2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery

Predicting User Behavior in E-Commerce Based on Psychology Model

Lei Shen, Yiming Zhou, Chao Xu, Xia Hu, Biyun Hu
School of Computer Science and Engineering
Beihang University
Beijing, China, 100191
lorashen@126.com, zhouyiming@buaa.edu.cn,
chaoxu@263.net, huxia0426@126.com, byhu8210@yahoo.com.cn

Abstract

Web sites are making great effort to understand the user's behavior in order to make the web sites easy to use and further increase their profits. This paper presents a method to predict the user's buying behavior based on psychology model. We employ the method to analyze online store data and treat the clicking and buying as user's attitude and behavior. Then, a new model, that is used to predict the user's future buying behavior, is built based on the attitude-behavior relationship theory. We then verify the model and simultaneously estimate its parameters by path analysis. Our method is evaluated by comparing with traditional naive bayes classification algorithm. Experiments results show that our model is more effective in predicting buying behavior and finding out users who are more profitable to web sites.

1. Introduction

The web sites are making great effort to understand user's behavior and make the web sites easy to use. To achieve this goal, researchers proposed lots of approaches to use web usage data.

Researchers studied this topic from different points of view. Some researchers studied the psychology model that may simulate the consumers' thought. Mostly they use the questionnaire to collect data and then verify the model, such as [1]. Economists concerned more about statistical methods. Poel and Buckinx[2] investigated the contribution of different types of predictors to the online purchasing behavior, including clickstream data, customer demographics, and customers' historical purchase behavior. After selecting the best subset of predictors, they used logit model to predict users' purchasing behavior during their next visit to the website. To find out useful patterns on the web, computer scientists proposed more advanced data mining methods and algorithms to process web usage data, such as association rules, sequential pattern discovery, clustering, and classification. Eirinaki[3] introduced this kind of work.

This paper proposes a new approach to predict user's online behavior. First, we build a new model based on the attitude behavior relationship theory. Second, we use our model to analyse the website real time data, and predict which product the user will probably buy. Through this method, we hope that we can better understand the user's behavior, which is especially useful for web sites trying to be user friendly.

The rest of the paper is organized as follows: Section 2 describes the attitude-behavior relationship theory. Section 3 depicts the proposed user behavior predicting model. Section 4 gives the experiment methodology and the results. Section 5 concludes the paper.

2. Attitude-Behavior Relationship Theory

Social psychologists have been investigating the attitude-behavior relationship for many decades. They studied many factors affecting attitude-behavior relationship. We use the model proposed by [4], in which 29 research reports were analyzed, involving 41 studies and 128 study conditions. These researches first took psychological experiments on subjects then analyzed how the factors might affect the attitude-behavior relationship.

Glasman and Albarracin[4] concludes that there are mainly six observed factors listed below affecting the attitude-behavior correlation.

1. Motivation,
2. Repeated expression or report of attitudes,
3. Direct behavioral experience,
4. Confidence,
5. Behavior relevance of attitudes,
6. One-sidedness of information.

978-0-7695-3735-1/09 $25.00 © 2009 IEEE
DOI 10.1109/FSKD.2009.515

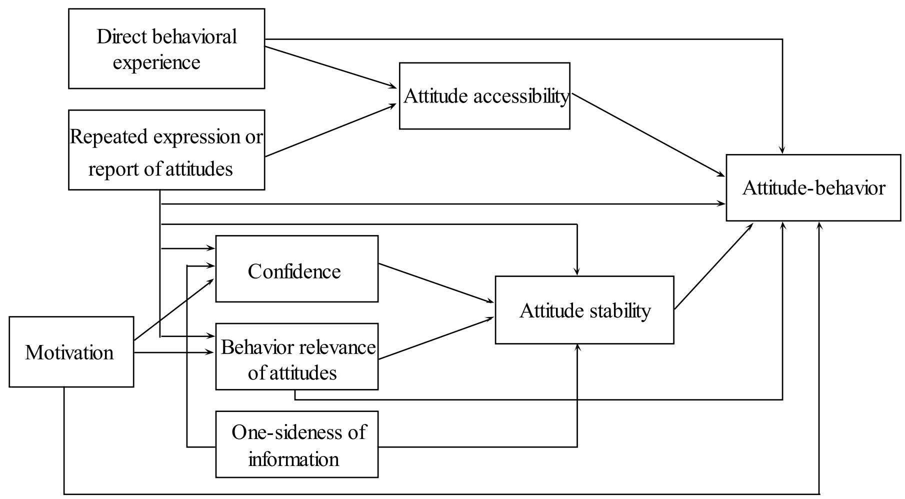

Figure 1. Attitude-behavior .

The attitude-behavior relationship model is shown in Figure 1 (attitude accessibility and attitude stability are latent factors).

From Figure 1, we get the following relationships: direct behavioral experience and repeated expression or report of attitudes are associated with attitude accessibility. Repeated expression or report of attitudes, motivation, and one-sidedness of information have influences on confidence. Repeated expression or report of attitudes, confidence, behavior relevance of attitudes and one-sidedness of information, have affects upon attitude stability. Besides direct behavioral experience, repeated expression or report of attitudes, motivation, greater attitude accessibility and attitude stability are also associated with greater attitude-behavior correspondence.

We can formulate the relationship as follows:

$$X_3 = \alpha_{31}X_1 + \alpha_{32}X_2 + \epsilon_3 \quad (1)$$
$$X_5 = \alpha_{51}X_1 + \alpha_{57}X_7 + \epsilon_5 \quad (2)$$
$$X_6 = \alpha_{61}X_1 + \alpha_{64}X_4 + \epsilon_6 \quad (3)$$
$$X_8 = \alpha_{81}X_1 + \alpha_{85}X_5 + \alpha_{86}X_6 + \alpha_{87}X_7 + \epsilon_8 \quad (4)$$
$$X_9 = \alpha_{91}X_1 + \alpha_{92}X_2 + \alpha_{93}X_3 + \alpha_{94}X_4 + \alpha_{97}X_7 + \alpha_{98}X_8 + \epsilon_9 \quad (5)$$

Where X_1 is repeated expression or report of attitudes, X_2 direct behavioral experience, X_3 attitude accessibility, X_4 motivation, X_5 confidence, X_6 behavior relevance of attitudes, X_7 one-sideness of information, X_8 attitude stability, X_9 attitude-behavior.

In [4], motivation denotes the feature arousing the participants toward engaging in the behavior and giving the purpose and direction to the behavior. Repeated expression or report of attitudes is the number of times that participants expressed their attitudes to the object before the behavior. Direct behavioral experience refers to whether the participants had previous experience with the object. Confidence refers to how confident the participants held the attitude to the object. Behavior relevance of attitudes denotes to the correlation between the attitudes the participants held and the behaviors they engaged in later. The meaning of one-sidedness of information is that the received information is biased, i.e., favorable information. Attitude accessibility evaluates how easily the attitude can be accessed, and attitude stability refers to the difficulty degree to change the attitude. These two factors are unobservable factors.

3. The New User Behavior Predicting Model

The major work of this paper is to build a new model to predict the user's buying behavior. We take the e-Commerce web site as an example. We regard a user's browsing a product as having an attitude towards it, buying it as having a behavior. Then we study the relation between the user's browsing and buying in the same session(the time during the user logs in). We apply the attitude-behavior relationship theory in predicting the user's future behavior. In order to achieve this, we find the variables representing the factors in attitude-behavior relationship. The variables are described as follows.

If a user browsed a product page, we consider that the user has motivation to buy it. So we regard browsing as motivation. As the data is from click data, we set browsing as constant and do not take it into account.

Viewing the product shows the user's attitude to it. Viewing it more than once means the user expresses his attitude to it repeatedly. So we regard the number of times that the user views the product as repeated expression of

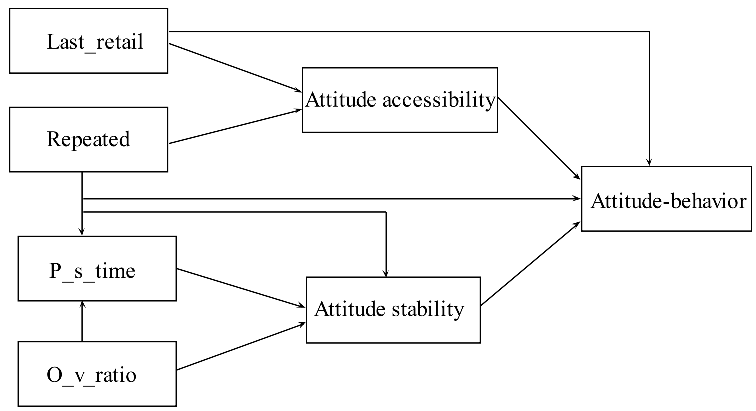

Figure 2. The new behavior predicting model.

attitudes. We name the variable as repeated.

Confidence varies with the absolute amount of information available and with the perceived strength of the evidence supporting the judgment, in other words, attitude confidence ought to vary with the quantity of the information and the quality of the information on which the attitude is based[5]. We consider that the user gets the product's information from browsing the product page. If a user views the product page for a long time, it means that the user gets enough information to hold high confidence. So we regard the time of viewing the product page as confidence. We name the variable p_s_time.

If a user logs in and browses the product, we consider the user's attitude toward the product is relevant to the buying behavior. So we see the behavior relevance of attitudes also as constant and do not take it into account.

The favorable comments to the product can be seen as one-sided information. So we regard the ratio of the number of times buying the product to that of viewing it as one-sideness of information. We name the variable o_v_ratio.

If a user has bought the product before, we consider the user has direct behavioral experience with the product. So we take the previous buying record as direct behavioral experience. We name the variable last_retail.

Finally, we check the buying record of the product in the session. If after the browsing session, the user buys the product, i.e. there is a buying record in the order data in the same session, we perceive the user has the behavior.

After finding the correspondence between the web variables and the attitude-behavior relationship factors, we build our model, see Figure 2. During the sessions, there are six variables affecting the user's buying behavior, and we can describe the model as follows:

$$Y_0 = \beta_{01}Y_1 + \beta_{02}Y_2 + \beta_{03}Y_3 + \beta_{05}Y_5 + \delta_0 \quad (6)$$

$$Y_1 = \beta_{12}Y_2 + \beta_{13}Y_3 + \delta_1 \quad (7)$$

$$Y_4 = \beta_{43}Y_3 + \beta_{46}Y_6 + \delta_4 \quad (8)$$

$$Y_5 = \beta_{53}Y_3 + \beta_{54}Y_4 + \beta_{56}Y_6 + \delta_5 \quad (9)$$

Where Y_1 is latent factor attitude accessibility, Y_2 the previous record the user bought the product(Last_retail), Y_3 how many times the user repeated browsing the product(Repeated), Y_4 the time the user spent browsing the product(P_s_time), Y_5 latent factor attitude stability, Y_6 the ratio of the users ordering and browsing the product(O_v_ratio), Y_0 the buying record of the product in the session (User's buying behavior).

4. Experiments

In this section, we provide experimental evidence to verify the effectiveness of our model. We will introduce our experimental methodology and the results.

4.1. Data Preprocessing

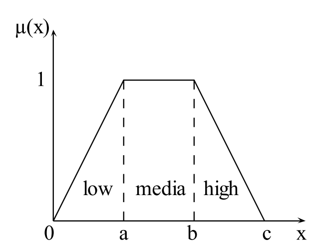

Figure 3. Membership function.

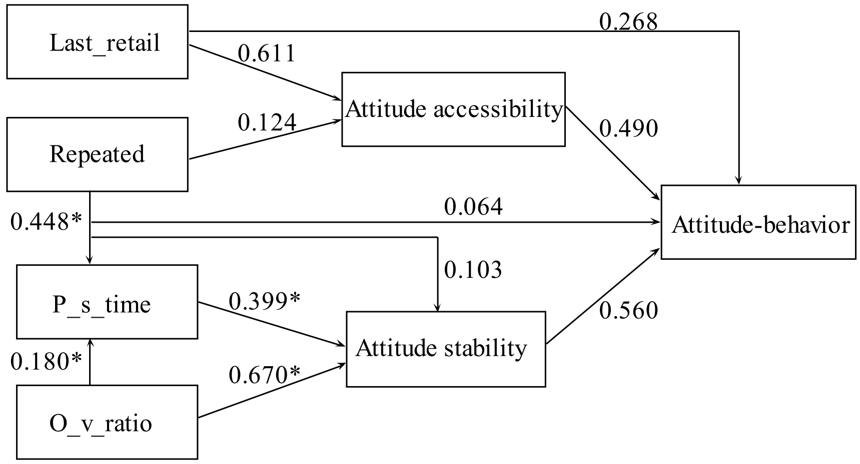

Figure 4. Path analysis. $*p < .05$.

We used the KDD CUP 2000 data. The data is from Gazelle.com, a legwear and legcare web retailer that closed their online store on 8/18/2000. The data records the information of the users' clicks and orders. More details about the data can be found in [6]. We only used the log-in users' clicking and ordering data.

We preprocessed the data to get the variables' values in our model. After assigning values for the observed variables, we normalized them between 0 and 1. As the meanings that the variables represent are fuzzy, we processed the variables as fuzzy sets and assigned a membership function for each. We used the fuzzy statistical method[7] to obtain the membership function curve. Approximately the curve is trapezoids as shown in Figure 3, so we assigned the membership function as below. Then we randomly divided the data into two sets, one for training and the other for testing. Each set has 5665 samples.

$$\mu(x) = \begin{cases} x/a, & 0 < x \leq a, \quad (10a) \\ 1, & a < x < b, \quad (10b) \\ (c-x)/(c-b), & b \leq x \leq c. \quad (10c) \end{cases}$$

4.2. Model Verification

The model we built in section 3 can be seen as a casual model, so we adopted a path analysis on the training data. Path analysis is a popular statistical analysis to verify casual models. It can estimate the magnitude and significance of hypothesized causal connections between sets of variables[8]. So this analysis allowed us to examine whether the four observed variables have effect on each other and whether they directly or indirectly influence user's buying behavior.

Through path analysis, we get the coefficients of our model as shown in Figure 4. Correlations between independent variables are as follows: repeated and o_v_ratio, $r = -.003, p < .001$, o_v_ratio and last_retail, $r = .001, p < .001$, repeated and last_retail, $r = .000, p < .001$. Fit measures of this model are: $\chi^2(1, N = 5665) = 12.415, p < .001$, normed fit index = .991, comparative fit index = .992, incremental fit index = .992, root-mean-square residual = .05. The chi-square indicates a good fit when the associated significance value is lower than .001. According to [9], the normed fit index, the comparative fit index, and the incremental fit index reflect good fit when they exceed .90, and the root-mean-square residual represents adequate fit when it is equal to or less than .10.

So we get that: repeated is highly related to p_s_time; o_v_ratio is also related to p_s_time; repeated and last_retail both proved directly associated with user's buying behavior; and repeated, last_retail, p_s_time and o_v_ratio proved indirectly associated with user's buying behavior through attitude accessibility or attitude stability. We can further conclude that our model is supported by path analysis.

4.3. User Behavior Prediction

We labeled the actual data by two categories. Category 0 means that the user didn't buy the product(non-buying) and category 1 means that the user bought the product(buying). As mentioned in Section 4.2, we also get the coefficients of the variables in our model through path analysis. Then we used the coefficients to calculate the values of Equations(6-9) on the test set and classify the users in the test set. We compared our method with naive bayes classification algorithm[10] and the experiments results are given in Table 1.

Although the overall precision of naive bayes is higher than our method, its performance for predicting buying behavior which is actually more important is really poor. The problem may originate in the feature of the data: the sample number of non-buying is three times as big as that of buying. And that makes sense for e-Commerce web: users browse, but only a limited number of them buy. Our method resolved the problem that naive bayes had. Our method obtains much higher predicting precision for buying behavior than naive bayes, while maintaining good precision for non-buying. We can see our method is very promising.

Table 1. User Behavior Predicting Results.

	Our Method	Naive Bayes
Overall Precision	0.6849	0.7541
Precision for buying	0.6231	0.2468
Precision for non-buying	0.7056	0.9241

5. Conclusion

In this paper, we propose a user behavior predicting model based on attitude-behavior relationship theory. We use path analysis to verify our model and the result demonstrates that the variables we extract from the e-Commence web site data do have relationship with the users buying behavior. Experiments results show that our model is better than naive bayes in predicting the users buying behavior, that is, our model has a better understanding of the users buying behavior.

Our method can be used in many cases. Using our method, we can better understand the e-Commerce users, so as to improve the recommendations. In future, we will try to apply it to the recommendation system.

Acknowledgement

The work is supported by the National Basic Research Program of China under Grant No.2007CB310803.

References

[1] Angel H. Crespo, Ignacio A. Rodriguez Del Bosque Rodriguez, ”Explaining B2C E-commerce Acceptance: An Integrative Model based on The Framework by Gatignon and Robertson”, *Interacing with Computers*, 2008, 20, pp. 212-224.

[2] Dirk Van den Poel, Wouter Buckinx, ”Predicting online-purchasing behaviour”, *European Journal of Operational Research*, 2005, 166, pp. 557-575.

[3] Magdalini Eirinaki, ”Web Mining: A Roadmap”, *Technical Report IST/NEMIS*, 2004.

[4] Laura R. Glasman, Dolores Albarracin, Forming Attitudes That Predict Future Behavior: A Mete-Analysis of the Attitude-Behavior Relation, *Psychological Bulletin*, 2006, Volume 132, Number 5, pp. 778-822.

[5] Ida E. Berger, ”The Nature of Attitude Accessibility and Attitude Confidence: A Triangulated Experiment”, *Journal of Consumer Psychology*, 1992, Volume 1, Number 2, pp.103-123.

[6] http://www.ecn.purdue.edu/KDDCUP/.

[7] Sivanandam S. N., Sumathi S. and Deepa S. N., *Introduction to Fuzzy Logic using MATLAB*, Springer Berlin Heidelberg, 2007.

[8] http://people.exeter.ac.uk/SEGLea/multvar2/pathanal.html.

[9] Peter M. Bentler, *EQS structural equations program manual*, Encino, CA: Multivariate Software, 1995.

[10] Harry Zhang, ”The optimality of naive Bayes”, *In Proceedings of the 17th International FLAIRS Conference*, AAAI Press, 2004.