

Summary 9-1, 9-2

Shaun Pritchard

Florida Atlantic University

CAP 6778

October -07-2021

M. Khoshgoftaar

Summary 9-1 - An Empirical Study of the Classification Performance of Learners on Imbalanced and Noisy Software Quality Data

The researchers compared the performance of seven sampling techniques compared to those without sampling techniques by using 11 classification algorithms on 12 measurement datasets with varying levels of noise and imbalance on the CCCS(military command, control, and communications system) dataset.

The purpose of the study was to determine whether the 11 learners could classify the unbalanced noisy data used in Software Quality assurance testing. According to the results of this study, specific sampling techniques are more appropriate when using noisy data.

Sampling techniques were implemented with the 11 learners were random undersampling (RUS), random oversampling (ROS), one-sided selection (OSS), cluster-based oversampling (CBOS), Wilson's editing (WE), SMOTE (SM), and borderline-SMOTE (BSM). The 11 learners used are as follows: Two K nearest neighbors (kNN) learners, Naive Bayes (NB), Logistic regression (LR), (Jrip)RIPPER (Repeated Incremental Pruning to Produce Error Reduction), random forests (RF), two different versions of C4.5 (J48), Multilayer perceptron(MLP), Radial basis function networks (RBF), and the support vector machine (SVM) classifier.

According to the study, software noise affected data sampling techniques such as RUS, WE, and BSM which outperformed the other four techniques. Results showed that the benefit of these techniques applied to noisy datasets led to OSS holding up well, ROS and SM showing to be worse, and CBOS showing to be the worst. WE performed better with high imbalance and noise, and BSM performed better with no noise and low imbalance. Data sampling techniques were shown to be affected by the combination of imbalance level and noise level in the study.

Also, when imbalance was high, the study found that some sampling techniques were more affected by noise.

Summary 9-1 - A Comparative Study of Filter-based Feature Ranking Techniques

This experiment was designed to identify the best classification model performance for software quality assurance. At the time, this was a groundbreaking study that had never been done before. Five different types of classification models were built using six filter-based feature ranking techniques evaluated in this study.

Using subsets of selected attributes from the datasets the researchers implemented five classification models with use of 8 performance metrics real under the Receiver Operating Characteristic (ROC) curve (AUC), the area under the Precision-Recall curve (PRC), Default F-Measure (DFM), Best F-Measure (BFM), Default Geometric Mean (DGM), Best Geometric Mean (BGM), Default Arithmetic Mean (DAM), and Best Arithmetic Mean (BAM)

The classifiers used were Naive Bayes (NB), multilayer perceptron (MLP), k-nearest neighbors (KNN), support vector machine (SVM), and logistic regression (LR). six different filter-based feature rankers were used with the classifiers chi-square (CS), information gain (IG), gain ratio (GR), symmetrical uncertainty (SU), and two forms of ReliefF (RF and RFW). In order to evaluate the effectiveness of these methods.

The experimental results highlight the importance of metrics selection and how performance metrics are crucial for assessing classification performance. In the experiments, the results found that the SVM classification method was the worst performer in reference performance metrics. According to the results, when models are built using the SP1 data set and NB classifier, RFW performed best in terms of AUC, SU performed best in terms of PRC, and

GR performed best on the performance metric DGM. This study is very interesting to me, and it offers some guidance in the selection of performance metrics for classifiers I will be implementing going forward.