2012 International Conference on Information and Computer Applications (ICICA 2012)
IPCSIT vol. 24 (2012) © (2012) IACSIT Press, Singapore

Prediction of Customer Purchase Probablity for Online Recommendation Systems

Song-Yi Han and Jong Woo Kim +

School of Business, Hanyang University, Seoul, Korea

Abstract. Tracking of online customers' behavior is essential part in online recommendation systems. Specifically, for online recommendation at online stores, it is necessary to understanding customers' purpose through their real-time activity data. This study suggests state probability methods that predict customers' purchase probability using their clickstream data. Also, it verifies usefulness of the proposed method by using real clickstream data of an online book store. From experimental results, state probability models show better performance. Also, 2-state model with weight show best performance among proposed methods

Keywords: Purchase probability, Clickstream data, Online Recommendation System

1. Introduction

The advance of Internet make it possible to overcome time and space restraints, and it became a background of the new formation of business called e-commerce. The vitalization of e-commerce augmented the importance of an immediate response about customers' needs. Therefore, studies on sorting out potential customers by analyzing in individual level rather than popular level or studies such as recommendation service for decision making for purchase are vitalizing[8,1].

Unlike existing recommendation service, online recommendation systems is focused on enhancing outcome of recommendation by using customers' real time data. Thus, for online recommendation systems, it is required to track customers' behavior online and recommendation activity based on that. However, existing recommendation technologies make recommendation according to past data rather than real time data. Therefore, this study is focused on understanding customers' intention to purchase through real time activity data at online storefronts as the part of the online recommendation systems. The study aims to suggest methods that predict customers' purchase probability by using their clickstream data. Also, the usefulness of the proposed methods is verified by using real clickstream data of an online bookstore.

2. Session Path

Based on clickstream data, we can generate session paths. A session path is a record of a user's search movement route and used to know user's movement in an Internet storefront [1,10]. Because it is a tracking of a user's route, it can be used to determine the purpose of users' visit or to predict next movement of users. In this study, chilckstream data of a domestic online bookstore is collected through a company that collects website visit data of a panel group.

3. Online Recommendation System

Recommending some products to a customer is an important feature of Internet storefronts that has to consider current trend of sales, customer's preference and emotion, and the ability to pay[3]. Since there

---

+ Corresponding author. Tel.: +82-2-2220-1067; fax: +82-2-2220-1169
E-mail address: kjw@hanyang.ac.kr
This work was supported by Seoul Industry-Academy-Research Collaboration Project (JP100080) funded by Seoul City.

exists various requirements to consider of online recommendation systems, there is diverse researches to solve the issue from different perspectives.

The representative recommendation technologies are collaborative filtering, content-based recommendation, and rule-based recommendation[5,7]. The collaborative filtering is a way to predict a user's interest automatically through other users' information whose interests are similar to the user. The content-based recommendation is a way to recommend products which have higher preference by comparing user's preference information and each attribute data of products[7,2]. The rule-based recommendation makes related rules and uses it for the recommendation service, and find relation of purchased product through visit path analysis and shopping basket analysis

4. Analysis Data

The pages of an online store are classified with several types of page. The 11 types of pages are classified in the study like Table 1. Because clickstream data has not specific session classification, we need to define session classification rules. In this study, one session is consists of page visits within 20 minutes. For example, Table 2 shows that information about the 5 sessions. That is, in the case of the first session, the customer visit homepage at first, search pages for 2 times, and product pages for 3 times after that.

Tab. 1: Cutting parameters of simulation milling

SORTING	EXPLANATION
Home	Website main page
Account	Personal information management and information check
Search	Search of products
Classification	Classification page according to standard of classification
Product	Introduction of product and information
Information	Information for use and customer service center
Promotion	Pages for promotion
Social	Social space
Storage	Storage of product
Cart	Shopping cart
Order	Order

Tab. 2: Cutting parameters of simulation milling

SESSION	PAGE
1	HO SE SE PT PT PT CA SE PT SE PT SE SC
2	HO SE SE PT PT PT PT SE SE SE SE PT
3	OT HO PR SE SE PT HO
4	HO PR AC SC SC SC SO SO HO SE SE PT SE PT PT SC HO AC SE PT SC SE SE SE SC PT SC
5	PT CA CA CA SC CA CA CA SC AC AC AC AC AC AC AC OD

Home=HO; Account = AC; Search = SE; Category = CA; Product = PT; Information = IN; Promotion = PR; Social = SO; Shopping Cart = SC; Order = OD; Other sites = OT.

5. Purchase Probability Prediction Methods

This paper proposes 2 methods to predict customer's purchase probability in real time at each session as follows.

Method 1 : State Model

- **1-State Model**
  In this method, the purchase probability of each page type is calculated from clickstream data. The purchase probability of a customer is predicted as the average purchase probability of the page types where she/he has visited. Since it does not consider state transition, it is called as the 1-State Model.
- **2-State Model**

In this method, state transition is considered to get the purchase probability. That is, the purchase probability of state transition is calculated from clicksstream data. The purchase probability of a customer is predicted as the average purchase probability of last state transition, which is called as 2-State Model.

Method 2: State with Weight Model

- **1-State with Weight Model**
  In this method, the purchase probability of each page type is calculated from clickstream data. The purchase probability of a customer is predicted as the average purchase probability of the page types where she/he has visited. Since it does not consider state transition, it is called as the 1-State Model.
- **2-State with Weight Model**
  This is similar to 2-State Model with similar modification of 1-State with Weight Model.

6. Experimental Results

The experiments are performed with real clickstream data of an Internet bookstore for 6 months. All sessions extracted from clikstream data are 1842. The sessions that include only one page visit are excluded in the experiments. The processed session paths are divided randomly to training data set and test data set. The training data set and test data set have the same size. The study gets the purchase probability as page type and the purchase probability at state transition by using the training data set, and compares the outcome of the 2 prediction methods.

Figure 1 shows purchase probability prediction of two sessions (one is a purchase session and the other is non-purchase session) using 2 proposed methods. In Figure 1, (a) and (b) shows results to apply State Model, in which we can find 2-State model shows better performance than 1-State model. In Figure 1, (c) and (d) shows results to apply State with Weight Model, in which 2-State With Weight Model shows better performance than 1-State With Weight Model. The results imply that state transition probabilities of pages are better indicators than probabilities of page types. Also, weighted models show better performance than no-weighted models.

7. Conclusion

This study suggests four methods to predict customers' purchase probability through customers' clickstream data analysis in real time at online store. In the first method, a customers' purchase probability is calculated by averaging purchase probabilities of web page types which she/he had visited. In the second method, it is calculated by average purchase probabilities of web page type transitions which she/he had visited. The third and fourth methods use weighted average as extensions of the first and second method, respectively. In the experimental results using real Internet bookstore clickstream data, the cases considering state transition show better performances than those of considering state probability. Also, weighted averaging methods show better performance than without weighting. As further research topics, we need to develop state transition models with more than 2 steps, and to develop methods to know customers' purchase intent at the early point of time. Also we need to perform more experiments to generalize the results of the study.

8. References

[1] A.L. Montgomery; S. Li; K. Srinivasan; J.C. Liechty, "Modeling Online Browsing and Path Analysis Using Clickstream Data", Marketing Science, Vol.23, No.4, PP579-595, 2004

[2] B. Justin and H. Thomas, "Unifying Collaborative and Content-Based Filtering", Proceedings of the 21st International Conference on Machine Learning, 2004

[3] C.H. Lim; K.Y. Chung, "Development of Human Sensibility Based Web Agent for On-line Recommendation Service", Journal of the Ergonomics Society of Korea, Vol.23, NO.3, PP1-12, 2004

[4] G. Adomavicius; A. Tuzhilin, "Personalization Technologies: a Process-oriented Perspective", Communications of the ACM, Vol.37, No.10, PP83-90, 2005

[5] H.J. Lee, "Effects of Product Recommendations on Customer Behavior in e-Commerce: An Empirical Analysis of Online Bookstore Clickstream Data", Journal of The Korean Operations Research and Management Science Society, Vol33, No.3, PP59-76, 2008

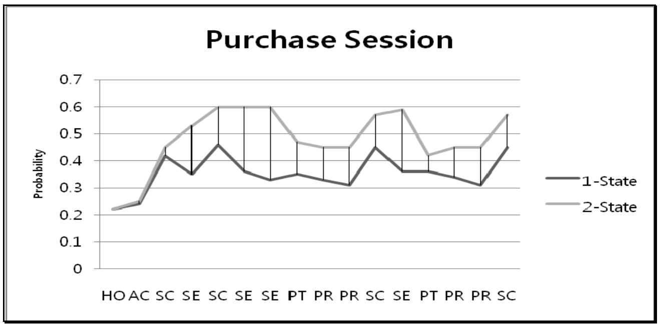
[6] J.A. Konstan; B.N. Miller; D. Maltz; J.L. Herlocker; L.R. Gordon; J. Riedl, “GroupLens: Applying Collaborative Filtering to Usenet News”, Communications of the ACM, Vol.40, No.3, PP77-87, 1997

[7] J.S. Breese; D. Hecherman; C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering", Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, Madison, WI, USA, 1998

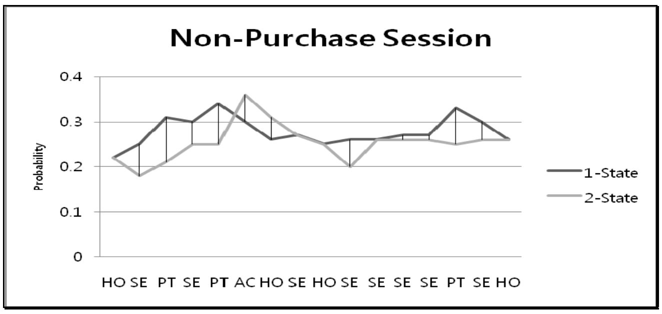
[8] J.W. Kim; H.J. Joo, "Empirical Validation of Customer Characteristics on Internet Shopping Mall Usage", Journal of The Korean Operations Research and Management Science Society, Vol.27, No.4, PP149-165, 2002

[9] M.W. Kim; E.J. Kim, "Performance Improvements in Collaborative Recommendation Using Multi-Layer Perceptron", LNCS, vol.4234, PP350-359, 2006

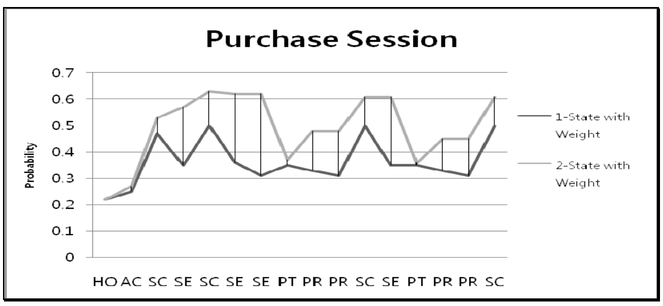
[10] R.E. Bucklin; J.M. Lattin; A. Ansari; D. Bell; E. Coupey; S. Gupta; J.D.C. Little; C. Mela; A. Montgomery; J. Steckel "Choice and the Internet: From Clickstream to Research Stream" Marketing Letters, Vol.13, No.3, PP 245-258, 2002.



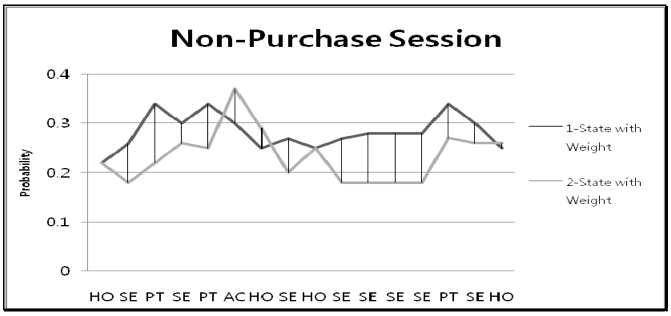
(a) Method 1-Example of the purchase probability of purchase session



(b) Method 1-Example of the purchase probability of non-purchase session



(c) Method 2-Example of the purchase probability of purchase session



(d) Method 2-Example of the purchase probability of non-purchase session

Fig. 1: Experimental results of two sessions