Expert Systems with Applications 38 (2011) 7151–7157

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey

Pınar Kisioglu, Y. Ilker Topcu *

Industrial Engineering Department, Istanbul Technical University, Istanbul, Turkey

ARTICLE INFO

Keywords:
Churn analysis
Bayesian Belief Networks
Telecom

ABSTRACT

In telecommunication industry, for many organizations, it is really important to take place in the market. As competition increases between companies, customer churn becomes a great issue to deal with by the telecommunication providers. For an effective churn management, companies try to retain their existing customers, instead of acquiring new ones. Previous researches focus on predicting the customers with a propensity to churn in telecommunication industry. In this study, a model is constructed by Bayesian Belief Network to identify the behaviors of customers with a propensity to churn. The data used are collected from one of the telecommunication providers in Turkey. First, as only discrete variables are used in Bayesian Belief Networks, CHAID (Chi-squared Automatic Interaction Detector) algorithm is applied to discretize continuous variables. Then, a causal map as a base of Bayesian Belief Network is brought out via the results of correlation analysis, multicollinearity test and experts' opinions. According to the results of Bayesian Belief Network, average minutes of calls, average billing amount, the frequency of calls to people from different providers and tariff type are the most important variables that explain customer churn. At the end of the study, three different scenarios that examine the characteristics of the churners are analyzed and promotions are suggested to reduce the churn rate.

1. Introduction

There is a continuous change in telecommunication industry all over the world. In 1996, US government removes the obstacles in local and long distance calling, cable TV, broadcasting, and wireless services. In Europe, markets are formed by the deregulations in England, Sweden, and Finland. Then, 15 European countries ended the restrictions in telecommunication industry. Chile, Malaysia, and Peru discontinued the telephone monopolies. All of these applications open new markets in those countries and put an end to limited communication (http://www.sequent.com).

As the new markets are developed, competition between companies increases sharply. Since the competition gets hard and telecommunication becomes a selling product, companies encounter to minimize costs, add value to their services, and guarantee differentiation. Now, the customers can choose their service providers, so companies pay attention to customer care in order to keep their situation in the market (http://www.sequent.com).

Under the hard conditions of competition, companies try to focus on the behaviors of customers. According to needs of customers, telecommunication companies decide their service offers, give a shape to their communication network and in addition change their organizational structure (http://www.iec.org/online/tutorials/acrobat/bus_int.pdf).

If a customer ends up his subscription from the existing provider and subscribe to another one, the customer is called a 'churner'. Churn is a major problem for companies with many customers, e.g., credit card providers or insurance companies. In telecommunication industry, the sharp increase of competition makes customer churn a great concern for the providers (Richeldi & Perrucci, 2002).

For wireless telecommunication industry, the monthly predicted churn rate is about 2.2%. This means that a company loses about 27% of its subscribers every year (Wei & Chiu, 2002). On the other hand, the cost of churn is a huge problem for the companies. In European and US markets, the cost of churn is about $4 billion per year, and throughout the world, it costs $10 billion per year (SAS International, 2001). Furthermore, acquiring new customers is much more expensive than retaining the existing ones. For this reason, paying attention on current subscribers is more efficient than acquiring new customers (Richeldi & Perrucci, 2002). The scope of competition changes into reducing the churn rate of the subscribers and keeping them from other competitors (Kim & Yoon, 2004).

With an effective churn management, a company can decide whether its customers have a churn propensity or not (SAS

* Corresponding author.
E-mail addresses: pinarkisioglu@gmail.com (P. Kisioglu), ilker.topcu@itu.edu.tr (Y.I. Topcu).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.12.045

International, 2001). The aim of churn management is to minimize the lost due to churn while maximizing the profits by retaining valuable customers (http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V03-4).

Previous studies about churn prediction in the telecommunication industry mainly applied data mining techniques such as neural networks, decision trees, and cluster analysis to predict churn rate. However, to our knowledge, there has been no research using Bayesian Belief Network to identify the factors that have effects on customers to churn.

Hung, Yen, and Wang (2006) studying customer churn in Taiwan market, showed that using neural network gives better results than decision trees. Another research about Taiwan telecommunication industry is studied by Wei and Chiu (2002). They designed a churn prediction model to estimate churn rate of customers using subscriber contractual information and call pattern changes.

Ferreira, Vellasco, Pacheco, and Barbosa (2004) studying the loss of valuable customers in Brazilian mobile telecommunications industry, analyzed the potential savings and profits as a result of the research. A similar article to Ferreira et al.'s research, studied by Mozer, Wolniewicz, Grimes, Johnson, and Kaushansky (2000), identified possible churners and calculated savings.

Nath and Behara (2002) studied customer churn using Naive–Bayes algorithm with a database of fifty thousand customers in American mobile telecommunication industry. Karahoca and Kara (2006) studying the segmentation of Turkish customers according to their profitability, compared a new technique with other clustering techniques. Jahanzeb and Jabeen (2007) showed the opinions of customers about churn management strategies by a survey and compared two telecommunication companies in Pakistan.

This paper is organized as follows. In Section 2, there is a brief explanation about Bayesian Belief Network. Section 3 exhibits the model constructed to identify the factors that cause customers to churn. Section 4 presents three different scenarios and comment. Finally, in Section 5, the results of the model and further suggestions for future studies are discussed.

2. Bayesian Belief Network

Bayesian Belief Network (BBN) is a graphical model that represents the casual relationships between the variables with their conditional probabilities (Heckerman, 1995). Due to its cause and effect diagram, it is used in many real-world problems. It is applied to large engineering project risk management by Lee, Park, and Shin (2009) and is used for performance prediction of box-office success on Korean movies by Chang and Lee (2009). A model is constructed to be applied in Turkish transportation system (Ülengin et al., 2007), and another is used to analyze the complex structure of inflation of Turkey (Şahin, Ülengin, & Ülengin, 2006). Since BBN is one of the most powerful methods for reasoning under uncertainty, it has different application areas such as diagnosis of schizophrenia (Ouali, Cherif, & Krebs, 2006), innovation performance in R&D collaborations (Kim & Park, 2008), and customer lifecycle slope estimation for customer relationship management (Baesens et al., 2004).

For telecommunication industry, most of the studies have proposed customer churn prediction by using data mining techniques as mentioned in previous section. However, these methodologies have some disadvantages. Heuristic-based approach and analytical methods are inconvenient for complicated problems, and also it is hard to collect pure data for statistical methods. Moreover, a mathematical model is required for simulation, correlated variables are unsuitable for decision trees, and a clear confidential data set is needed for neural networks (Lee et al., 2009).

In this study, a Bayesian Belief Network (BBN) is used because of the following advantages. First of all, having missing data in the data set does not make any problem for BBNs. Samples with incomplete data can be fixed by adding or integrating the probabilities over all possible values of the variable. Second, causal relationships are identified by applying a BBN. Thus, BBN makes it easier to understand the problem domain and estimate the results. Another advantage of BBN is that building a model does not consume time and need much effort. The structure of the network is developed, and then, it is easy to add new variables to the model. Finally, BBN is a combination of both a data set and users' prior knowledge; so, the model, constructed through BBN, is sufficiently dependable (Kim & Park, 2008).

3. Proposed model

In this paper, we specify the relationship between variables and use BBN as a prediction model of customer churn in telecommunication sector. Fig. 1 shows the flow chart of our research. First, raw data are collected from the database of one of the Turkish telecommunication providers. The variables are prepared, and a new data set is built for the study. Since continuous variables cannot be used in BBN networks, each continuous variable is discretized using CHAID algorithm. To show the interactions of variables, we check the independence of each variable by correlation analysis and multicollinearity test. A casual map is constructed as a base of BBN network according to the results of the tests and experts' opinions. Finally, sensitivity analysis is applied, and three different scenarios are developed. In each scenario, the possible savings are calculated with suggested promotions. We use the software Netica (www.norsys.com) to construct BBN network and make the sensitivity analysis. Also, SPSS 13.0 is used for decision trees, correlation analysis, and multicollinearity test.

3.1. Data and variables

All data, used in this study, are collected from a telecommunication company in Turkey with their permission. The data set contains data of 2000 subscribers, including 534 churners, from January 2008 to July 2008. Initially, there are 23 different variables in data set. After data preparation, the number of variables is 9. In the initial data set, the variable billing amount has six-month data such as the other two variables, minutes of usage and frequency of usage. The average of six-month data of billing amount is calculated and is inserted to the data set as a new variable instead of six variables in total. The same procedure is also applied for the other variables, minutes of usage and frequency of usage. In addition, a new variable is added to the data set: trend in billing amount. It is obtained from billing amounts of the customers and shows whether their billing amounts are in upward, downward, or constant trend. The detailed explanation of variables is as follows:

Place of residence: the place of residence, where the subscribers live, is important for churn analysis. Subscribers in the data set live in rural and urban areas. Subscribers living in rural areas are more loyal than subscribers living in urban areas due to the fact of limited number of telecom providers in rural areas.
Age: analysis shows that customers under the age of 35 have a high propensity to churn in Turkey.
Tenure: subscribers with less than 1865 days tenure have a high propensity to churn.
Tariff type: there are three different types of tariff: one for commercial companies, institutions, etc. and two for individual subscribers. We consider individual subscribers rather than others