Stacked Generalization for More Accurate Prediction of Patient Survival

Shotaro Ishihara

Asami Matsumoto

Abstract

This paper presents how stacked generalization works for getting more accurate prediction of hospital patient survival. Stacked generalization is a common technique in machine learning competitions, which utilizes multiple predictions to get better performance. It was the core of our solution in WiDS Datathon 2020, ranked 14th. Case study reveals that model diversity can contribute to superior results.

1 Competition overview and findings

The challenge is to create a model that uses data from the first 24 hours of intensive care to predict patient survival (WiDS, 2020). In machine learning competitions, it is common to verify their performance using techniques such as cross validation (Kohavi, 1995) to prevent over-fitting. The scores given by cross validation and the leaderboard seemed to be correlated strongly in this competition. This observation leads us to utilize the technique called stacked generalization (David, 1992).

2 Model pipeline and performance

2.1 Model pipeline

Figure 1 shows the model pipeline of one of the final submissions: (1) Select 7 subsets from 73 models (2) Stacked generalization by ridge classifier for each subset (3) Stacked generalization by ridge classifier. It is empirically known that the diversity of models is important for raising cross validation score with stacked generalization, and a wide variety of models has been created.

2.2 Performance

Fugure 2 summurizes the performance of indivisual predictions. Models with poor performance by themselves also contributed to the performance improvement. For example, removing the predictions of Decision Tree and

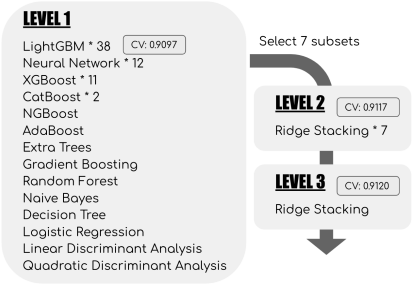

Figure 1: The overview of model pipeline of one of the final submissions.

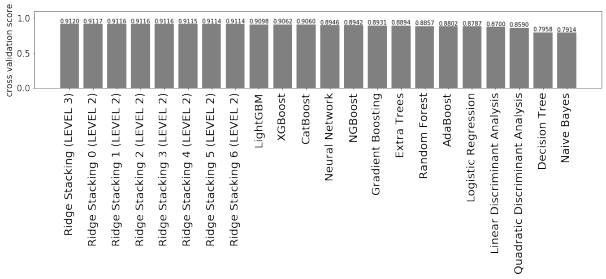

Figure 2: The performance of individual predictions. The best scores are used as the representative for LEVEL 1 predictions.

Naive Bayes droped the score. The same trends can be seen for the score of public and private leaderboard.

3 Conclusions

The presented case study demonstrates the utility of stacked generalization. Model diversity is important and drives our final results.

References

WiDS Datathon 2020 | Kaggle. 2020. available at: https://www.kaggle.com/c/widsdatathon2020/ (accessed 7 March 2020).

Kohavi. Ron: A study of cross-validation and bootstrap for accuracy estimation and model selection 1995. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, 2(12): 1137–1143.

David H. Wolpert: Stacked Generalization 1992. Neural Networks, 5: 241-259.