International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Amalgamating Ensemble Methods for Image Classification

P. Siva Prasad[1], *A. Senthilrajan*[2]

[1]Research Scholar, Dept. of Computer Science, Alagappa University, Tamil Nadu, E-mail id: prasadsiva_17@yahoo.com

[2]Professor and Director, Dept. of Computational Logistics, Alagappa University, Tamil Nadu, E-mail id: agni_senthil@yahoo.com

ABSTRACT:

Bagging and Boosting are most popular ensemble methods. Boosting method is stronger than bagging method on noise data. Bagging is more robust than boosting in noisy settings. For this reason, we have proposed and addressed an ensemble using a voting methodology of bagging and boosting ensembles with sub-classifiers in each one. We have performed a comparison with simple bagging and boosting ensembles with sub-classifiers, as well as other well-known combining methods, on standard datasets and the proposed technique was the most accurate.

Keywords: classification, bagging, boosting, machine learning, deep learning

Introduction:

There are differences between bagging and boosting. Boosting changes adaptively the distribution of the training dataset based on the performance of previously created classifiers while bagging changes the distribution of the training dataset. Boosting method use a function of the performance of a classifier as a weight for voting. Bagging uses equal weight voting. Boosting algorithms are considered stronger than bagging on noise-free data. However, bagging is much more robust than boosting in noisy settings. For this reason, in this work, we built an ensemble combing bagging and boosting version of the same learning algorithm using the sum voting methodology. We have performed a comparison with simple bagging and boosting ensembles as well as other known ensembles on standard benchmark datasets and the proposed technique had the best accuracy in most cases. For the experiments, representative algorithms of well-known machine learning techniques, such as decision trees, rule learners and Bayesian classifiers were used. In this paper, we have presented the most well-known methods for building ensembles that are based on a single learning algorithm and also discussed the proposed ensemble method. Experiment results using number data sets and comparisons of the presented combining method, using different base classifiers, with other ensembles are presented. Finally we have concluded with summary.

Ensemble Methods

Ensemble methods combine different decision trees to deliver better predictive results, afterward utilizing a single decision tree. The primary principle behind the ensemble model is that a group of weak learners come together to form an active learner. Bagging and boosting are most popular ensemble methods. Boosting method is stronger than bagging method on noise data. Bagging is more robust than boosting in noisy settings. Bagging is used when our objective is to reduce the variance of a decision tree. Here the concept is to create a few subsets of data from the training sample, which is chosen randomly with replacement. Now each collection of subset data is used to prepare their decision trees thus, we end up with an ensemble of various models. Boosting is another ensemble procedure to make a collection of predictors. In other words, we fit consecutive trees, usually random samples, and at each step, the objective is to solve net error from the prior trees. If a given input is misclassified by theory, then its weight is increased so that the upcoming hypothesis is more likely to classify it correctly by consolidating the entire set at last converts weak learners into better performing models. Bagging is a parallel ensemble, while boosting is sequential.

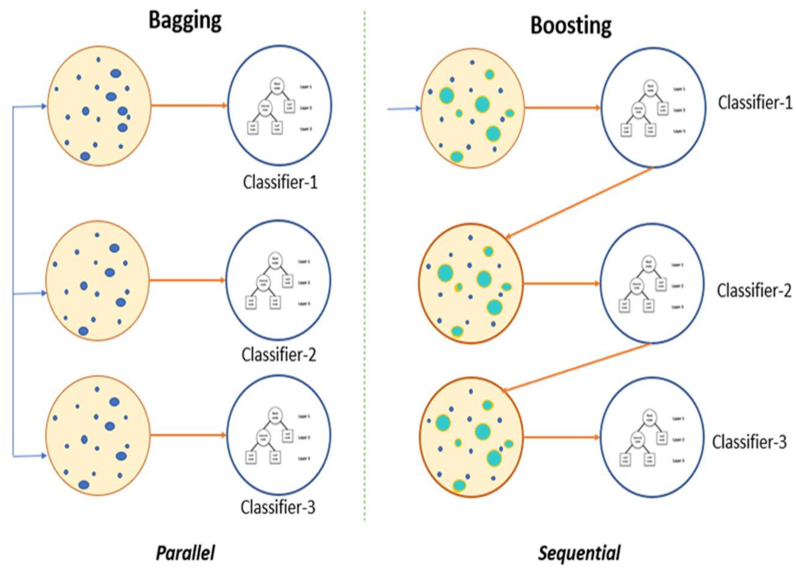

Fig.1 Comparison of Bagging and Boosting

Proposed Model

In this work, we have amalgamated ensemble methods namely bagging and boosting methods with sum rule voting. When the sum rule is used each sub-ensemble has to give a confidence value for each candidate. In our algorithm, voters express the degree of their preference using as confidence score the probabilities of subensemble prediction. Next all confidence values are added for each candidate and the candidate with the highest sum wins the election. The proposed ensemble is schematically presented in Fig. 1, where hi is the produced hypothesis of each sub-ensemble, x the instance for classification and y* the final prediction of the proposed ensemble.

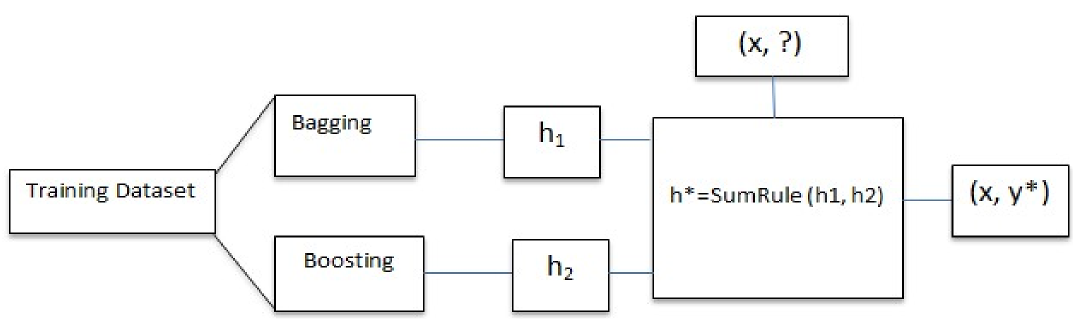

Fig.2 Proposed Architecture

Performance of Proposed Model

In this work, we have built an ensemble using a voting methodology of bagging and boosting ensembles. It was proved after a number of comparisons with other ensembles, that the proposed methodology gives better accuracy in most cases. The proposed ensemble can achieve an increase in classification accuracy compared to the tested base classifiers. The proposed ensemble achieved lower error than either boosting, bagging, multi-boost and decorate combining methods when applied to a base learning algorithm and learning tasks for which there is sufficient scope for both bias and variance reduction. The proposed ensemble can achieve a reduction in error rate about 9% compared to previous models. So, the performance of the presented ensemble is more accurate than the other well-known ensembles.

CONCLUSION:

Boosting method is considered stronger than bagging on noise-free data; however, bagging is much more robust than boosting in noisy settings. In this work, we have built an ensemble using a voting methodology of bagging and boosting ensembles. It was proved after a number of comparisons with other ensembles, that the proposed methodology gives better accuracy in most cases. The proposed ensemble has been demonstrated to (in general) achieve lower error than either boosting or bagging when applied to a base learning algorithm and learning tasks for which there is sufficient scope for both bias and variance reduction. The proposed ensemble can achieve an increase in classification accuracy of the order of 9% to 16% compared to the tested base classifiers. Our approach answers to some extent such questions as generating uncorrelated classifiers and control the number of classifiers needed to improve accuracy in the ensemble of classifiers.

REFERENCES:

1. Bauer, E. & Kohavi, R., An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. Machine Learning, Vol. 36, 1999, pp. 105–139
2. P. Siva Prasad, Dr. A. Senthilrajan, "Amalgamate Classification of Plant using Deep Learning," International Journal of Scientific & Technology Research, vol. 9, Issue 02, February 2020.
3. Webb G. I., MultiBoosting: A Technique for Combining Boosting and Wagging, Machine Learning, Vol. 40, 2000, pp. 159–196.
4. P. Siva Prasad, Dr. A. Senthilrajan, "An Incremental Learning Method for Classification of Plant Leaves using Deep Learning," Advances and Applications in Mathematical Sciences, vol 20, Issue 11, September 2021, Pages 2607-2611.
5. Mr. P. Siva Prasad, Dr. A. Senthilrajan,"A Novel CNN-KNN based Hybrid Method for Plant Classification," Journal of Algebraic Statistics, vol. 13 No. 1, pp. 498-502, 2022
6. Schapire, R. E., Freund, Y., Bartlett, P., & Lee, W. S., Boosting the margin: A new explanation for the effectiveness of voting methods. The Annals of Statistics, Vol. 26, 1998, pp. 1651–1686.
7. Mr. P. Siva Prasad, Dr. A. Senthilrajan," An Adaptive Boosting Approach for Plant Classification through Deep CNN," International Journal of Research Publication and Reviews Vol (2) Issue (3) (2021) Page 1-4
8. Mr. P. Siva Prasad, Dr. A. Senthilrajan," An Ensemble Deep Learning Technique for Plant Identification," International Journal of Computer Sciences and Engineering, Vol.8, Issue.4, Apr 2020
9. Freund Y. , Robert E. Schapire. Experiments with a New Boosting Algorithm, Proceedings of ICML'96, pp. 148-156.
10. Marco A Wiering , Hado Van Hasselt. "Ensemble algorithms in reinforcement learning," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 38(4):930–936, 2008
11. Mr. P. Siva Prasad, Dr. A. Senthilrajan, "Leaf Features Extraction for Plant Classification using CNN," International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), Volume 2, Issue 2, February 2021
12. Alexander Hans, Steffen Udluft," Ensembles of neural networks for robust reinforcement learning. In 2010 Ninth International Conference on Machine Learning and Applications, pages 401–406. IEEE, 2010.
13. Kristin P Bennett, Ayhan Demiriz, and Richard Maclin," Exploiting unlabeled data in ensemble method,". In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 289–296, 2002.
14. Yue Cao, Thomas Andrew Geddes, Jean Yee Hwa Yang, and Pengyi Yang,"Ensemble deep learning in bioinformatics,"Nature Machine Intelligence, 2(9):500–508, 2020.