Applied Soft Computing 24 (2014) 994–1012

Contents lists available at ScienceDirect

Applied Soft Computing

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

Improved churn prediction in telecommunication industry using data mining techniques

A. Keramati[a,*], R. Jafari-Marandi[a], M. Aliannejadi[b], I. Ahmadian[c], M. Mozaffari[a], U. Abbasi[a,d]

[a] School of Industrial and Systems Engineering, University of Tehran, Tehran, Iran
[b] Department of Computer Engineering, Amirkabir University of Technology, Tehran, Iran
[c] Department of Industrial Engineering, K.N. Toosi University of Technology, Tehran, Iran
[d] Design and Manufacturing Systems-Arts et Métiers ParisTech, Paris, France

ARTICLE INFO

Article history:
Received 20 November 2012
Received in revised form 29 July 2014
Accepted 18 August 2014
Available online 29 August 2014

Keywords:
Telecommunication
Churn prediction
ANN
KNN
SVM
Decision tree

ABSTRACT

To survive in today's telecommunication business it is imperative to distinguish customers who are not reluctant to move toward a competitor. Therefore, customer churn prediction has become an essential issue in telecommunication business. In such competitive business a reliable customer predictor will be regarded priceless. This paper has employed data mining classification techniques including Decision Tree, Artificial Neural Networks, K-Nearest Neighbors, and Support Vector Machine so as to compare their performances. Using the data of an Iranian mobile company, not only were these techniques experienced and compared to one another, but also we have drawn a parallel between some different prominent data mining software. Analyzing the techniques' behavior and coming to know their specialties, we proposed a hybrid methodology which made considerable improvements to the value of some of the evaluations metrics. The proposed methodology results showed that above 95% accuracy for Recall and Precision is easily achievable. Apart from that a new methodology for extracting influential features in dataset was introduced and experienced.

1. Motivation and significance

Nowadays business managers have started to appreciate the important role of churn prediction in their way of prosperity. In the literature, it has been repeatedly indicated that customer retention in comparison to absorbing new customers is significantly more achievable and less expensive. In today's competitive business environment, losing a customer should be considered as a real disaster. Loss of a customer can be contemplated in three different aspects. First, losing existing customers is figuratively equivalent to have a critical machine irreparably broken down due to the fact that they are any company's most precious assets. Furthermore, by the same imaginary assumption, losing a customer would mean passing our asset intentionally to our competitor. Finally it is too laborious a task to gain a new customer. To make the matters worse, even if a new customer were absorbed, they even would not be as loyal as the old customers. It may take some time for just a proportion of them to become slightly loyal. Therefore, the prevention strategy is absolutely worthwhile. Customer retention plays a major role in many enterprises, especially matured ones, including telecommunications and finances [1]. Acquiring it requires and rears churn prediction, which is another term and keyword in customer retention. It can be explained as predicting customers' probable tendency to switch to our competitor.

In today's telecommunication business environment, competition is tremendously fierce. The services and customers' options also have become more comparable and more competitive. This is the reason why customer loyalty tends to erode. It costs customers figuratively nothing to switch from a service provider to another. They are customers after all and they have freewill to switch to a better and probably more inexpensive service in a competitive market. We, as company managers, ought to take every necessary step so as to get in their way of leaving. It is imperative to distinguish customers who are not reluctant to move toward another competitor before they actually consider so. Therefore, dealing with the probability of customers' churn has become an inevitable issue in telecommunication industry [2]. The telecommunication service companies are annually facing with loss of valuable customers to competitors. Due to the changes and improvements in

* Corresponding author. Tel.: +98 9122388846.
E-mail addresses: keramati@ut.ac.ir (A. Keramati), ruholla.jafari@ut.ac.ir (R. Jafari-Marandi).

http://dx.doi.org/10.1016/j.asoc.2014.08.041
1568-4946/

telecommunications' services technologies in the last few years, customers' churn has resulted in magnificent losses and it has proven itself as a real issue [3].

2. Introduction

The number of mobile phone users during past years has perceptibly multiplied. Statistical report for the end of 2010 indicated that the number of mobile phone users in Iran has surpassed 70 million, more than 90% of the country population. Therefore, telecommunication market is in the point of being saturated, especially for big cities. Mobile phone penetration rates for certain city have gone above 100%, which means there are more subscriptions than inhabitants. As a result, heat of the competition in today's telecommunication market is distinguishably high. Proposed products and service offerings are becoming more and more similar and replaceable. The fact that customers, in most of the cases, are able to self-centrically prefer a service provider better brings about the eroding of customer loyalty. Then, Iran telecom operators in near future need to start giving a great amount of attention to customer churn prediction and customer retention strategies and should they fail to do so they would not survive. Furthermore, it has been repeatedly shown that taking customer retention strategy can be profitable for a company [4].

Not so long ago data mining techniques have been in use to tackle the challenging customer churn problems in telecommunication service field [3]. Due to the aforementioned heat of the competition in telecommunication market, these data mining techniques are mainly employed to cope with the churn prediction issues which has been receiving an unprecedented attention in the telecommunication industry and research. They are mostly and prevalently applied using the customer log-files or questionnaires so as to come up with some knowledge helping to determine the customers who are likely to churn. Although most of these data mining techniques use essentially different techniques to achieve approximately same result, the knowledge is totally worth spending and beneficial for almost any telecommunication company. The following paragraph is devoted to illustrate the importance of exact churn prediction.

Applying data mining techniques itself can be time-consuming and extravagantly expensive. Data Gathering, Data Cleaning, and Data Preprocessing in Iran rather immature telecommunication companies also can be painstakingly hard. It is important to understand that the knowledge extracted for data mining probably will be employed to make important decisions about the customers who are repeatedly known as the most important asset of any business. Therefore, the more accurate and reliable the extracted knowledge is, the more proper and appropriate decisions can be made. Making appropriate decision about such a vulnerable matter needs tremendous certainty and that is achievable by reliable knowledge extracted by reliable techniques. On the other hand, improper decision based on falsified, bias or flawed knowledge may lead to devastating situations where no one can make restitution.

Contemplating previous researches, they can be categorized by two main aspects. First is the studies which were centrally concerned about churn determinants. They analyzed well-known churn determinates and verified them by using customer behaviors in telecommunication. Some attributes, such as customer satisfaction, switching costs, customer demographics, tendency to change behavior, and service usage, have been prevalent among churn determinants [1]. Second, there are other studies such as this paper whose authors' interest has been to improve the outcome of churn prediction by state of the art computational methods. Having said that, although this paper's main part is to propose a novel methodology for more accurate prediction, a feature extraction method for dimensionality reduction is also presented.

A rather distinguished and important part of this paper is the idea of using intelligent algorithms for prediction of future matters. Intelligent algorithm has proven to be versatile and applicable for different tasks. They have been in use in variety of problems, the subject matter be Internet such as [5,6], ATM management [7], or the most basic statistical problem estimating missing value [8]. There have been many research works which employ neuro-fuzzy inference system to combine the advantages of fuzzy logic and neural network for classification and regression problems. There are many articles as an example for this adaptation: [9–14]. The combination of neural network and fuzzy logic in these studies benefits mostly from the human-like logic behind fuzzy systems and the connections of the neural networks. There are, moreover, approaches in which the use of meta-heuristic algorithms such as Genetic Algorithm (GA) for the parameter estimation of neural network or fuzzy system had been employed: [15,16]. Taking this strategy has many advantages because these types of problems are not closed form and more than too often the researcher opts to use an iterative algorithm to estimate their parameters. Meta-heuristic algorithms converge at a local minimum/maximum which is proven to be often sufficient. All the aforementioned works try to utilize intelligent algorithm to improve their estimates. These estimates in many works have come to natural phenomena. Such approaches have made great advances in different discipline such as production and inventory, or even natural hazards and risk management: [17–22].

However this paper's most important contribution is the churn prediction improvement by hybridizing some well-known algorithm. After experimenting with four mentioned prominent algorithms and coming to know their special features, a hybrid algorithm using all four is introduced. Not only did we prove that the suggested methodology is competitive with the best of the four, but also it has a tuning parameter that can be manipulated to predict the way its users need. If a decision maker needs to exactly know which customers will be churned, or if they want to realize who are the customers that have even slightest tendency to leave, the algorithm can be tuned to return the desired outcomes. Computational achievement in this paper is really strong since it has been able to push recall and precision measures, two of the most important measures for evaluating binary prediction, not to return less than 95% accuracy. Moreover, as it is an essential part of any data mining task we have investigated into the influence of the columns in our data. In real world setting computational limitation can bring about many restrictions. Thus, knowing the most influential columns can save many times an effort. To that end, a novel feature extraction methodology is suggested. The suggested method is experienced and consequently compared to traditional statistical means.

The rest of this paper is organized by 5 other sections. Section 3 is a short review of other research papers concerning telecommunication churn prediction. Section 4 describes in details the four classifiers algorithm used in our experiments. Section 5 is to show our experiences and results, whereas in Section 6 a hybridized methodology is explained and illustrated. Last but not least, Section 7 is the final discussion and conclusion of this paper experiments.

3. Literature review

Accurate and reliable prediction of churn customer is important in the development of appropriate retention strategies. Huang et al. [3] proposed a method based on ordinal regression to predict time of churn and tenure of customer in mobile telecommunication industry. They treated customer tenure as an ordinal outcome variable and take advantage of ordinal regression to form