Abstract

Churn prediction represents one type of data mining's task, called classification, that aim to predict potential churn customer at cellular telecommunication industry. Problem in churn prediction is imbalance class, where the distribution of training data isn't balance, sum of one class is much greater than another class. Classifier tends to assume that data is in balance condition, so it will bias the prediction for minority class belong to prediction for majority class, and possibly will judge minority class only as outlier (for minority class in too litte amount).

Churn prediction will be solved by implementing K-Nearest Neighbor for imbalance class method which modify K-Nearest Neighbor in order to more consider to minority class. At learning process, K-Nearest Neighbor will build dynamic cluster according to its class distribution and choose its K-Nearest Neighbor priority from the same nearest cluster. Wish this modification able to increase minority class prediction accuracy without decreasing majority class prediction accuracy. Classification result accuracy from K-Nearest Neighbor for imbalance class will be compared with the classfication result from lazy learner classifier Ibk on Weka 3.5.6 tool and another populer classifier on Clementine 10.1 tool in top decile lift, lift curve also gini coefficient forms.

Key words : churn prediction, imbalance class, K-Nearest Neighbor