initially a 10 fold cross validation was implemented to identify the classification accuracy of each classifier and the voting ensemble created before hyperparameter tuning. Then the hyperparameters of the base learners were tuned using grid search. The classification accuracy of the base learners as well the ensemble produced after tuning clearly shows an increase by minimum 3% to a maximum of 23% as compared to classification before tuning. Similarly the prediction accuracy of individual base learner has increased. There is a prediction accuracy increase of 2% for DT, 4% for MLP, 1% SGD and 10% for the propose OVE. Table. 1 summarizes the classification and prediction accuracy of each base learner and the ensemble created by combining them.

Classifiers	**Classification Accuracy**		**Prediction Accuracy**	
	Before Tuning	**After Tuning**	**Before Tuning**	**After Tuning**
DT	0.67 (+/- 0.05)	0.70 (+/- 0.04)	0.75	0.77
MLP	0.58 (+/- 0.16)	0.70 (+/- 0.05)	0.76	0.80
SGD	0.47 (+/- 0.09)	0.70 (+/- 0.04)	0.77	0.78
Ensemble	0.61 (+/- 0.13)	0.72 (+/- 0.06)	**0.73**	**0.83**

Table: 1 Classification and Prediction accuracy of classifiers in OVE

Optimization of hyperparameter of the base learners has considerably increased the performance of the prposed voting ensemble, OVE. Since the dataset used has a multi class target value a much more accurate way of identifying the performance of classifiers would be to analyse their precision, recall and Fscore. The micro and macro averaging of precision, recall and fscore is calculated for a multi-class imbalanced problem as given in Table 2. Macro average (Macro-avg) would represent the model best in a class imbalance dataset.

Prediction accuracy	**Precision**		**Recall**		**Fscore**	
	Macro-avg	**Micro-avg**	**Macro - avg**	**Micro-avg**	**Macro-avg**	**Micro-avg**
0.8333	0.8333	0.8333	0.8417	0.8333	0.8356	0.8333

Table:2 Performance Metrics of OVE

Macro-averaging (Micro-avg) better represents a multi-class imbalanced problem since average of all classes would be computed taking into account individual class average. The confusion matrix in fig.2 shows the classification outcome of OVE. There are three target values 0, 1 and 2 that represent the class H, L, M which is High, Low and Medium. The diagonal elements state the correctly labelled instances. The off-diagonal elements are the misclassified elements. The higher values in the diagonal matrix show the number of correctly predicted instances by OVE.

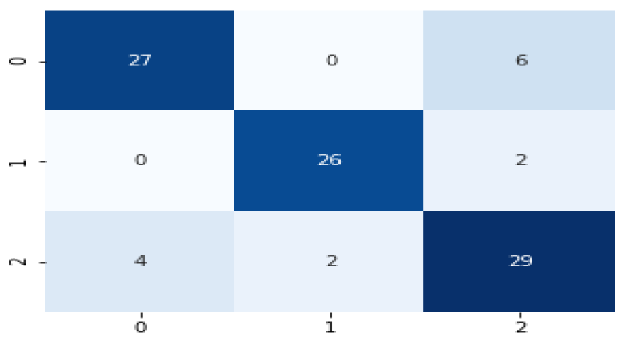

Figure 2: Confusion Matrix_OVE

V. CONCLUSION AND FUTURE ENHANCEMENTS

The OVE approach proposed in this article has shown to increase the accuracy of prediction of the multi-class labels.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

Soft voting has been used since it produces accurate results as compared to majority voting. The results clearly show that classification and prediction accuracy of the ensemble can be clearly enhanced by hyperparameter tuning in addition to the enhancing capability of ensemble of classifiers themselves. In future various high level hyperparameter optimization techniques like evolution algorithms, Bayesian optimization and population based optimization can be implemented to make the optimization of hyperparameter even more robust and accurate. In addition to this other ensemble development approaches like stacked generalization and blending can be explored.

REFERENCES

1. Princy Christy. A , N. Rama, " Enhancing Prediction Capability of Classifiers by Hyperparameter Optimization", Asian Journal of Computer Science and Technology, Vol.8 No.2, 2019, pp. 127-129.
2. Aytuğ Onan, Serdar Korukoğlu, Hasan Bulut, "A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification", Expert Systems with Applications, Volume 62, 15 November 2016, Pages 1-16.
3. Bashir, S., Qamar, U. & Khan, F.H. Qual Quant, "Heterogeneous classifiers fusion for dynamic breast cancer diagnosis using weighted vote based ensemble", Quality & QuantityVolume 49, Issue 5, pp 2061–2076, September 2015, https://doi.org/10.1007/s11135-014-0090-z.
4. Ade, R., & Deshmukh, P. R, "Classification of students by using an incremental ensemble of classifiers:, Proceedings of 3rd International Conference on Reliability, Infocom Technologies and Optimization. 2014 doi:10.1109/icrito.2014.7014666.
5. Ashraf M., Zaman M., Ahmed M, "To Ameliorate Classification Accuracy Using Ensemble Vote Approach and Base Classifier", Advances in Intelligent Systems and Computing, vol 813. Springer, 02 September 2018.
6. P. Kumari, P. K. Jain and R. Pamula, "An efficient use of ensemble methods to predict students academic performance," 2018 4th International Conference on Recent Advances in Information Technology (RAIT), 2018, pp. 1-6.
7. Levesque, Julien-Charles, Christian Gagné and Robert Sabourin. "Bayesian Hyperparameter Optimization for Ensemble Learning", CoRR abs/1605.06394 ,2016.
8. Stapel, Martin; Zheng, Zhilin; Pinkwart, Niels, "An Ensemble Method to Predict Student Performance in an Online Math Learning Environment", Proceedings of the 9th International Conference on Educational Data Mining, pp. 231-238, Jun 29-Jul 2, 2016.
9. Amrieh, E. A., Hamtini, T., & Aljarah, I, "Mining educational data to predict Student's academic performance using ensemble methods", International Journal of Database Theory and Application, 9(8), 119–136, 2016.
10. Satyanarayana, Ashwin, and Mariusz Nuckowski. "Data Mining using Ensemble Classifiers for Improved Prediction of Student Academic Performance." In ASEE
11. Mid-Atlantic Section Spring 2016 Conference, George Washington University, Washington D.C, April 8-9, 2016.
12. S.Kotsiantis, K.Patriarcheas, M.Xenos, "A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education", Knowledge-Based Systems, Volume 23, Issue 6, Pages 529-535, August 2010.
13. Mrinal Pandey, S. Taruna, A Comparative Study of Ensemble Methods for Students' Performance Modeling, International Journal of Computer Applications, Volume 103, No.8, pp. 26 – 32, October 2014.
14. Hyunjoong Kim, Hyeuk,Kim, Hojin Moon, Hongshik Ahn, "A weight-adjusted voting algorithm for ensembles of classifiers", Journal of the Korean Statistical Society, Volume 40, Issue 4, December 2011, Pages 437-449