Summary 14

Shaun Pritchard

Florida Atlantic University

CAP 6778

November 2, 0-2021

M. Khoshgoftaar

Output Thresholding for Ensemble Learners and Imbalanced Big Data

The aim of this study is to use output thresholding strategies to improve the binary classification of imbalanced and highly imbalanced big data sets. The study compares four thresholding strategies using four ensemble learners and two Medicare fraud classification data sets (part B and part D), both of which contain millions of negative class samples with positive class sizes as small as 0.8%.

This study compares four thresholding strategies four classification performance were Lambda is at a default of 0.5, Prior(P) was the prior probability of a positive class, Geometric Mean (G-Mean) were the optimized training data by maximizing G mean, and F-measure were the training data is optimized to maximizing the F-measure.

The study implemented four tree-based ensembles with six rounds of 5-fold cross validation to learners which included Random Forest (RF), Extreme Gradient Boosting (XGB), Light Gradient Boosting Machine (LightGBM), and CatBoost learners. This was followed by averaging the results across the learners and data sets and using Tukey's Honestly Significant Difference (HSD were $\alpha = 0.01$) test to identify meaningful differences between each thresholding strategy.

In the part B dataset, an adjustment from the prior threshold to the G-Mean using XGB showed the best results for TNR of 0.7978 and TPR of 0.7692, showing that a small change in the threshold can have a significant impact on the classification. All three non-default thresholds outperform the default threshold of 0.5 in the part D dataset. Based on G-Mean and F-Measure results, all three non-default thresholds perform statistically the same. Ultimately the study showed that the default threshold performed consistently worse. The prior threshold showed more consistency in the scores for TPR. The G-mean and F-measure prove a better balance of the TPR and TNR. Finally, small changes to the threshold (<0.01) yield significant changes to classification results overall.