Paper Review : BOOSTING THE MARGIN: A NEW EXPLANATION FOR THE EFFECTIVENESS OF VOTING METHODS

Eddie Ramirez, Minhaj Fahad, and Oluwasola Ogundare

Operations Research Department, Cornell University

March 18, 2024

Abstract

One of the surprising recurring phenomena observed in experiments with boosting is that the test error of the generated classifier usually does not increase as its size becomes very large, and often is observed to decrease even after the training error reaches zero. In "Boosting the Margin : A New Explanation for the Effectiveness of Voting Method", the authors relate this phenomenon to the distribution of *margins* of the training examples with respect to the generated voting classification rule, where the margin of an example is simply the difference between the number of correct votes and the maximum number of votes received by any incorrect label. They show that techniques used in the analysis of Vapnik's support vector classifiers and of neural networks with small weights can be applied to voting methods to relate the margin distribution to the test error. They also show theoretically and experimentally that boosting is especially effective at increasing the margins of the training examples. Finally, the authors compare their explanation to those based on the bias-variance decomposition.

1 Summary

The research paper "Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods" by Robert E. Schapire, Yoav Freund, Peter Bartlett, and Wee Sun Lee, presents a technical exploration of why boosting algorithms continue to perform well even as the complexity of the generated classifier increases significantly. Their investigation reveals that the phenomenon can be explained by the distribution of margins, the differences between the number of correct votes and the maximum number of votes for any incorrect label across training examples. The paper establishes "rigorous" bounds on the generalization error that relate to the margin distribution, independent of the number of base classifiers used. This provides a novel understanding that has challenged traditional beliefs about classifier complexity and generalization error, suggesting that the key to the success of boosting methods lies in their ability to manipulate the margin distribution rather than merely reducing bias or variance without overfitting.

2 Main Results

In the first experiment, when bagging with C4.5, Breiman's bagging method was applied on top of C4.5, which involved rerunning the C4.5 algorithm multiple times on different bootstrap sub samples of the training data and then combining the resulting trees through a voting mechanism. The results showed a large improvement in prediction accuracy where the test error of a single run of C4.5 on the dataset was 13.8%. After combining 1000 trees using bagging, the test error was reduced to 6.6%.

In the second experiment, when boosting with C4.5, Freund and Schapire's AdaBoost algorithm was used along with C4.5 as the base learning algorithm. Similar to bagging, AdaBoost runs C4.5 multiple times but selects training subsamples differently. AdaBoost focuses on the examples that are the most difficult to classify correctly by adjusting the distribution of examples based on the performance of classifiers from before. After combining a small number of trees, the training error drops to zero, but the test error continues to decrease as more trees are added, going from 8.4% to 3.1% after 1000 rounds. This shows that AdaBoost can continue to improve the generalization of the model even after gaining perfect training accuracy and show more improvement over the single classifier than bagging. The experiments demonstrate that combining up to 1000 trees leads to classifiers with more than two million decision tree nodes and does not result in increased test error. Usually, complex classifiers are expected to overfit the training data and perform worse on unseen data and yet boosting and bagging maintained low test errors even with complex combined classifiers.

The next section revolves around two theorems, the first being a finite base-classifier space and the second theorem being the infinite base-classifier space. With the generalization error bounds of the theorems, it is possible to construct majority vote classifiers with bounded errors. On top of this, a larger margin implies a more confident classification since the analysis shows that increasing the margins across the training set can lead to a decrease in the upper bound of the generalization error. However, the paper acknowledges that the bounds are loose and not directly applicable to making quantitative predictions in practice. The authors test these bounds in the Waveform dataset, reducing the error from 44.7% to 19.6%, showing a significant decrease in variance from 5.6% to 2.8%, the Twonorm dataset, decreasing error from 33.3% to 5.3%, with a variance reduction from 28.5% to 2.3% and in the Threenorm dataset, lowering the error from 41.9% to 22.0%.

3 Insights

Given the paper and the findings presented above, one is able to draw to the conclusion that the classification becomes more accurate as the margin increases. The margins serve as a manner to distinguish and evaluate the confidence of an ensemble classifier when making predictions. A study is done to analyze the relation between the distribution of the margins throughout a dataset and the frequency of the error of generalization within the classifier. A close observation is maintained throughout the study to better understand how the algorithms work to optimize the distribution of the margins and under what constraints, drawing parallel to how the classifier improves and handles data that is not foreseen in the training data, but introduced in the testing data. This is done by predicting the error given what is observed in the dataset based upon its marginal distribution. An approach to construct this would be applying these boosting algorithms to datasets and taking note of the interconnection between the distribution of the margins and recurrence of error to better understand how the theory relates to practice. To conclude, this project pivots towards exploring the relationship of optimizing the margins via the use of boosting algorithms and how it relates to reducing the error of generalization, how the use of ensemble learning methods is supported when used for creating a classification model for different problems and datasets.

References

[1] Schapire, Robert E., et al. "Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods." The Annals of Statistics, vol. 26, no. 5, 1998, pp. 1651–86. JSTOR, http://www.jstor.org/stable/120016. Accessed 10 Mar. 2024.