Churn Prediction in Telecommunications Using MiningMart

Timm Euler

Computer Science VIII, University of Dortmund
D-44221 Dortmund, Germany
euler@ls8.cs.uni-dortmund.de,
http://www-ai.cs.uni-dortmund.de

Abstract. This paper summarises a successful application of Knowledge Discovery in Databases (KDD) in an Italian telecommunications research lab. The aim of the application was to predict customer churn behaviour. A critical success factor for this application was clever preprocessing of the given data, in particular the construction of derived predictor features. The application was realised in the MiningMart KDD system, whose particular strength is data preprocessing on a conceptual level. Since MiningMart provides a declarative, yet executable model of the presented application, this model could be published in a central repository of KDD models, where it is publicly inspectable, which complements the descriptions in this paper.

1 A case study in churn prediction

A major concern in customer relationship management in telecommunications companies is the ease with which customers can move to a competitor, a process called "churning". Churning is a costly process for the company, as it is much cheaper to retain a customer than to acquire a new one [1]. The objectives of the application to be presented here were to find out which types of customers of a telecommunications company are likely to churn, and when. The task was solved using decision trees which achieved a predictive accuracy of 82%. This good result was only possible due to the introduction of relevant derived features for prediction which were not available in the original data, and due to a re-representation of the data so that temporal aspects could be included. Thus data preprocessing was a key success factor in this application.

During preprocessing, the available data tables were transformed so that a classification algorithm could be applied. In the resulting data set, each row (that is, each example for classification) corresponded to one customer of the company, and contained many features describing their telecommunication behaviour for *each* of five consecutive months. Whether or not the customer left the company in the sixth month determined the binary classification label or target. Once a learned classifier is available it can be applied every month on data from the current and past four months, to predict churn for the following month.

A different prediction horizon can be easily used, since it requires only a few parameter changes in the declarative model of the application (see section 2).

The main preprocessing operations concerned a table that provided call detail records. Since it includes very detailed data, for every phone call made by a customer, the data had to be aggregated. Aggregation is done with the aim of creating a customer profile that reflects the customer's phoning behaviour over the last five months. The details of this aggregation process are complex, interesting and important, but cannot be described here due to space limitations. In essence, many aggregated attributes containing the lengths of calls made by every customer in these five months were created. On this table a decision tree was trained to predict the binary target. Since the first results were not satisfactory, more derived attributes were added for learning. The decisive additions in terms of resulting predictive accuracy were the attributes that indicated *changes* in the telecommunication behaviour over the five months represented.

2 Publishing the case study using MiningMart

The particular, time-based character of this application requires to repeat it every month to predict a group of customers likely to churn soon. This also involves repeating the preprocessing phase every month. Therefore a particular criterion during the choice of the supporting KDD system was excellent support in preprocessing. Further, for reasons of scalability, an environment to operate directly on a relational database was needed. Therefore, MiningMart [2] was chosen, which employs a declarative metamodel (called M4) of mining processes and the data they use, and provides an SQL compiler for KDD application models expressed in M4. Thus all data processing in this application took place inside the relational database. A unique feature of MiningMart is the opportunity to publish these models in the internet without further editing. To this end, the process model as set up with the MiningMart client is exported to a file based on XML syntax, and sent to the central "Case Base" that is maintained on the MiningMart web site[1]. The reader is invited to inspect the case study presented here in the case base under "Model Case Telecom". From the case base a process model can be downloaded and reused on similar data. Schema matching is used in MiningMart to map the data model from the case base to local data.

References

1. Richeldi, M., Perrucci, A.: Mining Mart Evaluation Report. Deliverable D17.3, IST Project MiningMart, IST-11993 (2002)
2. Morik, K., Scholz, M.: The MiningMart Approach to Knowledge Discovery in Databases. In Zhong, N., Liu, J., eds.: Intelligent Technologies for Information Analysis. Springer (2004)

[1] `http://mmart.cs.uni-dortmund.de`