

# Users shopping behavior mining of B2C e-commerce

Gan Ping

Department of Finance and Economics, Guangxi University of science and technology, China

**Keywords:** B2C e-commerce; Users shopping behavior; Data mining; Basket model

**Abstract.** Rich transaction information and customer data can be recorded by the back-end database of e-commerce sites. By the application of the basket model in data mining analysis the correlation between data and predict customer loyalty. Then combined with the customer's shopping behavior, sales promotions pattern based prediction model is proposed and implemented. The research results will enable the B2C e-commerce to make the appropriate commercial promotion, and to improve the profitability of the company in the B2C mode by customer behavior, personal preferences.

## Introduction

Today, e-commerce is a new business field, a lot of people in the industry are still busy with the technical considerations, rarely analyze the buying behavior of the customer network. Ultimately, the "customer" is the company's core. What customers of the customer base is that online shopping? What characteristics of their web shopping behavior have? What differs are from their traditional business customers? These questions' answer has a vital role to the commercial success [1]. The data customers on shopping processes and outcomes are demonstrated very useful to merchants by various studies [2]. The order data and on online shopping other data are all very important to merchants. Between them there are difference that is final shopping decisions are made by the former while the customer behavior that making shopping decisions are provided by the latter information [3]. In addition to this, one of the main difference is that, browsing and shopping cart data information embedded in the latter more closely related to the last shopping decision [4]. Data on customer browsing behavior have been discussed in many studies, such as find the path travel mode, the website flow, the product is the most visited, the hot areas of a web page, the customer pro file based on these browsing data. However, there is little research data involves customers shopping cart [5]. In this paper, by the application of the basket model in data mining analysis the correlation between data and predict customer loyalty through the analyze of user's shopping behavior in B2C e-commerce. The research results will enable the B2C e-commerce to make the appropriate commercial promotion, and to improve the profitability of the company in the B2C mode by customer behavior, personal preferences. At the same time, the research method can to provide reference for other e-commerce companies to analyze customer data.

## Description of user shopping behavior

In the online shopping behavior data collection is one of the methods to understand a customer are interested in every step and the possible idea fundamental this process. Traditionally, it is only through the POS collection of data is the most important data, reflecting the final shopping decision, not in front of the Settings in Fig. 1. Although know consumers have already bought is useful marketing personnel, this does not let them learn why consumers make a particular purchase decisions, not bought some other projects.

In the Internet age, it is relatively easy to collect data related to the decision making process such as browse and shopping cart data. Browse data may provide information that thinking process from consciousness set to consider setting and selection, and shopping cart data can provide more information in the process of thinking from consideration set to choice set, and then make the final decision. It should be noted that the shopping cart data may be different from order data, because the shopping cart items can be changed at any time made before purchase.

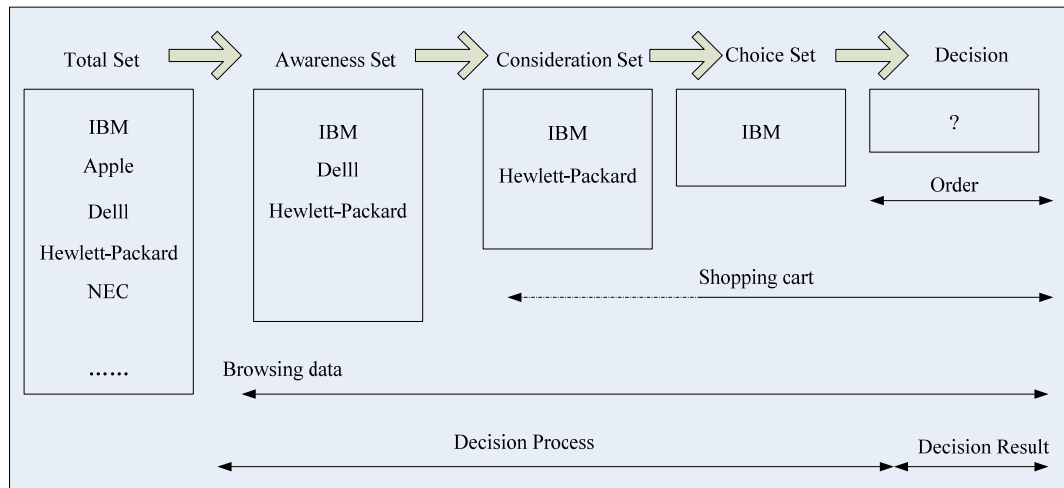


Fig. 1 The relationship between the continuous set in consumer decision-making and their online behavior

## Data mining

### Basket Analysis

In recent years, basket model in statistical theory has wide application with the development of the theory of data mining. Its main purpose is to study the purchase behavior of consumer in the purchase goods, which goods are more likely to be buy at the same time. There are two main statistical indicators: the degree of support (support), confidence (confidence). Assuming that, we have N-consumer purchase records to the study the N consumers ( $C_1, \dots, C_N$ ) to buy M kinds of commodities ( $P_1, \dots, P_M$ ) on the preferences. The degree of support is a basic indicator that measure the popularity of a mix product The support definition of products combination  $P_1$  and  $P_2$  is as  $(100 * N_{12} / N)\%$ , where  $N_{12}$  indicate the number of consumers in the N consumers that purchase the product of  $P_1$  and  $P_2$ . Confidence is an indicator to measure two commodities' associated degrees. It means that the probability of consumers buy some others goods after buying a certain commodity. Suppose that there are  $N_1$  customers that buy products the  $P_1$ . There are  $N_2$  customers that buy products  $P_2$  among the  $N_1$  customers. Then the degree of confidence  $P_1 \rightarrow P_2$  is  $(100 * N_2 / N_1)\%$ . Noted that confidence direction,  $P_2 \rightarrow P_1$  confidence will be different from the  $P_1 \rightarrow P_2$  confidence. Gazelle.com has hundreds of sale goods that of different brands such as DKNY, HOSO and so on. Further, we analysis the purchase data by the applications of shopping basket model according to the 1804 commodities data from customers. Table 1 shows 10 groups of mix product with the number combinations of 2 and 3, both a high degree of support, and a relatively high degree of confidence. These results have important reference value on how to develop a business promotion strategy. For example, aim to the first commodity mix ELT- of commodities  $D \rightarrow$  ELT- commodities E, their support is 60.61% and confidence level is 25%. In other words, 60.61% of the customers also bought two commodities, and in the purchase of the "ELT-commodity D" customers, 25% of customers buy the commodity "ELT - commodity E" According to this result recommended Gazelle.com when customers buy the "ELT-commodity E, recommend can make to customers to buy" ELT-commodity D "in the future.

Another example, aim to the sixth commodity mix ORO- of commodities  $O \rightarrow$  DON-commodities B and DON- commodities C, their support is 12.11% and confidence level is 84.44%. In other words, 12.11% of the customers also bought these three commodities, and in the purchase of the ORO- of commodities  $O \rightarrow$  DON- commodities B customers, 84.44% of customers buy the commodity DON- commodities C. According to this result recommended Gazelle.com when customers buy the ORO- of commodities  $O \rightarrow$  DON- commodities B, recommend can make to

customers to buy DON- commodities C or give as a present in the future which will receive good business.

Table 1 The results table of basket model analysis

Number	Combination number	Commodity combination	Support degree	Confidence degree
1	2	ELT Commodity D→ELT Commodity E	60.61	25.00
2	2	NMW-Commodity M→NMW-Commodity N	38.57	25.36
3	2	HAUS Commodity G→HHT-Commodity F	33.06	30.00
4	2	HHT-Commodity F→HAUS-Commodity G	33.06	24.00
5	2	HOSO-Commodity H→HOSO-Commodity I	27.55	50.17
6	3	ORO-Commodity ODON-Commodity B→DON-Commodity C	11.02	83.56
7	3	HPK-DON Commodity B→ORO-Commodity -ODON-Commodity C	11.02	22.22
8	3	Commodity J→HPK Commodity LHPK-Commodity K	10.87	66.67
9	3	HPK-Commodity L→HPK Commodity K→HPK-Commodity J	10.87	28.57
10	3	DKNY-Commodity A→NMW-Commodity -NELT-Commodity E	8.63	35.00

### Forecast modeling

The argument of the prediction model is customers' personal characteristics, as well as the behavior variables of web browsing and purchase. Figure 2 shows the flow chart of the application of enterprise miner to build predictive models. The first module is data reading, the second and third modules is the data preprocessing, and the next paratactic three modules is make predictions to data by application "regression", "decision trees and neural network model, and the finally module draws the lift chart by comparing the performance of the three models. The lift chart drawing method is as follows: (1) Predictive model response variable is an argument 0,1 binary variable  $y$ , and model of each customer is given a probabilistic prediction value of  $y = 1$ , the predicted value is called the client score. (2) Sort customers according to the customer scores in accordance with this model, then the customers were divided into 10 groups, each group were calculated the actual number percentage that the value  $Y = 1$ , and then cumulative percentage to make enhance table.

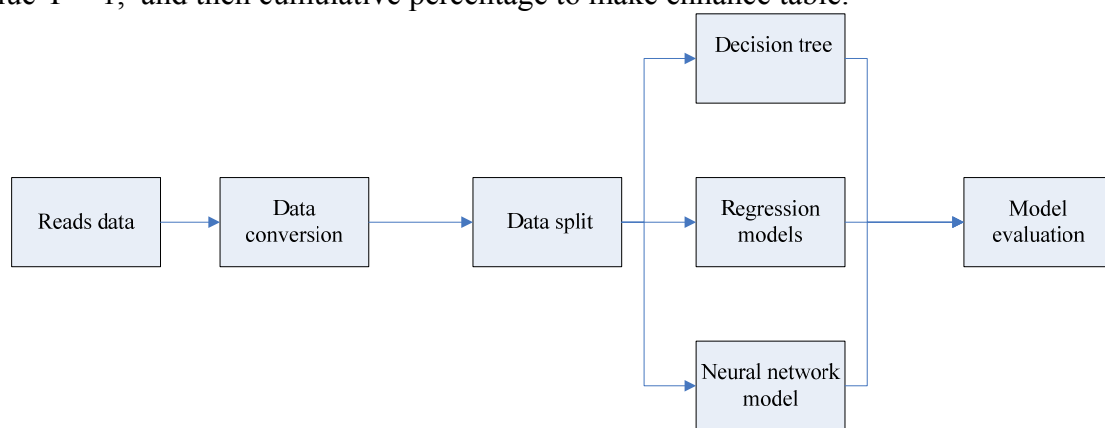


Fig. 2 SAS enterprise miner predictive model flow chart

In accordance with the predicted results of the regression model, select the first group of 200 customers, which will be 140 (70%) of customers' shopping value excess of \$ 12. The accuracy promoted 2.64 (70/26.5) times than randomly selected 200 people without any model. Gazelle.com can put the existing customers' corresponding argument into the substituted model to predict the probability that customers buy the products of more than 12 U.S. dollars. According to the predictive value promote marketing promotion to the targeted customers with higher probability.

### Users shopping behavior mining of B2C e-commerce

In the field test, publishers provides 6 months, bind strategy is based on the manager's professional knowledge. In this period, a mean 2.7 customers, orders, and nine book 2.8 monthly sales. Nine bound strategy be promoted for 1 month in our field test, in this period, there are 22 customers, 33 orders, and sold 78 copies, it shows that the performance has improved remarkably. We choose to buy the number of books is measurement variable; that is to say, more books were purchased because

discount bound strategy shows that a better performance. Mean value and standard deviation value to buy the number of books for each method are listed in Table 2. Two-way analysis of variance is used to stop me whether performance with two browsing and shopping cart data to determine product bundling is still far higher than other two strategies, the application of a type of data only. Results of summing up are in Table 3.

Table 2 Mean and standard deviation of the purchase discount bound strategy

Deffirent Strategy	The mean	The standard deviation
Based on order data	0.812	1.301
Based on browse data	0.601	1.010
Based on two browsing and shopping cart data	2.214	2.689

Table 3 The results of variance analysis for discount two-way binding strategy

Source	Sum of squares	DF	Mean square	F	p
Users	46.05	22	2.24	0.57	0.930
Methods	31.45	3	15.34	3.80	0.030
Error	171.00	32	4.15		
Total	248.13	57			

In this experiment, any product can be purchased by every customer to bundles extracted from the above model. Therefore, between the user and the different strategies have variable source included in different preferences. The results show that there is no significant by the user's preferences in the of the discounted bundles (  $F = 0.57, p = 0.930$ ) but three product bundling strategies gave significantly affected to the purchasing (  $F = 3.80, p = 0.030$ ). Next, in order to compare means that each pair of strategy we use male multiple comparison test estimated 95% confidence interval and different strategies. From simulation result, show that, the best strategy is based on two browsing and shopping cart data only significantly outperform the strategy based on order data and strategy only based on browse data. And through analysis we know that there are little significantly differ in the performance only between the strategy based on order data and strategy only based on browse data.

## Conclusion

In this paper, we analyze the user's personal characteristics, online shopping behavior of B2C e-commerce. The main conclusion is that: most of the clients are women, middle-aged customers (31 to 40 years old) is the highest proportion of all ages, married proportion higher than unmarried. In the two-month study period, their consumption of online shopping average is \$ 14.97. Female consumers expense higher than the male, and consumer with child is higher the clients than has no child. By the application of the basket model in data mining analysis the correlation between data and predict customer loyalty. Then combined with the customer's shopping behavior, sales promotions pattern based prediction model is proposed and implemented. The research results will enable the B2C e-commerce to make the appropriate commercial promotion, and to improve the profitability of the company in the B2C mode by customer behavior, personal preferences.

## References

- [1] Jin, D.et.al. Winning in emerging-market cities[M]. A guide to the world's largest growth opportunity. 2011. Boston, MA.
- [2] D.L. Hoffman et.al. Marketing in hypermedia computer-mediated environments: conceptual foundations[J]. Journal of Marketing, 1996, 60 (3): 50–68.
- [3] Bhattacharyya, S., Jha, S.. Data mining for credit card fraud: A comparative study[J]. Decision Support Systems, 2010, 50, 602–613.
- [4] J. Balla, G. Desai. Understanding e-commerce[J]. Information, 1999, 13(8): 28–30.
- [5] H. Lai, T.-C. Yang, Applying shopping cart data to web customers clustering, in: Proceedings of the Western Decision Sciences, Vancouver, Canada, 2001.