NOMBRE DEL CURSO/CLAVE:	TC4030 Inteligencia Artificial y Aprendizaje Automáti
NOMBRE DEL PROFESOR TITULAR:	Luis Eduardo Falcón Morales
NOMBRE DE LA REVISTA:	5th International Conference on Cloud Computing and Big Data Analytics
NOMBRE DEL AUTOR:	Xin Hu, Yanfei Yang, Lanhua Chen y Siru Zhu
NOMBRE DEL ARTÍCULO:	Research on a customer churn combination prediction model based on decision tree and neural
EDITORIAL:	IEEE
EDICIÓN:	1a edición
NÚMERO DOI:	10.1109/ICCCBDA49378.2020.9095611
FECHA DE PUBLICACIÓN:	2020
RANGO DE PÁGINAS:	129 a 132
TOTAL DE PÁGINAS:	4

2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics

Research on a Customer Churn Combination Prediction Model Based on Decision Tree and Neural Network

Xin Hu
Basic Department,
Air Force Early Warning Academy,
Wuhan, Hubei, China
e-mail: sstcookie@163.com

Yanfei Yang
Basic Department,
Air Force Early Warning Academy,
Wuhan, Hubei, China
e-mail: sstcookie@163.com

Lanhua Chen
Basic Department,
Air Force Early Warning Academy,
Wuhan, Hubei, China
e-mail: sstcookie@163.com

Siru Zhu
Basic Department,
Air Force Early Warning Academy,
Wuhan, Hubei, China
e-mail: sstcookie@163.com

***Abstract*—Customer churn is a prominent issue facing companies. Preventing customer churn, trying to retain and retain customers has become an important issue for business operations and development. Most of the current customer churn predictions use a single prediction model, which makes it difficult to accurately predict customer churn. Based on the prediction results and confidence of decision tree and neural network model, this paper designs a combined prediction model of customer churn and conducts empirical research on the effectiveness of the model. The prediction results show that compared with the single customer churn prediction model, the combined prediction model has higher accuracy and better prediction effect, and can more intuitively display the basic characteristics of the churn customers.**

Keywords-customer churn; decision tree; prediction model; neural network

I. INTRODUCTION

Since China's entry into the WTO, many industries have opened to the outside world, which has led to a more fierce market competition environment for enterprises. The intensified market competition environment has made the problem of customer churn faced by enterprises more and more serious. Generally, the original customers of an enterprise no longer purchase the company's products or receive corporate services are called corporate customer churn. Different industries have different definitions of customer churn [1]. Generally speaking, it can be divided into two categories: active customer churn and passive customer churn. Usually excessive customer churn can have a significant impact on a company's performance. How to retain new customers while developing new customers has become a subject that related staff of the company have to study [2].

Data mining technology is the most commonly used method to predict customer churn. Data mining technology uses decision tree, neural network, classification regression tree and other technologies to build a model for predicting customer churn. Through the model, information useful for predicting customer loss can be extracted from a large number of customer data so that enterprises can formulate relevant customer work plans based on this information [3]. At present, domestic and foreign customer churn prediction algorithms include predictions based on traditional statistics and predictions based on combined classifiers [4]. Based on machine learning methods and statistical theory, Y. Hang et al. [5] used customers' demographic statistics to correlate indicators to predict churn customers. Based on the transaction time of retail customers, Miguéis et al. [6] established a predictive model based on Logistic regression. Yin Ting et al. [7] combined the prior information method of Bayesian classification with the information entropy gain method of decision tree classification and applied it to the analysis of customer churn in the telecommunications industry. Du Gang and Huang Zhenyu [8] used the improved decision tree model to predict customer purchase behavior, and analyzed the effect comparison before and after optimization to verify the effectiveness and efficiency of the improved algorithm in the customer purchase behavior. Cui Yongzhe [9] uses the C4.5 algorithm in the decision tree algorithm to establish a churn early warning model for telecommunications customers.

However the traditional parametric model or a single artificial intelligence-based method cannot achieve relatively high-precision prediction, so the establishment of a combined prediction model to improve the prediction accuracy is an inevitable trend to solve the problem of customer churn [10]. Based on this, this paper designs a combined prediction model based on two models of decision tree C5.0 and neural network. The confidence of the two prediction models is used as the weight, and the weighted score is used as the customer churn probability. The combined prediction model is used to predict customer churn in a supermarket. By comparing the prediction accuracy of the three models, the validity of the combined prediction model is verified.

II. CUSTOMER CHURN PREDICTION MODEL

A. Decision Tree Customer Churn Prediction Model

Based on the information gain theory, the decision tree is one of the most widely used data classification algorithms at present. The decision tree structure contains several nodes and branches, where nodes represent tests on a certain attribute and branches represent the results of the tests. Common decision tree algorithms include ID3, C5.0, etc., which are mainly used for predictive analysis of events. The decision tree prediction process is performed in two steps: one is to build and evolve a decision tree using the training set; the other is to test the attribute values of each node, classify the input data, and use the attribute values of this class to complete the estimation of the prediction object [11].

This article adopts the C5.0 decision tree method, which is widely used in the industry for dirt tolerance and strong interpretation ability, to build a customer churn model. The C5.0 model uses the field with the largest information gain to split the sample. The sample subset obtained from the first split is usually split according to another attribute. This is repeated until the sample subset can no longer be split. Finally, we remove or trim out a subset of the samples that do not contribute significantly to the model.

The model uses the bootstrap method to improve the accuracy of the C5.0 algorithm. The entire model set is used for sample classification, and each decentralized prediction is integrated into a comprehensive prediction through a weighted voting process. When the node's flow is executed, two new fields are added: the predicted value of each record and the confidence.

B. Neural Network Customer Churn Prediction Model

As a data analysis mode of human brain thought simulation, neural networks are based on massive data parallel processing and calculation, and are used to describe cognitive, decision-making and other intelligent control behaviors. The model structure of a typical neural network includes input layer, hidden layer, and output layer, which are connected by several neurons. BP neural network is the most widely used neural network algorithm, and its output expression [12] is:

$$H = f_i(\sum \omega_{ij} x_i + \theta_j)$$

where ω_{ij} is the connection weight coefficient, f_i is the excitation function, θ_j is the threshold of the neuron, and x_i is the input of the neuron. BP neural network is trained using a teacher-learning method and can implement any complex non-linear mapping function. The training process is based on the principle of minimum output error, and the connection weight coefficients and thresholds are modified layer by layer.

This paper selects the customer related characteristic attributes as the input of the neural network, and the output is whether the customer is lost. In order to facilitate the effective evaluation of the model later, 2/3 of the data is randomly selected as the training set and 1/3 as the test set. By default, every time a neural network node is executed, a brand new network is created. But when the "Continue to train the existing model" option is selected, the training will continue to use the network successfully generated by the previous node. When a stream containing the generated neural network model is executed, for each output field in the original training data, a new field is added to the stream. This new field contains the network prediction for the corresponding output field. For the symbol output field, another new field is added that contains the confidence of the prediction.

C. Combined Customer Churn Prediction Model

The so-called combined prediction model takes the confidence of the decision tree and the neural network prediction model as the weight, the prediction result as the variable, recalculates the probability of customer churn by weighting, and gives the prediction result of customer churn based on the probability of churn. The formula for calculating the probability of customer churn is as follows:

$$P = \frac{1}{2}(aX_1 + bX_2) \quad (1)$$

where a, b are the confidence of each sample record in the decision tree and neural network customer churn prediction respectively, and X_1, X_2 are the prediction results of customer churn in the decision tree and neural network model respectively. The value of X_1, X_2 is either 0 or 1. If the customer is lost, the value is 1; if the customer is not lost, the value is 0.

According to formula (1), if the prediction results of both models are not churn, that is, the values of X_1 and X_2 are both 0, then the customer churn probability P of the combined prediction is also 0, that is, the customer is not churn. If the prediction results of both models are churn, that is, the values of X_1 and X_2 are 1, and the accuracy of the predictions of both models is above 90%, then the customer churn probability P of the combined prediction will be greater than 0.5, that is, the customer is predicted churn. If the prediction result of one model is churn and the prediction result of the other model is not churn, then the churn probability P of the combined prediction will be less than 0.5. Further, if the confidence level predicted for the churn record is greater than 0.8, indicating that the model predicts the customer churn with a greater probability, then the combined model will give the probability of customer churn. If the confidence of the churn record is less than 0.8, it is considered that the possibility of customer churn is small, then the combination model predicts that the customer is not churn.

It can be seen that the biggest advantage of the combined customer churn prediction model is that it can integrate the results of the two models to clearly distinguish between churn customers and non-churn customers, and for customers in between, it gives the probability of churn. This can help marketers to more accurately identify the churn status of customers with different values, and select customers with different churn probability for marketing purposes.

III. Empirical Analysis of the Combined Prediction Model

A. *Data Collection and Selection of Attributes*

The data comes from the customer customer information of a supermarket from June 2018 to April 2019. After missing value processing, 2681 member customers were finally used for customer churn prediction analysis.

The member database has accumulated a lot of information about members, such as gender, occupation, age, average monthly income, consumption amount, etc. We call them attributes. Some of these attributes are closely related to customer loyalty, while others are not. The first step in customer churn prediction is to choose the most reasonable customer characteristic attributes. The membership characteristic attributes selected in this article include membership card level, age, trend value of the number of purchases, sum of non-shopping points, unit shopping points and average shopping interval. Among them, the membership card level refers to the level of our store's membership card owned by member customers, which are ordinary members, silver members, gold members, and platinum members in descending order. The trend value of the number of purchases indicates the speed at which customers' purchases increase or decrease within a certain period of time. The sum of non-shopping points refers to points earned by members through activities such as sharing products. Unit shopping points represent the ratio of the total points accumulated by customers to the total amount of purchases over a period of time. The average shopping interval is the average shopping interval of a member customer during a certain period of time.

B. *Data Pre-processing and Churn Judgment*

Before the analysis, first of all the indicators are dimensionless. For continuous numeric variables, all data is mapped between [0, 1]. The next step is to summarize the discrete variables and divide the customer age into six age groups. Finally, some new derived variables are generated, including unit shopping points, average shopping interval and trend value of the number of purchases. The trend value of the number of purchases indicates the speed and direction of the increase or decrease of the number of purchases by the customer within 11 months, and is expressed by the slope of a linear regression. The slope is greater than 0, indicating that the number of customer purchases is increasing, and the larger the value, the faster the increase. On the contrary, if the slope is less than 0, it means that the number of customer purchases is decreasing, and the smaller the value is, the faster the decrease is. The characteristic attribute values of a member customer are shown in Table I. In Table I, an ordinary member customer with a card number of 1800193, is 34 years old, the sum of non-shopping points is 0 points, the average shopping interval is 41.6 days, and the unit shopping point is 0.504 points / yuan. The number of customer purchases is increasing and the trend of purchases is 0.112.

TABLE I. Characteristic Attributes of a Member Customer

Member card number	Membership Card Level	Age	Non-shopping points	Average shopping interval	Unit shopping points	Trend value of purchases
1800193	Ordinary member	34	0	41.6	0.504	0.112

In order to predict customer churn based on these attributes, this article compares the latest shopping interval of member customers with the maximum shopping interval as a criterion for judging whether customers are churn. Among them, the latest shopping interval refers to the length of time that the member customer's last shopping time is away from the current time, and the maximum shopping interval refers to the maximum interval between two purchases made by member customers within a certain period of time. When the customer's latest shopping interval is greater than the maximum shopping interval, the customer is considered to be churn, otherwise the customer is considered to be not churn. According to this customer churn judgment standard, out of 2,681 member customers, there were 669 churn customers and 2012 unchurn customers.

C. *Application of Combined Prediction Model*

After the data is prepared and pre-processed, the data can be input into the model. First, the customer churn prediction is performed using the decision tree prediction model and the neural network prediction model respectively. This article uses the C5.0 algorithm in SPSS modeler 14.1 to build a decision tree customer churn prediction model, and selects "Expert Mode" in the "Model" option, and sets "Construction Severity" to 80%. In order to improve the accuracy of predicting churn customers, this article sets the misclassification loss that actual churn customers are predicted to be churn to 10. After the model is run, the prediction result and the confidence level are obtained. When using the neural network customer churn prediction model, the membership characteristic attribute is used as an input field of the neural network customer churn prediction model, with "whether churn" as the target, and the proportion of samples selected for modeling is 67%. After the model is output, the prediction result and the confidence level are obtained. Tables II and III are the confusion matrices predicted by the two models respectively.

TABLE II. Prediction Results of the Decision Tree Model

	Predicted churn customers	Predicted unchurn customers
Actual churn customers	562	107
Actual unchurn customers	68	1944

TABLE III. Prediction Results of the Neural Network Model

	Predicted churn customers	Predicted unchurn customers
Actual churn customers	618	51
Actual unchurn customers	45	1967

Then the prediction results and confidence of the customer churn model of the decision tree and neural network are summarized, and the prediction results are converted: not churn is recorded as 0, churn is recorded as 1. Substitute into formula (1) for weighted calculation, and the customer churn probability P is between 0 and 0.99. When P does not exceed 0.4, the prediction result of the combined prediction model is that the customer is not churn. When P is not less than 0.5, the prediction result of the combined prediction model is the customer churn. And when P is between 0.4 and 0.5, the combined prediction model gives the probability value of the customer churn. Table IV shows the prediction results of the loss of some member customers.

TABLE IV. PREDICTION RESULTS OF CHURN OF SOME MEMBER CUSTOMERS

Member card number	Decision tree prediction results	Decision tree confidence	Neural network prediction results	Neural network confidence	Customer churn probability	Whether churn
1800200	0	0.912	0	0.878	0.0000	not churn
1800371	1	0.721	0	0.543	0.3605	not churn
1800433	1	0.823	0	0.521	0.4115	0.4115

D. Model Evaluation

The evaluation index for predicting the quality of the customer churn model is the accuracy rate. The so-called accuracy rate refers to the ratio of the number of predicted correct customers to the total number of customers. The combined churn model is used to predict member customers, and the results are shown in Table V. Of the 2681 customers, 21 have a churn probability between 0.4 and 0.5. Excluding these 21 customers, there are 2660 customers left. Among these 2660 customer churn predictions, there are 2630 correct predictions and 30 incorrect predictions. The accuracy of the model prediction is as high as 98.87%.

TABLE V. PREDICTION RESULTS OF THE COMBINED PREDICTION MODEL

	Predicted churn customers	Predicted unchurn customers	Customers with a churn probability of (0.4,0.5)
Actual churn customers	640	17	12
Actual unchurn customers	13	1990	9

TABLE VI. COMPARISON OF PREDICTION ACCURACY OF THE THREE MODELS

Prediction Model	Prediction accuracy
Decision tree customer churn prediction model	93.47%
Neural network customer churn prediction model	96.42%
Combined customer churn prediction model	98.87%

In order to further analyze the prediction accuracy of different model algorithms, the prediction results of the decision tree churn model, the neural network churn model and the combined churn model are compared, as shown in Table VI. The results show that the accuracy of customer churn prediction using the combined model is higher than that of decision tree and neural network customer churn prediction models. This is mainly because the combination model makes full use of the advantages of a single prediction model and improves the accuracy of the prediction. Enterprises can formulate corresponding countermeasures to avoid customer churn based on the prediction results.

IV. CONCLUSION

Aiming at the problem that a single model is difficult to achieve high-precision customer churn prediction, this paper uses the prediction results and confidence of decision tree prediction model and neural network prediction model to build a combined customer churn prediction model. The empirical results show that the combined prediction model can not only have a better interpretation ability like a decision tree model, but also a higher prediction accuracy rate of a neural network model, which can better make up for the shortcomings of a single prediction model, and can also get more stable and accurate prediction results.

ACKNOWLEDGMENT

Grateful acknowledgement is made to Ms. Shi Ziyan who provided me with the data, and Mr. Fang Chensheng who gave me considerable help by means of suggestion, comments and criticism. Without their help, the completion of the article would be impossible. In addition, I deeply appreciate the contribution to this thesis made in various ways by my friends and colleagues.

REFERENCES

[1] Chen Mingliang. Discussion on the framework of the basic theoretical system of customer relationship management[J]. Journal of Management Engineering, 2006,20 (4): 36- 41.

[2] Li Yang. Neural network-based customer churn data mining prediction model[J]. Journal of Computer Applications, 2006,33 (S1): 48-51.

[3] Zeng Yaohui. Application of Data Mining in Customer Relationship Management of Communication Industry[J]. Telecommunication Engineering Technology and Standardization, 2006,19 (7): 63- 66.

[4] Yu Xiaobin, Cao Jie, Kong Zaiwu. A review on customer churn[J]. Computer Integrated Manufacturing System, 2012(10).

[5] Zhang X, Zhu J, Xu S, et al. Predicting Customer Churn through Interpersonal Influence[J]. Knowledge-Based Systems, 2012, 28(6):97-104.

[6] Miguéis V L, Van Den Poel D, Camanho A S, et al. Modeling Partial Customer Churn: On the Value of First Product-category Purchase Sequences[J]. Expert Systems with Applications, 2012, 39(12):11250-11256.

[7] Yi Ting, Ma Jun, Qin Xizhong, et al. Application of Bayes Decision Tree in Prediction of Customer Churn[J]. Computer Engineering and Applications, 2014(7).

[8] Du Gang, Huang Zhenyu. Prediction of Customer Buying Behavior in Big Data Environment[J]. Management Modernization, 2015(1).

[9] Cui Yongzhe. Application of data mining technology in early warning of customer churn[J]. Journal of Yanbian University: Natural Science Edition, 2008, 34 (2).

[10] Chen Chen. Analysis of Telecom Customer Churn Forecast Based on Combination Forecast [D]. Hunan University, 2011.

[11] Yu Lu. A Combined Prediction Model of Telecommunications Customer Churn[J]. Journal of Huaqiao University (Natural Science Edition), 2016, 37 (5): 637- 640.

[12] Li Aiqun, Qiao Yan, Wang Ruchuan, et al. Analysis of Telecom Customer Churn Based on Distributed Hybrid Data Mining [J]. Computer Technology and Development, 2010, 20 (10):43-46.