

Using Combined Model Approach for Churn Prediction in Telecommunication

Fa-Gui LIU, Zhi-Jie ZHANG*, Xin YANG

School of Computer Science and Engineering, South China University of Technology, Guangzhou, China

* Zhijie Zhang (corresponding author) email: zhgzjz@126.com

Keywords: Combined model; Churn prediction; Hybrid data; FKP; SVM

Abstract: To solve the prediction problem of hybrid data in users' consumption information of telecommunication, the paper use the fuzzy K-Prototypes (FKP) and support vector machine (SVM) combined model to improve the accuracy of users churn prediction. In the combined model, FKP is adopted to cluster hybrid large data volume effectively, and then the samples nearby cluster center in each cluster as the input of SVM to promote the prediction efficiency. As shown in the experience validation results, the proposed FKP-SVM combined model has excellent performance in predicting churn, due to reduce the training time of hybrid large-scale data set and save system resources.

1 Introduction

With the development of mobile network, researchers pay more attention to the off-network of mobile users. In accordance with the factors of basic information, consumption information and individual preference of users, the off-network trend of users can be effectively and reasonably predicted. Namely, users are divided into two types: churn and non-churn [1-2]. In the research on the situation of users churn in mobile network, Pareto principle is also applicable. The influence created by 80% ordinary (low ARPU value, APRU which refers to average consumption per user) users will be lower than that of 20% high-value (high ARPU value) users. The effectiveness of retention policy customized for high-value users on prevention of users churn is more obvious [3]. In the research, this principle will also be followed to analyze the churn of high-value users, so that the operator can clearly understand the degree of satisfaction of users for the current mobile service.

At present, in the research of users churn, focusing on the following aspects: (1) hardly considering the establishment of corresponding users churn prediction model for different groups; for example, subdividing users based on the value of users and establishing prediction models respectively; (2) the data type in the users' consumption information has not been considered yet, and effective strategies are not taken to deal with the classification of hybrid data; (3) main algorithms used to predict churn are mostly single model which has not the better stability and effect of the prediction. Such as regression analysis, decision tree, neural network, and SVM (Support Vector Machine) and so on. Meanwhile, the above-mentioned prediction algorithms have their own advantages and disadvantages, and the scope of application has certain limitations. For example, SVM has good classification accuracy in small data set and high time complexity in big data set.

Therefore, for the above-mentioned problems and the characteristic of users consumption data, a fuzzy K-Prototypes and SVM combined model (Fuzzy K-Prototypes Cluster-based Support Vector Machine, FKP-SVM) is proposed to solve the instability and limitations of single prediction model and predict churn trend of high-value users. The purpose is to reduce operation time and improve prediction accuracy. In the combined model, fuzzy K-Prototypes algorithm is firstly adopted to effectively deal with hybrid data and cluster the mixed sample with large data volume; then, selecting samples nearby the cluster center in each cluster as the input of SVM to solve the poor training efficiency of SVM, which caused by the randomness in handling large-scale data set; finally, establishing rapid and accurate prediction model. The experimental

results show that FKP-SVM combined method can reduce the training time of large-scale data set, save system resources and improve the prediction accuracy for hybrid data.

2 Backgrounds

In the prediction research of users churn, attentions are mainly focus on two aspects: firstly, adopting single model method to predict users churn; secondly, adopting combined model method to predict users churn.

There are rich research results of the users churn prediction based on single model, researchers adopt different single models to analyze the users churn. Hadden J et al. [4] compared the effect of methods such as neural network, regression tree and regression model in the users churn prediction. Ying et al. [5] proposed the SVM method with different types of weighting parameters based on the characteristic of unbalanced size of positive and negative samples in actual customers churn data. Xia et al. [6] compared methods such as artificial neural network, decision tree and Bayes classifier from aspects such as accuracy, hit rate, coverage rate and lifting coefficient. Gopal R K et al. [7] firstly adopted ordinal regression method to model the users churn situation. Sato T et al. [8] used principal component analysis method to predict the customers churn and compared this method with Bayes, decision tree and SVM method. Owczarczuk M et al. [9] adopted the logistic regression method to conduct research on mobile and telecom customers churn. Huang et al. [10] proposed a new characteristic set and adopted logistic regression, linear classification, Naive Bayes, decision tree, multilayer perception and SVM to predict customer churn.

It has become an important strategy and method to model users churn with combined model. Tsai C F et al. [11] designed two mixed models to predict customers churn: the first model is the mixed model of neural network + neural network; the second model is the mixed model of self-organizing mapping + neural network. Pendharkar P C [12] proposed the neural network combined model based on genetic algorithm to model on the users churn. Wojewnik P et al. [13] combined K-means clustering algorithm and classical single model classification algorithm to design an integrated model to model for the customers churn and obtain higher hit rate of prediction. He et al. [14] adopted C5.0 decision tree algorithm, CART decision tree algorithm and RBF neural network to make research on the customers churn. Lu et al. [15] adopted logistic regression to establish the churn prediction model for each type respectively. Idris A et al. [16] proposed the feature selection approach based on filter and wrapper to establish the combined prediction model of users churn.

3 Combined Model

For the characteristic of complex factors on users churn, the data set includes various types of attributes which is characterized with the hybrid data type. Therefore, the fuzzy K-Prototypes and SVM combined model method is proposed in this Section to deal with the above-mentioned problems to improve the accuracy of prediction and obtain stable prediction effect. The thought of this combined model will be specifically described below.

3.1 Fuzzy K-Prototypes

Huang et al. propose K-Prototypes algorithm to solve the problems of numeric and symbolic hybrid data[17]. Fuzzy K-Prototypes is expanded from K-Prototypes algorithm. It obtain the minimum value of the objective function by assigning membership to cluster objects and updating the cluster center and dividing matrix to cluster hybrid data effectively.

Given $X = \{X_1, X_2, \dots, X_N\}$, setting the cluster number K and dividing N data sample to K cluster. The cost function is minimized as a clustering criterion:

$$\text{Minimize. } F(W, Z) = \sum_{k=1}^K \sum_{i=1}^N w_{ki}^{\phi} d(X_i, Z_k)$$

Subject to. $w_{ki} \in [0,1]$

$$\sum_{i=1}^K w_{ki} = 1 \quad (1 \leq k \leq K, 1 \leq i \leq N) \quad (1)$$

$$0 < \sum_{i=1}^N w_{ki} < N$$

In Eq.1, $Z = \{Z_1, Z_2, \dots, Z_K\}$, $Z_K = [z_{k1}, z_{k2}, \dots, z_{km}]$ are the cluster center, $Z_{kj} \in DOM(A_j)$, $1 \leq j \leq m$; $\partial > 1$ is fuzzy index; $d(X_i, Z_k)$ is difference measure; $W = \{w_{ki}\}$ is membership of X_i in cluster K .

In the attribute of X_i , $[x_{i1}, x_{i2}, \dots, x_{ip}]$ is numeric attribute, $[x_{ip}, x_{ip+1}, \dots, x_{im}]$ is classification attribute, the difference measure is defined as follows:

$$d(X_i, Z_k) = d_1(X_i, Z_k) + \gamma d_2(X_i, Z_k) = \sum_{\tau=1}^p (x_{i\tau} - z_{k\tau})^2 + \gamma \sum_{\tau=p+1}^m \delta(x_{i\tau}, z_{k\tau}) \quad (2)$$

In Eq.2, $\delta(x_{i\tau}, z_{k\tau}) = \begin{cases} 0 & (x_{i\tau} = z_{k\tau}) \\ 1 & (x_{i\tau} \neq z_{k\tau}) \end{cases}$, $x_{i\tau}$, $z_{k\tau}$ is the value of the τ attribute; γ is the

classification attribute weight of each class.

The division matrix is updated by using the following method:

$$w_{ki} = 1(Z_k = X_i) \quad w_{ki} = 0(Z_l = X_i, l \neq k)$$

$$w_{ki} = \frac{1}{\sum_{l=1}^K \left[\frac{d(Z_k, X)}{d(Z_l, X)} \right]^{\frac{1}{\partial-1}}} (Z_l \neq X_i, 1 \leq l \leq K) \quad (3)$$

The cluster center is updated by the following method:

When $A_j (1 \leq j \leq p)$ is a numeric attribute,

$$Z_{kj} = \frac{\sum_{i=1}^N w_{ki}^{\partial} x_{ij}}{\sum_{i=1}^N w_{ki}^{\partial}} (1 \leq k \leq K) \quad (4)$$

When $A_j (p+1 \leq j \leq m)$ is a classification attribute and $Z_{kj} = a_j^{(\gamma)} \in DOM(A_j)$,

$$\sum_{i=1}^N (w_{ki}^{\partial} | x_{ij} = a_j^{(\gamma)}) \geq \sum_{i=1}^N (w_{ki}^{\partial} | x_{ij} = a_j^{(\tau)}) (1 \leq \gamma, \tau \leq n_j) \quad (5)$$

In Eq.5, n_j is the number of classification attributes A_j . That is, the eigenvalues with the highest frequency appear as the new cluster center eigenvalues.

3.2 SVM

SVM is a kind of supervised learning method. It selects the hyperplane with the maximum distance from member to non-member under the situation of hyperplane with several partitioned data so as to establish the optimal hyperplane to partition the two types of samples. The characteristic of SVM is that it is used to solve non-linear mapping problems and adopted kernel function method to map the complex problem in high-dimensional space to low-dimensional space[18]. The specific algorithm description is as follows:

Minimize. $\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j a_i a_j K(x_i, x_j) - \sum_{i=1}^l a_i$

Subject to. $\sum_{i=1}^l a_i y_i = 0$

$$0 \leq a_i \leq C, i = 1, 2, \dots, l \quad (6)$$

In Eq.7, using the Gaussian kernel function $G(x, y) = e^{-\gamma \|x-y\|^2}$, $\gamma > 0$ is used to control the width

of the kernel, $\|x-y\|^2$ is the square Euclidean distance between eigenvector x and y , C represents the classification interval cost function to solve the noise or error in the sample Tag-induced misclassification.

3.3 FKP-SVM

The FKP algorithm is used to cluster the mixed consumption data, and then the samples near each cluster center as input of SVM to deal with the mixed large-scale data problem effectively. Specific thoughts are as follows:

Firstly, divide original dataset into training set and test set. At the time of selecting training set and test set, training set is 60% of original dataset, and test set is 40% of original dataset.

Secondly, it is the construction of FKP-SVM combined model. Specific process is as follows:

(1) Cluster training set. Initialize cluster number and cluster center of FKP algorithm and carry out cluster processing.

(2) Build FKP-SVM combined model. The cluster center of each cluster in training dataset can be obtained by using FKP cluster algorithm and select the samples closed to cluster center as input of SVM. The selection method is to calculate the matrix of distance between each sample and cluster center, and select N samples with the shortest distance in distance matrix to form input training set of SVM.

(3) FKP-SVM prediction. After the network training is converged, the test set is used to predict the users churn.

Based on the, the cluster results of FKP algorithm is the input of SVM to deal with the mixed data for reducing the running time and improving the prediction accuracy. The FKP-SVM combined model is shown in algorithm 1 and the algorithm flow is shown in Fig. 1.

Algorithm 1: FKP-SVM combined model (x_1, \dots, x_n)

- 1: Dividing the original data set into training sets and test sets.
 - 2: Using the algorithm 1 to cluster the training set.
 - 3: Getting the output value of algorithm 1, that is, the cluster center of each cluster.
 - 4: Calculating the distance matrix of all samples in each cluster, and the input training set of SVM is selected from the distance matrix to construct the FKP-SVM combined model.
 - 5: Calculating the training error of the FKP-SVM combined model. When the training error satisfies the requirement, the iteration is stopped and the training is completed.
-

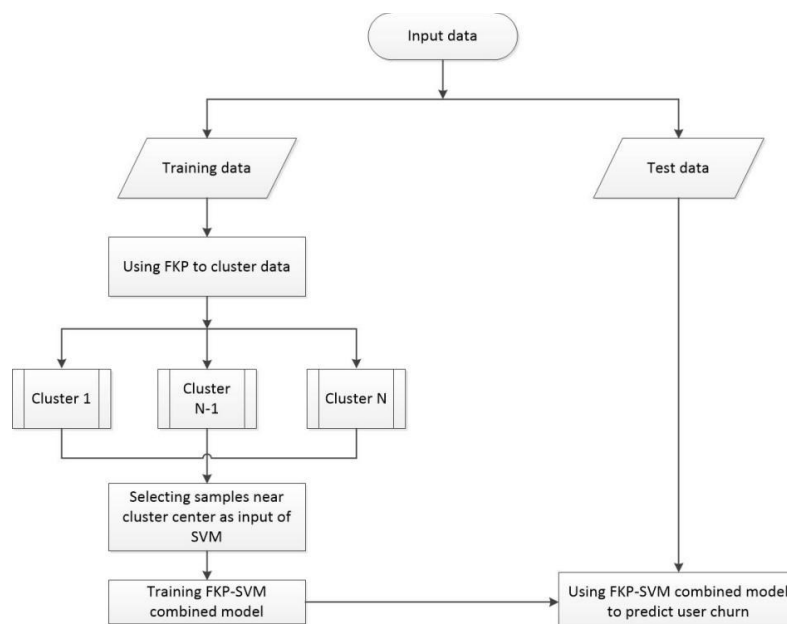


Fig. 1 The flow chart of FKP-SVM combined model

4 Experiment and Analysis

4.1 Data attribute analysis

In this section, combining existing users' consumption information with users' information, consuming behavior, billing data to form the important attributes which are extracted from the mobile core network to form primitive attribute set A ; by analyzing consuming behavior changes of users on time window to expand new attribute set B from primitive attribute set A , the details are in Table 1. The period from September 1, 2015 to October 31, 2015 as time window to discrimination of user churn. Wherein, randomly select 28,504 users with ARPU value higher than 50 yuan. After the preprocessing operation on the data set, the numbers of valid samples are 28,379, which account for 99.56% of total samples. Among them, 3901 are churn users, churn rate is 13.75%. This dataset is denoted as "churn dataset".

Table 1 The attribute set of consumption information

No.	A	B
1	Call number	Call number
2	Type (3G, 4G)	Type (3G, 4G)
3	ARPU per month	ARPU per month
4	Sex	Sex
5	Age	Age
6	Network age	Network age
7	Package type	Package type
8	Monthly call duration	Monthly call duration
9	Time	Time
10	Site	Site
11	Monthly number of messages	Monthly number of messages
12	Monthly internet access	Monthly internet access
13	Value added and total cost	Value added and total cost
14		call duration changes for nearly a month
15		bill change for nearly a month
16		basic fee changes for nearly a month
17		new fee changes for nearly a month
18		Balance for nearly a month
19		Number of downtime for nearly a month
20		Call the hotline for nearly a month

4.2 Model process

In this section, the model will be made on user churn and specific process is as follows:

- (1) Dividing "churn dataset" into training set and test set, the division proportion of dataset is 6:4.
- (2) Initializing the parameters of FKP algorithm. Wherein, the cluster initialization center is randomly generated, the number of cluster is 4, the minimum allowable error is 10^{-4} , classification attribute weight γ is 1.1 and the fuzzy index is 2.
- (3) In training set, using FKP to cluster dataset.
- (4) In cluster result, selecting 200 samples which are close to each cluster center as the input of SVM. In SVM, Gaussian kernel function $G(x, y) = e^{-\gamma \|x-y\|^2}$, γ and C is 1 respectively.
- (5) Training FKP-SVM combined model until network is converged.
- (6) In training set, using FKP-SVM combined model to predict users churn.

4.3 Model evaluation

In this section, in order to verify the performance of FKP-SVM combined model, K-means-SVM [13] and FCM-SVM [19] combined model methods are introduced to compare with it. The following evaluation model is adopted to assess the experiment results [6], true (actual) churn in datasets is compared with predicted churn, as shown in Table 2.

Table 2 Churn evaluation matrix

Churn status	Predicted churn	Predicted non-churn
True churn	A	B
True non-churn	C	D

According to Table 2, calculating the accuracy, hit rate and coverage rate of the model, the specific formula is as follows:

$$\text{Accuracy} = \frac{A+D}{A+B+C+D}, \text{ Hit rate} = \frac{A}{A+C}, \text{ Coverage rate} = \frac{A}{A+B}.$$

In the experiment, K-means-SVM, FCM-SVM, and FKP-SVM combined model methods will be respectively adopted to predict users churn. In these three models, cluster number is 4, cluster center will be generated randomly, 1-200 data sample of each cluster which is close to cluster center will be selected as the input of SVM. The prediction results are shown in Table 3 and Table 4.

Table 3 Test results of three combined models

Churn \ Model	K-means-SVM		FCM-SVM		FKP-SVM	
	1	0	1	0	1	0
1	1113	571	1353	262	1421	51
0	685	8982	284	9452	94	9785

Table 4 Evaluation results of three combined models

Combined model	Accuracy	Hit rate	Coverage rate
K-means-SVM	88.93%	61.90%	66.09%
FCM-SVM	95.19%	82.65%	83.78%
FKP-SVM	98.72%	93.80%	96.54%

Experiment results in Table 3 and Table 4 show that the accuracy, hit rate, and coverage rate of FKP-SVM combined model are superior to those of K-means-SVM and FCM-SVM combined model. The reason is that FKP can cluster hybrid data effectively, and it can treat numerical data attributes and categorical data attributes differently. Besides, different calculation methods and strategies are adopted to carry out reasonable calculation of various attributes. Therefore, the better cluster effect is obtained, and the accuracy of subsequent algorithm in combined model is also improved. However, K-means cluster algorithm and FCM cluster algorithm are not considered the difference between data attributes. However, fuzzy boundary and uncertain factors are considered in FCM clustering, the prediction effect and accuracy of subsequent algorithm in FCM are better than those of K-means method.

5 Conclusion

According to the Pareto principle, the impact of high-value users will be much higher than the average user and the effect of retention policy at high-value users on prevention of users churn is more obvious. Therefore, this research is mainly centered on churn of high-value mobile users. Based on the prediction problem of mixed attributes existing in big users' consumption data, FKP-SVM combined model is designed to predict users churn and solve prediction problems of hybrid large-scale data. In this model, FKP algorithm can be used to effectively cluster mixed attributes through carrying out different calculations of numerical attributes and categorical attributes. In cluster result, samples which are close to cluster center are selected to be input of SVM. In this way, the problem of classification accuracy of SVM for large-scale samples can be resolved to obtain the better classification result. In the experience, the attributes in original dataset have been extended to reflect the changes of users' consuming behaviors clearly. Besides, multiple evaluation indicators have been used to evaluate the performance of FKP-SVM combined

model. Comparing the proposed combined model with others from the aspects of accuracy, hit rate and coverage rate, FKP-SVM has a better effect in dealing with hybrid large-scale data and can predict users churn trend accurately.

Acknowledgement

The authors thank the anonymous reviewers and editors for their valuable comments on improving this paper. This paper is partially supported by the Engineering and Technology Research Center of Guangdong Province for Logistics Supply Chain and Internet of Things (Project No. GDDST[2016]176); the Provincial Science and Technology Project in Guangdong Province (Project No. 2013B090200055); the Key Laboratory of Cloud Computing for Super-integration Cloud Computing in Guangdong Province (Project No. 610245048129); the Engineering and Technology Research Center of Guangdong Province for Big Data Intelligent Processing (Project No. GDDST[2013]1513-1-11).

References

- [1]Zhang, Guozheng. "Customer Segmentation Based on Survival Character." International Conference on Wireless Communications, NETWORKING and Mobile Computing IEEE, 2007:3391-3396.
- [2]Chang Xiaoning. Mobile communication customer churn problem - Taking a company as an example. Tongji University, 2007.
- [3]Craft, Ralph C., and C. Leake. "The Pareto principle in organizational decision making." Management Decision volume 40.8(2002):729-733.
- [4]Hadden, John, et al. "Churn prediction using complaints data."Enformatika (2008).
- [5]Ying, Wei Yun, et al. "Support Vector Machine and Its Application in Customer Churn Prediction." Systems Engineering-Theory & Practice27.7(2007):105-110.
- [6]Xia, Guo En, and W. D. Jin. "Model of Customer Churn Prediction on Support Vector Machine." Systems Engineering - Theory & Practice28.1(2008):71-77.
- [7]Gopal, Rupesh K., and S. K. Meher. Customer Churn Time Prediction in Mobile Telecommunication Industry Using Ordinal Regression. Advances in Knowledge Discovery and Data Mining. Springer Berlin Heidelberg, 2008:884-889.
- [8]Sato T, Huang B Q, Huang Y, et al. Using PCA to Predict Customer Churn in Telecommunication Dataset[C]// Advanced Data Mining and Applications -, International Conference, Adma 2010, Chongqing, China, November 19-21, 2010, Proceedings. DBLP, 2010:326-335.
- [9]Owczarczuk, Marcin. "Churn models for prepaid customers in the cellular telecommunication industry using large data marts." Expert Systems with Applications 37.6(2010):4710-4712.
- [10]Huang, Bingquan, M. T. Kechadi, and B. Buckley. "Customer churn prediction in telecommunications." Expert Systems with Applications 39.1(2012):1414-1425.
- [11]Tsai, Chih Fong, and Y. H. Lu. "Customer churn prediction by hybrid neural networks." Expert Systems with Applications 36.10(2009):12547-12553.
- [12]Pendharkar, Parag C. "Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services ☆." Expert Systems with Applications 36.3(2009):6714-6720.

- [13]Wojewnik, Piotr, et al. "Social-Network Influence on Telecommunication Customer Attrition." *Kes International Conference on Agent and Multi-Agent Systems: Technologies and Applications* Springer-Verlag, 2011:64-73.
- [14]Yue, H. E., H. E. Zheng-Lin, and X. Y. Zhou. "Combined Early Warning of Major Customer Lose in Wireline Telecom Based on Data Mining." *Soft Science* (2012).
- [15]Lu, Ning, et al. "A Customer Churn Prediction Model in Telecom Industry Using Boosting." *IEEE Transactions on Industrial Informatics* 10.2(2014):1659-1665.
- [16]Idris, Adnan, and A. Khan. "Churn Prediction System for Telecom using Filter–Wrapper and Ensemble Classification." *Computer Journal* (2016):bxv123.
- [17]Huang, Zhexue, and M. K. Ng. *A fuzzy k-modes algorithm for clustering categorical data.* IEEE Press, 1999.
- [18]Cherkassky, Vladimir, and F. Mulier. *Statistical Learning Theory.* Encyclopedia of the Sciences of Learning. 2012:3185-3185.
- [19]Esme, Engin, and B. Karlik. "Fuzzy c-means based support vector machines classifier for perfume recognition." *Applied Soft Computing* 46.C(2016):452-458.