ICEB 2015 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-6-2015

Does A Customers Own Review Behavior Have An Impact On Its Purchase Behavior? Analyzing The Impact Of Review Platform On Group-Buying Platform——A Study Based On Dianping.Com

Huili Liu

Meng Zhao

Jiayin Qi

DOES A CUSTOMER'S OWN REVIEW BEHAVIOR HAVE AN IMPACT ON ITS PURCHASE BEHAVIOR? ANALYZING THE IMPACT OF REVIEW PLATFORM ON GROUP-BUYING PLATFORM-----A STUDY BASED ON DIANPING.COM

Huili Liu,
Meng Zhao,
Jiayin Qi,
Key Laboratory of Trustworthy Distributed Computing and Service (BUPT), Ministry of Education,China
School of Economics and Management, Beijing University of Posts and Telecommunications, China
yucailhl@163.com,
xyfzhm@bupt.edu.cn,
qijiayin@bupt.edu

ABSTRACT

With the development of Web 2.0, traditional customers have increasingly transferred to online purchase and created a large volume of User Generated Content (UGC) on the Internet. The changes brought traditional customer relationship management a great impact and forced companies to adapt, change and evolve. The previous researches have studied the influence of crowds' feedback on customer's purchase behavior, but little researches explore the impact of customer's own review behavior on its purchase behavior. In this paper, our study seeks insights into analyzing the impact of customer's own review behavior on its purchase behavior and discovering how this effect could be fully utilized to predict customer's next stage churn. Based on data from Dianping.com, a famous comprehensive website which contains review and purchase platforms, we build the Logit regression model, considering customer's own review and purchase behavior and finding the impact of user's own review behavior on purchase behavior. Finally, we also use ten-fold cross-validation to prove the stability of our model.

Keywords: User generated content,Logit regression model, Customer churn model, Customer review, Ten-fold cross-validation

INTRODUCTION

Does a customer's own review have an impact on its purchase behavior? With the continuous development of Web 2.0, more and more customers will post a comment to share their experience after they make some purchase online. New e-commerce sites with comments has arisen at home and abroad, such as Dianping.com and Douban.com in China, as well as Yelp.com in America. There are a mass of customer reviews on these website, which have offered information reference for both online and offline customers. The massive reviews has met the purchase requirement of online customers and attracted the offline customers. Data show that there are more than 23 million comments on Dianping.com till the last quarter of 2012. A number of researchers have discovered that plenty of customer reviews will make an effect on other customers' purchase behavior. Chevalier and Mayzline (2006) find that emotion in comments have an effect in the customer's decision [2]. Research shows that more online reviews and greater intensity will lead to greater impact on customer purchase intention [9]. Research also shows that review valence, other user's aggregated helpfulness rating of the review, and another user's verbal agreement or disagreement with the review will affect customer decision [3]. All these research study the review's effect on other user's purchase behavior. However, there is little literature related to whether review will influence the customer's own purchase behavior.

Customer review behavior is a sign of customer participation. From the behavior perspective of customer participation, it is identified as a customer behavior. Kellogg (1997) believe that customer participation include preparing for the purchase, contacting with the firm in the purchasing process, and giving some suggestion to the firm after the purchase [4]. As can be seen, customer review behavior is a process of giving suggestions to the firm after making some purchase, it is also a process of preparing for the customer's next purchase. From the result perspective of customer participation, customer participation is a process of value creation. Lloyd (2003) think that the process of customer participation contribute resources and capability, which will affect the quality of service received by the customer itself [5]. As a portion of customer participation process, customer review will affect the quality of service received by itself, and even affect its subsequent purchase behavior.

Above all, customer's own review behavior will not only affect other customer's purchase behavior, but also affect its own subsequent purchase behavior. In this paper, we will mainly consider the effect of customer's own review on its purchase behavior. With customer's churn behavior as the dependent variable, we will study the effect of customer review behavior on the next stage of churn behavior. Furthermore, we have built the model to predict the customer churn behavior more accurately.

We organize the rest of the article as follows: Firstly, the literature review is presented. Then, the research design and model building are described. Thirdly, the methodology is presented, including the data description, variables explanation, and result analysis. Finally, this research is concluded.

LITERATURE REVIEW

Factors Affecting Customer Churn Behavior

Currently, researchers have studied the factors affecting customer churn behaviors, mainly divided into two aspects: customer own factors and previous purchase behaviors. As for customer own factors, researchers have taken the effect of demographics on customer churn behaviors into consideration and bring it into the prediction models. For example, Zhu and Zhang (2010) introduced five demographic characteristics about customer gender, age, marital status, educational attainment and annual income [10]. Wang (2013) further treated customer category and profession as the key attributes in his model [7]. Cao, Xu and Shen (2012) found the significant effect on customer churn behaviors related to their gender and age, using so-called multidimensional commercial bank customer churn prediction model based on RFM model and demographic variable [1]. Ren and Zhang (2012) also considered customer age, gender and annual income in their prediction model [6].

Regarding the customer previous purchase behaviors, researchers have thought about the influence of indicators involved in RFM model and other purchase behaviors in prediction models. For instance, Wang (2013) studied three indicators of RFM models including total transaction amount, total transaction frequency and the time of last transaction as well as the basic customer reward points, phased transaction amount and quantity, etc. [7]. Cao, Xu and Shen (2012) used the time interval of last transaction and the total transaction amount and frequency to predict customer churn behaviors [1]. Ren and Zhang (2012) adopted the transaction amount, repeat transaction frequency, the number of transactions during the day and night in their model [6]. Zhu and Zhang (2010) also imported repeat transaction frequency, the time of last transaction, transaction amount, the number of transactions during the day, at night and midnight [10].

To sum up, we give an overall review about the main factors affecting customer churn behaviors by various researchers in Table 1:

Table 1. The factors affecting customer churn behaviors

dimensions	factors		
customer own factors	gender	profession	category
	age	marital status	
	income	educational attainment	
customer previous purchase behaviors	transaction amount	the number of transactions during the day	basic customer reward points
	repeat transaction frequency	the number of transactions at night	the number of transactions during worktime
	the total transaction frequency	the number of transactions at midnight	the time of last transaction
	the time of first transaction		

As shown in table 1, it is easy to find that researchers just consider the effect of customer's own factors and previous purchase behaviors on customer churn behaviors at the present stage. With the development of Web 2.0 era, more and more customers tend to publish their opinions and views of product on purchase platforms. Will this spontaneous review behavior affect the customer's own churn behaviors? As far as we know, no researchers have done similar works, and thus, we further consider the customer's own review behavior as well as customer's own factors and purchase behaviors in our prediction model, focusing on the effect of customer's own review behavior and the accuracy of the prediction.

Method for the Prediction of Customer Churn Behaviors

Researchers have recently adopted several methods to predict customer churn behaviors, such as Logit regression model [8], neural network [6] [7], SMC [10] and LSSVM [6]. Detailed comparison can be found in Table 2:

Table 2. comparisons among different prediction methods

model name	description	pros	cons
Logit regression model	econometric model	The introduction of multiple independent variables and their influence on dependent variables and significance level, high dynamic nature.	high complexity, hard to determine parameters
neural network	BP neural network (BPNN) as the most widely used feedforward type neural network, is composed of input, hidden and output layer which are all consist of nerve cell with different transfer functions.	Any complex nonlinear mapping functions can be implemented successful; self-learning ability; Multivariable system, i.e. the number of input and output variable is arbitrary	too long learning time; Number of hidden layer neurons is uncertain, needing cut-and-trial
SMC	calculate the activity, identifying those customers who are still active	directly calculate the activity of individual customers	use only a few key variable information, ignore a lot of explanation variable, prediction performance is not ideal
LSSVM	a modified SVM with the introduction of least square linear system, using equality constraints instead of inequality constraints, the solving process varies from the quadratic programming method to solve a set of equations, relatively faster.	very flexible nonlinear modeling capabilities, well to get the nonlinear mapping relationship between input and output variables.	high complexity

As described above, we adopt the Logit regression model to study the effect of customer’s own review behavior on their churn behaviors due to the pros mentioned in Table 2, building a comprehensive model with customer’s own factors, customer purchase behavior and customer’s own review behavior in consideration.

RESEARCH DESIGN AND MODEL BUILDING

Conceptual Framework

Consumer’s current behavior can indirectly reflect the possibility to purchase in the next phase, which can be used as an important basis to estimate the customer churn behaviors. In era of Web 2.0, more and more customers tend to publish their opinions and views of product after purchase, which can indirectly reflect the use viscosity of customers with respect to the product, enterprise or website and make the enterprise consider the factors of customer churn behaviors roundly with the aid of a modified prediction model. Herein, we further introduce the customer’s current review behavior on the basis of customer’s own factors and customer purchase behavior, focusing on the effect of customer ’s own review behavior on customer churn behaviors and the improvement of prediction accuracy. Our model conceptual framework can be seen intuitively in Fig. 1.

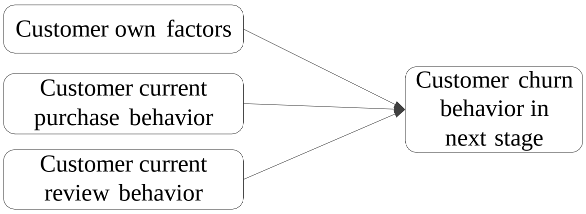

Fig. 1. conceptual framework

Model Building

Basic model

Customer churn model is used to analyze and predict the influence factors of their churn behaviors. The dependent variable of this model is the customer churn behaviors in the next stage while independent variables are customer's own factors, customer purchase and current review behaviors. As a binary variable (churn:1 and unchurn: 0), customer churn behaviors can be analyzed by a Logit regression model using a basic formula:

$$Churn_{\boldsymbol{it}} = \frac{e^{L_{it}}}{1+e^{L_{it}}} \qquad (1)$$

$$\boldsymbol{L_{it}} = \boldsymbol{\beta User_i} + \boldsymbol{\gamma} Purchase_{\boldsymbol{it-1}} + \boldsymbol{\delta} Comment_{\boldsymbol{it-1}} + \boldsymbol{\varepsilon_{it-1}} \qquad (2)$$

Where, $\mathrm{Churn_{it}}$ represents the possibility of customer churn behaviors in the next stage, is also a binary variable; $\mathrm{User_i}$ is the basic information of customer i, such as transaction amount, transaction frequency, etc; $\mathrm{Comment_{it-1}}$ is the current review behaviors of customer i, such as the number of reviews and the cumulative reviews contribution, etc. β, γ, δ are three independent variables, ε_{ijt} is the stochastic error term, representing the potential impact of dependent variable in the model and meeting the assumption of $\varepsilon_{ijt} \sim \mathrm{N}(0, \varphi)$.

Customer churn model without customer's own review behavior

The prediction model in this section just focus on the customer churn behaviors resulting from customer 's own factors and customer purchase behaviors. According to the references about E-commerce customer churn at home and abroad, considering various factors affecting customer churn and explaining the availability of variable data, we determine twelve indicators as selected variables to research in this paper, which includes age, gender, customer contribution value, transaction amount, highest transaction amount, transaction frequency, the time of last transaction, the time of first transaction, numbers of transaction during the work time, day, night and midnight as follows:

$$\begin{aligned} Churn_{it} = {} & \beta_1 age_i + \beta_2 gender_i + \beta_3 Contribution_i + \gamma_1 Gm_{it-1} + \gamma_2 Gm_{top_{it-1}} \\ & + \gamma_3 Gf_{it-1} + \gamma_4 G_{lasttime_{it-1}} + \gamma_5 G_{firsttime_{it-1}} + \gamma_6 G_{worktime_{it-1}} \\ & + \gamma_7 G\_midnight_{it-1} + \gamma_8 G\_day_{it-1} + \gamma_9 G\_night_{it-1} + \varepsilon_{ijt} \end{aligned} \qquad (3)$$

Customer churn model including customer own review behavior

Next, we introduce the customer's own review behaviors, giving a modified prediction model with the consideration of customer's own factors, customer purchase behaviors and customer review behaviors. On the basis of Formula (3), nineteen indicators have been further put forward with the addition of cumulative review contribution value, the highest review contribution value, review frequency, numbers of review during the work time, day, night and midnight as follows:

$$\begin{aligned} Churn_{it} = {} & \beta_1 age_i + \beta_2 gender_i + \beta_3 Contribution_i + \gamma_1 Gm_{it-1} + \gamma_2 Gm\_top_{it-1} + \gamma_3 Gf_{it-1} + \gamma_4 G\_lasttime_{it-1} \\ & + \gamma_5 G\_firsttime_{it-1} + \gamma_6 G\_worktime_{it-1} + \gamma_7 G\_midnight_{it-1} + \gamma_8 G\_day_{it-1} + \gamma_9 G\_night_{it-1} \\ & + \delta_1 Cm_{it-1} + \delta_2 Cm\_top_{it-1} + \delta_3 Cf_{it-1} + \delta_4 C\_worktime_{it-1} + \delta_5 C\_midnight_{it-1} + \delta_6 C\_day_{it-1} \\ & + \delta_4 C\_night_{it-1} + \varepsilon_{ijt} \end{aligned} \qquad (4)$$

EMPIRICAL STUDY

Data Collection

The data used in our work is obtained from a well-known Chinese online review website named "Dian Ping" at Shanghai. "Dian Ping.com" is one of the world's largest online website with a collection of professional reviews, group purchase and other functions, which can provide user's basic information, online reviews and group purchase information, etc. We randomly selected group purchase customers from January 1, 2011 to June 30, 2011 and extracted the data of these customer's purchase behaviors, customer reviews behaviors and demographic characteristics as the independent variables. Meanwhile, we selected the number of group purchase from all above customers from July 1, 2011 to September 30, 2011 and believed that customers without group purchase in this period of time have already churn, denoted as 1;while customers with group purchase in this period of time can be assigned to not churn, denoted as 0. Finally, we obtained 921 users' group purchase data and review data as the independent variable and dependent variable according to Fig. 2:

Fig. 2. Extraction time of independent variable and dependent variable

Variables and Measurement Values

Table 3. Description of model variables

Type	Names	Symbols	Description
Dependent variable	Customer churn	Churn	No purchase：1；Purchase：0, 2011.7.1-2011.9.30
Own factors	Age	Age	Male：1；Female：0
	Gender	Gender	user's age information
	Contribution value	Contribution	user's overall contribution value
Purchase behaviors	Cumulative amount	Gm	cumulative amount: 2011.1.1-2011.6.30
	Largest transaction amount	Gm_top	Largest single transaction amount: 2011.1.1-2011.6.30
	transaction frequency	Gf	cumulative transaction number: 2011.1.1-2011.6.30
	time of last transaction	G_lasttime	Days between the time of last transaction and 2011.6.30
	time of first transaction,	G_firsttime	Days between the time of first transaction and 2011.6.30
	numbers of transactions during the work time	G_worktime	total number of transactions: Monday to Friday & 9 AM - 5 PM
	numbers of transactions at midnight	G_midnight	total number of transactions during 0-6 o'clock
	numbers of transactions during the day	G_day	total number of transactions during 8-18 o'clock
	numbers of transactions during the night	G_night	total number of transactions during 18-24 o'clock
Review behaviors	cumulative review contribution value	Cm	cumulative review contribution value: 2011.1.1-2011.6.30
	the highest review contribution value	Cm_top	the highest single review contribution value: 2011.1.1-2011.6.30
	review frequency	Cf	total number of reviews: 2011.1.1-2011.6.30
	numbers of review during the work time	C_worktime	total number of reviews: Monday to Friday & 9 AM - 5 PM
	numbers of review at midnight	C_midnight	total number of reviews during 0-6 o'clock

	numbers of review during the day	C_day	total number of reviews during 8-18 o'clock
	numbers of review at night	C_night	total number of reviews during 18-24 o'clock

Dependent variable

"Dian Ping.com" can record each user's purchase history, thus the dependent variable of customer churn behaviors is judged by purchase frequency of customers. In detail, we use "Churn" to show the customer churn behaviors in the next stage, no group purchase denoted as 1 while group purchase denoted as 0.

Independent variable

In this paper, the independent variables are divided into three parts, customer's own factors, customer purchase behaviors and customer review behaviors, and model I just includes the first two parts while model II take all three factors into consideration. Based on the demographic characteristic factors from "Dian Ping.com", we believe that customer's own factors are consist of three variables, i.e. gender, age and overall contribution value. Similarly, there are nine variables for customer purchase behaviors, including cumulative transaction amount (Gm), largest transaction amount (Gm_top), transaction frequency (Gf), the time of last transaction (G_lasttime), the time of first transaction (G_firsttime), numbers of transaction during the work time (G_worktime), midnight (G_midnight), day (G_day) and night (G_night). Finally, we also think that there are seven variables for customer review behaviors, such as cumulative review contribution value (Cm), the highest review contribution value (Cm_top), review frequency (Cf), numbers of review during the work time (C_worktime), midnight (C_midnight), day (C_day), and night (C_night).

Data Description

As the description statistics of model variables shown in Table 4, the mean value of dependent variable Churn is 0.347 which means 65.3% of the customers still purchased in the next stage. In addition, the average age of observations is 28 with a standard deviation of 8.181 which also means that the age of sample is generally small. Meanwhile, the mean value of gender is 0.174 which also suggests that the major customers are female.

Table 4 Descriptive statistics of variables

variable	average	standard deviation	Minimum	maximum	observation
Churn	0.347	0.476	0	1	921
Gender	0.174	0.379	0	1	921
Age	28.129	8.181	0	112	921
Contribution	124.619	187.988	-271	3044	921
Gm	330.102	590.035	1	5773.8	921
Gm_top	151.7732	229.059	1	2580	921
Gf	3.664	4.286	1	42	921
G_lasttime	49.142	45.904	0	174	921
G_firsttime	97.701	51.810	0	180	921
G_worktime	1.993	2.687	0	23	921
G_day	2.457	3.057	0	29	921
G_night	0.733	1.388	0	18	921
G_midnight	0.733	1.388	0	18	921
Cm	9.219	31.609	0	415	921
Cm_top	1.195	1.713	0	5	921
Cf	3.483	11.592	0	183	921
C_worktime	0.691	3.798	0	53	921

C_day	1.011	5.398	0	64	921
C_night	0.524	2.949	0	42	921
C_midnight	0.093	1.044	0	27	921

Analysis of Results

Colinearity test

Before building the prediction model, we first carried out the multicollinearity test between the variables, i.e. judging by VIF factor. If VIF> 10, there exists colinearity among variables. As shown in Fig. 3, the VIF factors of transaction frequency (Gf), numbers of transaction during the work time (G_worktime), and day (G_day), cumulative review contribution value (Cm), review frequency (Cf), numbers of review during the work time (C_worktime) are all larger than 10. Thus, there actually exists colinearity among the variables. We adopt the method of stepwise regression for variable processing and selection in order to remove the colinearity among variables and get the key variables of influence factors.

Variable	VIF	1/VIF
gf	48.91	0.020444
g_day	41.97	0.023824
c_day	34.47	0.029013
c_worktime	25.53	0.039173
cm	17.51	0.057103
cf	14.40	0.069460
g_worktime	13.62	0.073434
g_midnight	8.44	0.118515
g_night	6.52	0.153456
c_night	3.69	0.271170
gm	2.51	0.398311
contribution	2.15	0.465960
g_lasttime	1.92	0.521244
g_firsttime	1.89	0.530021
cm_top	1.85	0.541742
mean_comen~e	1.65	0.605008
c_midnight	1.58	0.631007
gender	1.04	0.965061
age	1.03	0.967913
Mean VIF	12.14	

Fig. 3. Analysis of VIF factor

Analysis of model results

At the same time, we divide the overall data set into training set and validation set according to the proportion of 8:2, where 737 users in training set (churn:258, unchurn: 479) and 184 users in validation set (churn:62, unchurn:122).

Table 5. Model result - training set

Churn	Model I	Model II	Model III
Gender	0.5566**(0.2186)	0.5634**(0.2194)	.5320**(0.2217)
Gm_top	0.0010**(0.0004)	.0010**(0.0004)	.0010**(0.0004)
G_lasttime	0.0093***(0.0020)	.0093***(0.0020)	.0089***(0.0020)
Gf	-0.3009***(0.0506)	-.3046***(0.0517)	-0.3406***(0.1468)
Cf		0.0032(0.0083)	-0.0218*(0.0169)
Cf*Gf			0.0045**(0.0089)
_cons	-0.5104***(0.2080)	-.5134***(0.2082)	-0.3817*(0.1757)
obj.	737	737	737
LR chi2(5)	142.64	142.78	147.79
Count R2	0.712	0.707	0.712

AIC	1.115	1.118	1.113

Table 5 gives the analysis results of training set by Logit model, where model I corresponds to Formula 3 and model II, model III are related to Formula 4. By analysis the result of model I, we find that customer gender (β_1=0.5566), largest transaction amount (δ_2=0.001), the time of last transaction (γ_4=0.0093) and transaction frequency (γ_3=-0.3009) will significantly influence the customer churn behaviors in the next stage, if only considering the customer's own factor and current purchase behaviors. Moreover, the largest transaction amount ($\gamma_2 Gm\_top_{it-1}$) has a significant positive impact on customer churn behaviors. The higher the largest transaction amount is, the greater possibility a customer will churn in the next stage. However, transaction frequency has a significant negative impact, which also indicates that higher transaction frequency will result in higher use viscosity to websites and lower customer churn behaviors in the next stage. The time of last transaction also has a significant positive impact with a performance of lager time interval leading to higher customer churn behaviors possibility.

By the comparisons of model II and model III, we find that the introduction of customer review frequency has no significant impact on the customer churn behaviors in the next stage, when considering customer's own factors, current purchase and review behaviors together. However, the effect of customer review frequency on the review platform turns into significant, when we further consider the cross term of transaction frequency and review frequency. This result suggests that the effect of customer review frequency on customer churn behaviors strongly depends on customer transaction frequency in the platform of group-buying. In detail, the synthetic effect (separate influence coefficient + Cross influence coefficient * transaction frequency) of review frequency is -0.005, i.e., the higher review frequency is, the lower customer churn behaviors will happens. In summary, we build the ultimate prediction model as shown in Formula 5:

$$Churn_{it} = \beta_1 gender_i + \gamma_2 Gm\_top_{it-1} + \gamma_2 Gf_{it-1} + \gamma_3 G\_lasttime_{it-1} + \delta_1 Cf_{it-1} + \delta_2 Cf_{it-1} * Gf_{it-1} + \varepsilon_{ijt} \quad (5)$$

Comparison of model prediction accuracy

By far, the most direct way to evaluate the prediction results is to use the quantitative standards for assessment. Here, we adopt the hitting rate of prediction as the standard on the basis of actual conditions. As shown in Table 6, A, B, C and D can be differentiated by churn or un-churn both in fact and prediction, A and D stands for the successful predictions.

Table 6. Prediction evaluation matrix

	Churn in prediction	Un-churn in prediction
Churn in fact	A	B
Un-churn in fact	C	D

Thus the equation of prediction hitting rate can be written as:

$$\text{Hitting rate} = \frac{A}{A+C} \quad (6)$$

Obviously, we find the hitting rates of mode III is superior to those of mode I with respect to both training set and validation set in Table 7.

Table 7. Comparison of hitting rate

	training set				validation set			
	Model I Own + Purchase		Model III Own + Purchase + review		Model I Own + Purchase		Model III Own + Purchase + review	
	Churn	Unchurn	Churn	Unchurn	Churn	Unchurn	Churn	Unchurn
Churn	138	120	140	118	30	32	33	29
Unchurn	99	380	97	382	27	95	29	93
Hitting rate of predicting	58.2%	76%	59.1%	76.4%	52.6%	74.8%	53.2%	76.2%

Ten-fold Cross Validation

Due to the small amount of data, we use the Ten-fold Cross Validation method to further validate model, which is suit for small sample. The concrete method is as follow: firstly, the training set of 258 churn samples and 479 non-churn samples are divided into 10 parts, the first nine parts each contain 74 samples, and the last one contain 71 samples. Secondly, we calibrate the model using the first nine parts samples, and build the logit regression model using the last part sample, which leads to ten models. Thirdly, we apply the ten models into the ten training sets to get the error rate of each model. The overall training set error rate is the average of the above ten error rates. In the same way, we can get the overall error rate of validation sets. The results are reported in table 8.

Table 8. Ten-fold Cross-validation

Error nuber s (error	Training sets		Validation sets	
	Model I Own + Purchase	Model III Own + Purchase + review	Model I Own + Purchase	Model III Own + Purchase + review
(1)	16 （22.5%）	16 （22.5%）	61 （33.2%）	61 （33.2%）
(2)	23 （31.1%）	22 （29.7%）	61 （33.2%）	60 （32.6%）
(3)	17 （23.0%）	15 （20.3%）	63 （34.2%）	61 （33.2%）
(4)	28 （37.8%）	21 （28.4%）	62 （33.7%）	59 （32.1%）
(5)	28 （37.8%）	25 （33.8%）	61 （33.2%）	56 （30.4%）
(6)	21 （28.4%）	16 （22.5%）	63 （34.2%）	54 （29.3%）
(7)	29 （39.2%）	26 （35.1%）	62 （33.7%）	60 （32.6%）
(8)	17 （23.0%）	17 （20.3%）	61 （33.2%）	57 （31.0%）
(9)	18 （24.3%）	21 （28.4%）	54 （29.3%）	57 （31.0%）
(10)	21 （28.4%）	18 （24.3%）	62 （33.7%）	61 （33.2%）
Average error rate	29.55%	26.53%	33.16%	31.86%
Original error rate	29.72%	29.17%	32.07%	31.52%

To sum up, the difference between average error rate and original error rate is very small using the Ten-fold Cross-validation method to analyze the training sets as well as validation sets in Model I & Model III, which shows the stability of our customer churn prediction model.

CONCLUSION

The main purpose of this article is to study the impact of customer’s review behavior on its purchase behavior and try to understand how to utilize this effect to predict customer’s churn behavior in the next stage. Based on the data from Dianping.com, we build the Logit regression model to analyze the effect of review platform on group-buying platform, and use ten-fold cross-validation to prove the stability of our model. The main results are as follows: Firstly, different from the traditional research, we combine the customer’s own review behavior and purchase behavior to build the churn prediction model, considering customer’s own factors, review behavior and the effect of review behavior on its next stage purchase behavior. We discover that there is an effect of customer’s own review behavior on its purchase behavior, which depends on the cross term of purchase frequency and review frequency. The more review times, the less possibilities to churn in the next stage, which demonstrate the effect of review platform on group-buy platform. Secondly, we improve the customer churn prediction model, introducing the customer review frequency on group-buy platform to the prediction model, which promotes the accuracy of prediction.

There are some limitations for this research. Due to the amount of data, we did not consider the effect of public purchase and review behavior on customer's own churn behavior. In the future research, we should further expand the sample size, and add more factors to refine the prediction model for further analysis.

ACKNOWLEDGEMENTS

This paper was supported by the National Natural Science Foundation of China (No.: 71171023, 71231002), Major State Basic Research Development Program of China (973 Program) (No.: 2012CB315805, 2013CB329604), Research Fund for the Doctoral Program of Higher Education of China (No.: 20120005110015).

REFERENCES

[1] Cao, C., Xu, J. & Shen, L. X. (2012) 'The customer churn prediction model in commercial banks based on ERFM and binary logistic methods', *Communication of finance and accounting*, No. 11, pp. 147-149.

[2] Chevalier, J. & Mayzlin, D. (2006) 'The effect of word of mouth on sales: online book reviews', *Journal of Marketing Research*, Vol. 43, No. 3, pp. 345-354.

[3] Joseph, B., Jake, L. & Tina G. (2012) 'Online reviews, helpfulness ratings, and consumer attitudes: An extension of congruity theory to multiple sources in Web2.0', *Journal of Computer Mediated Communication*, Vol. 18, No. 1, pp. 97-112.

[4] Kellogg, D. L., Youngdahl, W. E. & Bowen, D. E. (1997) 'On the relationship between customer participation and satisfaction: two frameworks', *International Journal of Service*, Vol. 8, No. 3, pp. 206-219.

[5] Lloyd, A. E. (2003) 'The role of culture on customer participation in service', *dissertation*, Hong Kong Polytechnic University, HK.

[6] Ren, J. F. & Zhang, X.Y. (2012) 'Research on e-business customer churning modeling and prediction', *Computer simulation*, Vol. 29, No. 5, pp. 363-366.

[7] Wang, Z. J. (2013) 'Research on customer churn early-warning based on neural network', *dissertation*, Jilin University, Jilin.

[8] Xiao, X. (2006) 'Research on the methods in customer lose and customer value for mobile telecom enterprise', *dissertation*, Tianjin University, Tianjin.

[9] Zhai, L. K. (2011) 'Research on the effect of online consumer review on purchasing intention', *dissertation*, Northeast University of Finance, Dalian.

[10] Zhu, B. Z. & Zhang, Q. J. (2010) 'E-business customer churn prediction based on three-step integration of SMC and least square support vector machine', *Chinese soft science*, No. 6, pp. 186-192.