Expert Systems with Applications 40 (2013) 5635–5647

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

An effective hybrid learning system for telecommunication churn prediction

Ying Huang *, Tahar Kechadi

School of Computer Science and Informatics, University College Dublin, Belfield, Dublin 4, Ireland

ARTICLE INFO

Keywords:
Classification
Telecommunication
Churn prediction
Hybrid model
Clustering
Rule induction

ABSTRACT

Customer churn has emerged as a critical issue for Customer Relationship Management and customer retention in the telecommunications industry, thus churn prediction is necessary and valuable to retain the customers and reduce the losses. Moreover, high predictive accuracy and good interpretability of the results are two key measures of a classification model. More studies have shown that single model-based classification methods may not be good enough to achieve a satisfactory result. To obtain more accurate predictive results, we present a novel hybrid model-based learning system, which integrates the supervised and unsupervised techniques for predicting customer behaviour. The system combines a modified k-means clustering algorithm and a classic rule inductive technique (FOIL).

Three sets of experiments were carried out on telecom datasets. One set of the experiments is for verifying that the weighted k-means clustering can lead to a better data partitioning results; the second set of experiments is for evaluating the classification results, and comparing it to other well-known modelling techniques; the last set of experiment compares the proposed hybrid-model system with several other recently proposed hybrid classification approaches. We also performed a comparative study on a set of benchmarks obtained from the UCI repository. All the results show that the hybrid model-based learning system is very promising and outperform the existing models.

1. Introduction

With recent evolution in the Information and Communication Technology (ICT) sector, numerous new and attractive services have been introduced, and they put huge pressure on traditional services. Customer churn has emerged as one of the major issues in Customer Relationship Management (CRM) in telecommunication services around the world, for both wireless providers and long-distance carriers. For instance, in the U.S., telecom providers of long-distance and international services have been bearing the churn rates from 45% to 70% percent for some years (Mattison, 2001). Under the fierce competitive environment, it becomes very important for the telecom operators to retain their existing customers as acquiring new customers is much more expensive. Consequently, predicting which customers are likely to stop their subscription and switch to competitors (churn) is critical. Predicting the potential churners and successfully retain them, especially the valuable ones, can substantially increase the profitability of a company.

In the telecommunications industry, operators usually capture the transactional data, which reflects the service usage, and some static data such as subscriber's personal information and contract details. Data mining (DM) methods have emerged as a good alternative to study the customer behaviour. We can find various DM techniques, such as decision tree, logistic regression, support vector machine, artificial neural networks, inductive rule learning, etc. They have been applied to predict customer behaviour (Huang, Huang, & Kechadi, 2011; Hwang, Jung, & Suh, 2004; Larivire & Poel, 2005; Wei & Chiu, 2002; Xia & dong Jin, 2008). Most of the existing predictive modelling techniques, applied to customer churn, are based on supervised learning; very few of them have been based on unsupervised learning. In addition, most of the classifiers use single model (i.e., only one data mining technique). Many of the single model-based classifiers can predict potential churners to a large extent. However, either the accuracy is not good enough for some of the techniques or there is a room for improving the prediction accuracy for some others, and a hybrid model is a good alternative for better classification performance. Moreover, usually the entire training data instances are all used to build prediction models. However, it may be more effective to predict a new data instance based on partial training instances that are more similar to the test data than other training instances.

The advantages of the proposed model over the other commonly used modelling techniques in the domain of churn prediction concern the following aspects: Firstly, the prediction model

* Corresponding author. Tel.: +353 873145218.
E-mail address: ying.huang.1@ucdconnect.ie (Y. Huang).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.04.020

of the proposed system is a hybrid-model, which fully integrates supervised and unsupervised learning by employing a clustering method for dividing the training data and a rule induction method for generalising classification rules for each cluster. Secondly, the process for predicting a test data instance relies on the most appropriate sub-classifier, which is produced from the most similar cluster to the test instances rather than the entire training set. We argue that using only a sub-classifier may improve the prediction accuracy. Thirdly, the commonly used k-means algorithm is used for clustering the original training data; however, to enhance the clustering result, we apply a weighting technique, which makes full use of the causality between attributes and the target. The experimental results show that modified k-means enhances not only the clustering results but also the prediction results.

The hybrid-model based prediction system has four main phases, which are as follows:

(1) Data Discretisation: We apply a class-dependent discretization method on all the continuous attributes. The continuous data is transformed to the form of intervals, which is conducive to the rule induction.
(2) Weighted Clustering: We apply a weighted k-means clustering algorithm to divide the training data into a number of clusters. We use Path Analysis to calculate the weights of attributes.
(3) Rule Extraction: We apply a rule learning method (i.e., FOIL) to each cluster, extracted in the previous phase, to induce a set of classification rules, which constitutes a sub-classifier. Thus, each cluster corresponds to a sub-classifier.
(4) Prediction: We predict each test instance by choosing the most suitable sub-classifier, which corresponds to the *closest* cluster to the test data, according to a distance or similarity measure.

The remainder of the paper is organised as follows: Section 2 reviews some related work in the area. Section 3 describes the details of the proposed hybrid learning system. The experimental set-up and results are discussed in Section 4. We conclude and outline some future work in Section 5.

2. Related Work

Large number of machine learning and knowledge discovery techniques have been proposed and applied to the problem of customer retention in the domain of CRM. The techniques can be used in different phases of the data mining process, such as eliminating the noise and outliers, reducing the feature space by selecting most relevant attributes, predicting churn behaviour, etc.

In this section, we briefly introduce ways of building a classification model by reviewing several customer churn prediction models proposed in the literature. Hung, Yen, and Wang (2006) developed two classification methods by using three models, one is based on k-means clustering and decision tree (C5.0) and the other combines the back-propagation neural network (BPN) and decision tree. The models are evaluated using LIFT and hit ratio measures. Huang, Kechadi, and Buckley (2012) proposed a mining process that consists of feature extraction and classification. The authors implemented seven traditional classification modelling techniques (e.g., Decision Tree, Linear Regression, and Multilayer Perceptron Neural Networks, etc.) to build different predictive models, and for some models, they use different data processing methods. They have evaluated the performance of their approach using true positive and false positive measures.

Apart from the techniques described above, in fact, there is a large amount of literature reporting on the application of data mining techniques to study the behaviour of customers in telecoms, (Au, Chan, & Yao, 2003; Coussement & Den Poel, 2008; Huang et al., 2011; Mozer, Wolniewicz, Grimes, Johnson, & Kaushansky, 2000). Ngai, Xiu, and Chau (2009) reviewed more than 80 papers about the application of data mining to Customer Relationship Management, and a lot of them concern the domain of customer churn prediction. Nonetheless, most of the proposed modelling techniques use a single model.

Recently, many researchers have started to study hybrid models to improve the classification effectiveness. Usually, hybrid models combine two or more techniques. For instance, the classification or clustering techniques can be sequentially combined (e.g., Khashei, Hamadani, & Bijari (2012) proposed a hybrid classification model by integrating artificial neural networks and multiple linear regression to yield more general and accurate classification result than single traditional method). Tsai and Lu (2009) suggested that the hybrid techniques can provide better performance than many single model-based approaches in numerous different domains. One can refer to Lee and Lee (2006) for some common types and structures of hybrid model-based classification methods. The authors have built a hybrid model called *SePI* (**S**egmentation by **P**erformance **I**nformation). The framework of *SePI* is based on three models, main model, discrimination model, and support model. Decision tree (C5.0), which found to be one of the best single model for a given data, is chosen as the main model; the discrimination model uses the performance information of the main model on the training dataset; the support model uses the data for which the main model predicted incorrectly, and ANN is employed as the support model. The key idea of *SePI* is that if the test examples cannot be predicted correctly by the main model, then they will be predicted by the support model.

For the customer churn prediction problem, most of the related work focuses on using only one data mining method. Table 1 shows the related implementations that employ hybrid modelling techniques for customer churn prediction. Most of the hybrid model-based applications normally follow a common pattern, which consists of two stages: the first stage deals with pre-processing the input data (e.g., reduce the dataset, detect the outliers, etc.), the second stage is dedicated to the mining of the preprocessed data to extract useful patterns. In our study, we also build a hybrid learning system by integrating two stages. However, the difference is that the first stage is no longer considered as the step for data pre-processing. We re-design the system by firstly segmenting the training data into different groups; we build a set of sub-classifiers by applying rule induction method on each group of the data; the second stage predicts the categories of the test examples, each example (e.g., $test\_i$) is predicted by a sub-classifier that was produced by the closest sub-cluster to $test\_i$.

3. Methodology

Our main motivation is the design of a hybrid learning system for predicting the customer future behaviour. The main idea behind our hybrid learning system is to predict a customer instance according to the training examples that are more similar to it. We assume that customers having similar behaviour patterns (characteristics) are more likely to behave the same in the future. Thus, it might be more accurate if an unlabelled instance is predicted using partial training instance, which has similar characteristics with the tested instance, rather than the whole data. This can be achieved by dividing the training data into clusters, and the test instance is assigned to the closest cluster to it.

In this paper, the main concern is the effectiveness of classification, thus, we do not discuss some data pre-process steps, such as data cleaning, normalisation, and feature selection, as they have