Abstract

Dataset with imbalance class distribution or known as class imbalance problem can be found in real-world problems. In so many cases, minority class in dataset is more potential compared to majority class. Common classification technique cannot handle this problem because it considered that each class shares the same importancy. Because of that, a method is needed to classify the minority class well.

PNrule algorithm is a rule based algorithm designed to handle multi-class imbalance problem. Multi-class problem handled by constructing binary classifier for each class in dataset with the class target is the class which is being learned. This algorithm consists of P-stage, N-stage, and scoring mechanism. Each stage has their own role in making the final rule for each class.

There are five datasets used in this final project and retrieved from UCI Machine Learning Repository. Each dataset is multiclass with different class distribution. The performance of this algorithm is measured by recall, precision, and F_measure. The result of testing and analysis show that PNrule algorithm is able to classify the multi-class imbalance problem very well with adjusting threshold for each parameter.

Keyword : class imbalance, multi-class, PNrule algorithm