Summary 11

Shaun Pritchard

Florida Atlantic University

CAP 6778

October -26-2021

M. Khoshgoftaar

Threshold Based Optimization of Performance Metrics with Severely Imbalanced Big Security Data

The purpose of this study was to provide new results to see if other performance metrics besides the de-facto Area Under the Receiver the Operating Characteristic Curve (AUC) metric could do a better job of measuring the performance of highly imbalanced datasets. The researchers compared the performance of eight machine learning algorithms on imbalanced data. Using the OWASP Switch-blade 4.0, HTTP denial-of-service attacks were collected from the school's student resource server and analyzed. In this study, six different metrics were compared to determine which provided a better representation of a model's predictive performance and evaluate the impact of adjusting the classification threshold on the metrics.

In this study, the AUC is evaluated extensively in regard to imbalanced large data sets and the reasons it is misleading. As an example, the AUC metric can be used as a discrimination index to represent the chances that a majority instance will have a higher predicted value than a minority instance, regardless of the goodness-of-fit test. It can potentially lead to a poorly fitted model with discriminatory power.

The six evaluation metrics used in this study were AUC, AUPRC, TPR, TNR, F-Measure, and G-mean performance metrics using the C4.5N(J48 in Weka) learner. To determine significance, they calculate the metrics based on the best classification threshold values, compared to the default threshold of 0.5.

According to the researchers, even though the AUC measures all selected learners at high levels (average above 0.94), the other resulting metrics are rather variable. This is the primary reason why they questioned the use of AUC as a primary metric. In addition, they

observed that many metrics that performed poorly at default thresholds significantly improved when using new thresholds. In this way, it is beneficial to identify the classification thresholds for a preferred metric. They concluded that the C4.5N (J48 Weka) is the ideal learner for detecting Slow POST attacks within big datasets with severe imbalances.