



Abstract

The interest for data mining models (DMM) has increased tremendously the past decades, due to its potential for uncovering valuable information hidden in massive data sets.

There exist several categories of data mining tasks, such as e.g. clustering, regression, association analysis, etc., but this talk will focus on classification techniques. Two topics will be discussed: (1) the need for business driven performance metrics to measure the effectiveness of customer churn prediction models; and (2) the incorporation of (social) network effects into traditional classification models.

Telecommunication companies rely more than ever on data mining techniques for the prediction of customer churn. A range of traditional classification models is available (e.g. logistic regression, Bayesian network classifiers, support vector machines, etc.), from which the best performing model needs to be selected. Traditionally, this selection is based on statistical measures such as e.g. the area under the receiver operating characteristic curve (AUC). However, these measures do not take into account the business goal of the classification task, i.e. the generation of extra profit. A novel performance measure will be discussed, which purpose is to select the technique which generates the highest incremental profit, instead of simply maximizing a statistical property of the classifier.

The second topic concerns the increasing availability of (social) network data. Traditional data mining techniques generally do not incorporate network effects. The potential of network based classification techniques will be illustrated by a case study about customer churn prediction based on call graphs. The results suggest that traditional DMM and networked learners are complementary in the sense that they identify different individuals as potential churners. This leads to opportunities to enhance the predictive performance of customer churn prediction models.