Knowledge-Based Systems 28 (2012) 97–104

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

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

# Predicting customer churn through interpersonal influence

Xiaohang Zhang [a,*], Ji Zhu [b], Shuhua Xu [c], Yan Wan [d]

[a] School of Economics and Management, Beijing University of Posts and Telecommunications, 1318 Main Building, 10 Xitucheng Road, Haidian District, Beijing 100876, China
[b] Department of Statistics, University of Michigan, 439 West Hall, 1085 South University Ann Arbor, MI 48109-1107, USA
[c] School of Economics and Management, Beijing University of Posts and Telecommunications, 1319 Main Building, 10 Xitucheng Road, Haidian District, Beijing 100876, China
[d] School of Economics and Management, Beijing University of Posts and Telecommunications, 506 Mingguang Building, 10 Xitucheng Road, Haidian District, Beijing 100876, China

ARTICLE INFO

*Article history:*
Received 24 July 2011
Received in revised form 5 December 2011
Accepted 5 December 2011
Available online 13 December 2011

*Keywords:*
Churn prediction
Interpersonal influence
Propagation
Telecommunications
Social network

ABSTRACT

Preventing customer churn is an important task for many enterprises and requires customer churn prediction. This paper investigates the effects of interpersonal influence on the accuracy of customer churn predictions and proposes a novel prediction model that is based on interpersonal influence and that combines the propagation process and customers' personalized characters. Our contributions include the following: (1) the effects of interpersonal influence on prediction accuracy are evaluated while including determinants that other researchers proved effective, and several models are constructed based on machine learning and statistical methods and compared, assuring the validity of the evaluation; and (2) a novel prediction model based on interpersonal influence and information propagation is proposed. The dataset used in the empirical study was obtained from a leading mobile telecommunication service provider and contains the traditional and network attributes of over one million customers. The empirical results show that traditional classification models that incorporate interpersonal influence can greatly improve prediction accuracy, and our proposed prediction model outperforms the traditional models.



## 1. Introduction

Preventing customer churn is an important task for many enterprises, especially in matured industries, including telecommunications [15,22,1] and finances [42]. Achieving it requires churn prediction, which is defined as identifying customers who tend to switch to other service providers. Findings from previous research can be categorized according to two aspects. First, churn determinants are analyzed and verified using customer behaviors in various industries. Some attributes, including customer satisfaction, switching costs, customer demographics, tendency to change behavior, and service usage, have been found to be common churn determinants [15,22,1]. Second, researchers have proposed prediction models based on machine learning methods, including the decision tree, neural network, and support vector machine [18,41,8,7], or statistical methods, including logistic regression, survival analysis, and Markov chain [21,24].

Though many practices benefit from these valuable results, one limitation exists. Many researchers have assumed implicitly or explicitly that a customer's decision to switch service providers is independent of other customers' decisions. Most prior research focuses exclusively on individual customers, without accounting for any interpersonal influence [9], as they measure each customer's perception and behavior independently. Typically, many explanatory variables are collected on each customer and used in multivariate prediction modeling. In reality, customers' behaviors not only depend on their own perceptions and subjective desires but also interplay with each other. Thus, customers' choices are interdependent [45].

This paper investigates the effects of interpersonal influence on the accuracy of predicting customer churn and proposes a novel prediction model based on interpersonal influence that combines the propagation process and customers' personalized characters. Our contributions include the following: (1) the effects of interpersonal influence on prediction accuracy are evaluated while including traditional attributes (i.e., customers' personalized characters) that other researchers proved to be effective, and several models are constructed based on machine learning and statistical methods and compared, assuring the validity of the evaluation; and (2) a novel prediction model based on interpersonal influence and information propagation is proposed. The dataset used in the empirical study was obtained from a leading mobile telecommunication service provider and contains the traditional and network attributes of over one million customers. The empirical results show that traditional prediction models incorporating interpersonal influence can greatly improve prediction accuracy, and our proposed prediction model outperforms the traditional models.

* Corresponding author. Present address: 427 West Hall, 1085 South University Ann Arbor, MI 48109-1107, USA. Tel.: +1 7347096295, mobile: +86 13910390890; fax: +86 62282039.
*E-mail addresses:* zhangxiaohang@bupt.edu.cn (X. Zhang), jizhu@umich.edu (J. Zhu), xushuhua@bupt.edu.cn (S. Xu), wanyan@bupt.edu.cn (Y. Wan).

0950-7051/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2011.12.005

This paper is organized as follows. Section 2 discusses several related works. Section 3 provides the churn determinants used in our models. In Section 4, several classification models are separately constructed based on different attributes and methods, and a propagation-based model is proposed. Section 5 discusses the experimental results and their implications. Section 6 gives our conclusions.

## 2. Literature

Customer churn prediction can be regarded as a classification problem, in which each customer is classified into one of two classes, churn or non-churn. Machine learning and statistical methods are the most widely used approaches for classification problems. Popular churn prediction methods include logistic regression [33], decision trees [18], neural networks [41], support vector machines [8], and evolutionary algorithms [3]. However, many previous studies solely focused on customers' individual attributes when using these models. Though some studies accounted for interpersonal influence [9,37], they did not combine it with customers' personalized attributes.

Propagation models have also been widely used to describe the dynamic processes of viral marketing [11,36,19], trust formation [23,25], risk propagation [31], epidemic diffusion [10], and computer virus infection [34]. Popular propagation models include the SI, SIR, and SIS models [16], which are based on different diffusion mechanisms. In our proposed propagation model, we partially employ the SI model and combine its diffusion process with customers' personalized characters. In the SI model, each infected individual can pass information (or influence) to his/her neighbors in a network in which recovery is not possible, i.e., an infected individual continues transmitting information indefinitely.

To our knowledge, little research has been conducted on using propagation models to predict customer churn. Dasgupta et al. [9] proposed a spreading activation-based technique (SPA) that predicts potential churners by examining the current set of churners and their underlying social networks. Using the underlying topology of the customer contact network, the SPA initiates a diffusion process with the churners as seeds. Essentially, it models a "word-of-mouth" scenario, in which a churner influences a neighbor to churn, and the influence spreads from that neighbor to another neighbor, and so on. At the end of the diffusion process, the amount of influence received by each node is inspected. Using a threshold-based technique, a node that is currently not a churner can be declared to be a potential future churner based on the influence it accumulated. However, many propagation models assume that customer behaviors (i.e., churn or non-churn) depend solely on interpersonal influence, not accounting for customers' personalized attributes, i.e., customers are believed to be homogeneous. In reality, customers may behave differently, though they have similar neighbors or interpersonal influence. For example, if a customer has high switching costs, strong influences from other churn neighbors may not lead him or her to churn.

## 3. Determinants of customer churn

For clarity and comparison, we classify customer churn determinants into two categories: network and traditional attributes. Network attributes measure interpersonal influence and describes each customer's local topology in customer contact network and his or her relationships with their neighbors. The other attributes are included in the traditional attributes category, which has been frequently discussed in previous research.

All attributes used in our study were obtained from the telecommunication service database of a mobile telecommunication company. The database stores certain customer-related information and call detail records (CDR), which include the labels of calling and called customers and the start time and duration of the transaction. Our attributes differ from those used in previous research, in which data are based on questionnaires. However, actual customer transactions or billing data may fully represent customer's actual future decisions better than survey data [1]. Moreover, to paint a complete picture of interpersonal influence, the ideal dataset would have measurement of direct communication between customers [17], which can be extracted from the CDR.

### 3.1. Traditional attributes

To assure the validity and credibility of the evaluation of interpersonal influence on prediction accuracy, we include traditional attributes in our study. We focus on five prominent drivers in this literature: customer satisfaction, switching barrier, service usage, price sensitivity, and previous anomaly behavior. We do not discuss traditional attributes in detail, as they have been frequently discussed and used in previous research. Table 1 shows the descriptions of traditional attributes and their supporting references.

### 3.2. Network attributes

We represent interpersonal influence through network attributes, including neighbor composition, tie strength, similarity, homophily, structural cohesion, influence degree of neighbors, and order-2 neighbors (see Table 2).

#### 3.2.1. Neighbor composition

Customers who have direct links with customer $i$ are defined as customer $i$'s neighbors. Neighbor composition is important because different types of neighbors (i.e., churn neighbors or non-churn neighbors) may have different influences. Moreover, external neighbors who belong to different service providers may either explicitly influence customer churn by conveying their perceptions of satisfaction or implicitly influence customers through their choices of service provider.

#### 3.2.2. Tie strength

Tie strength measures the intensity of contact between a pair of customers. Dasgupta et al. [9] show that tie strength can improve churn prediction accuracy in the telecommunications industry. In this study, we quantify tie strength by the total number of calls made between a pair of customers over the studied period, as prior studies [35,9].

#### 3.2.3. Similarity

In this paper, we use the notation of similarity to denote overlap in a local neighborhood, as in other studies [17,35]. Similarity between customers $i$ and $j$ is defined as $C_{ij}/(k_i + k_j - C_{ij} - 2)$, where $k_i$ and $k_j$ denote the number of neighbors of customers $i$ and $j$ respectively, and $C_{ij}$ denotes the number of common neighbors of customers $i$ and $j$ [35].

#### 3.2.4. Homophily

Homophily is the tendency of customers to associate and bond with others who are similar. The presence of homophily has been discovered in a vast array of network studies [32]. In this paper, homophily refers to "churn homophily", which implies that customers with similar characters are more likely to influence each other when making churn decisions. The following measures churn homophily between customer $i$ and his or her neighbor $j$:

$$H_{ij} = \sqrt{\sum_p w_p (a_{ip} - a_{jp})^2}, \tag{1}$$