Hotness Analysis based on Users Online Shopping Behavior

Yu Chen [a], Lingling Shen [b,] *and Gang Qian [c]

School of Nanjing Normal University, Nanjing 210046, China

[a]cherrychen1999@163.com, [b]ushen509@163.comqgmail@vip.sina.com, [c]qgmail@vip.sina.com

Abstract

The world wide web produces vast amount of data every day.Retrieving information from such a huge database is very useful for the business.In personalized prediction, user's history and recent online behavior need to be considered at the same time.The HM combines three typical online behaviors,query, browsing and collecting,to calculate product's hotness. We choose different indicators to measure different behaviors. Then, we use a liner parametric function to connect the three behaviors. Finally, we carry out a simulation data to test our model. The result shows that our model is more accurate than that only considering browsing.

Keywords

Hotness, Online behavior, Action reliability, Time validity.

1. Introduction

With the continuous development of E-commerce platform,more and more people prefer shopping online.In this year's Double 11 event,taobao's transaction volume reached 120.7 billion.This is another new record.These online shopping behavior generate large amounts of data.Deep mining it will have great commercial value for the business.

Efforts to data mining have a long history that has sparked lively debate and generated much public interest.Although mining data is pivotal to researchers and marketers, investigations into this issue by such individuals have been few in number, and weak in theory.Jingdong,its approach is a replenishment for several clicks,for the same kind of commodity.However,this doesn't take into account the impact of users' other operations,such as query,collecting and browsing time.And also,it does not consider user's Individual behavior difference.This paper strives to reinvigorate such online behaviors to calculate product's hotness.

We calculate the hotness fully considering three typical behaviors,browsing,query and collecting.To reflect the personal difference ,we propose the action reliability and time validity.

The rest of this paper is organized as follows.Literature review is introduced in Section II . The HM and description of algorithm are presented in SectionIII . In Section IV,we take a simulation analysis.We conclude this study and address the future research directions in Section V.

2. Literature Review

2.1 Users online behavior.

A lot of research points out that the user's interest in product is closely related to online behaviors,such as browsing time,browsing times,collecting and query,etc.[1]-[4] all take the browsing time into account to calculate user's interest.Respectively, [1] considers the browsing behaviors and the same page browsing times,in [2],browsing rate and browsing speed of webpage are considered,[3] is from the view of click behavior and click rate,and [4] calculate the user's interest joining the interest importance. Literature [5][6] think that query behavior is an important factor in reflecting the user's interest. All these references only take into account the impact of current action. The impact of before and after this action is not considered.In this paper,we calculate the user interest considering the last time interest and the time validity

2.2 Time validity model.

In [8],it posits that an index is closer to a recent prediction than to a prior prediction. Therefore, they associate a higher weight to a recent prediction than to an old prediction.[9][10] use the exponential gradual forgetting function and time window method to capture the change of user interest.In [11] ,a slower decay is required and it uses the polynomial decay to describe the application.However,in this paper,we describe the longer browsing time,the higher weight about user's interest and a exponential growing is required .

3. Hotness Model(HM)

3.1 Model construction.

User online browsing behavior is usually with a strong purpose that is intended to buy a certain goods. A correct understanding of the user's shopping will find the user's focus on the product, to find the hidden rules, these are the key points of the user online shopping behavior analysis. Analysis of the user shopping behavior, can understand the user's shopping needs.

We define some terms for describing our approach as following:

Definition 1 Hotness: The popularity of products. It's also the degree of user interest in product.

Definition 2 Action reliability : It's a weight about users' behavior. Such as ,a user places an order after every collecting, so his action reliability is 1.However,if he never places an order after collecting, then his action reliability is 0.We can calculate user's action reliability from the history data.It is described in Eq.(1),

$$R = \frac{\sum O}{\sum A_1 + \sum O} \tag{1}$$

where O represents the number of order before, and A means action. Its value is 0 or 1.If this action is executed, it is 1, otherwise, it is o.

Definition 3 Time validity: When a user browsing a product, the browsing time can largely reflect this user's interest, the longer ,the more interest. We describe time validity as Eq.(2),

$$V(t) = \begin{cases} \exp(\frac{t}{\theta}) & t > 0 \\ 0 & t = 0 \end{cases} \tag{2}$$

where t represents the browsing time, θ is a coefficient. It can be adjusted according to different requirement.

Product's hotness is composed of three parts, collecting value, query value and browsing value.

(1) Collecting value: When collecting a goods, user has to log in, so we can use his history data to know his action reliability. We use C to describe the every user's collecting value for different goods. It's a m*n matrix. C_{ij} means the ith user's collecting value for the jth product. A certain goods' collecting value

$$C_j = \sum A_1 * R \tag{3}$$

(2) Query value: Every query add the product's hotness by 1.So we count the query time to describe the query value. The total query time is represented by Q.

(3) Browsing value: Browsing value is strongly reflected by the browsing time. So it is showed as below

$$B = A_2 * V(t) \tag{4}$$

We choose a liner parametric function to predict the hotness as below.

$$H = \alpha * normalized(C_j) + \beta * normalized(Q) + \theta * normalized(B) \tag{5}$$

α,ρ and θ are coefficient .It can be modified according to different requirements. C_j, $_Q$ and B, they have different factors. So we need to normalize them.

3.2 Architecture of the model.

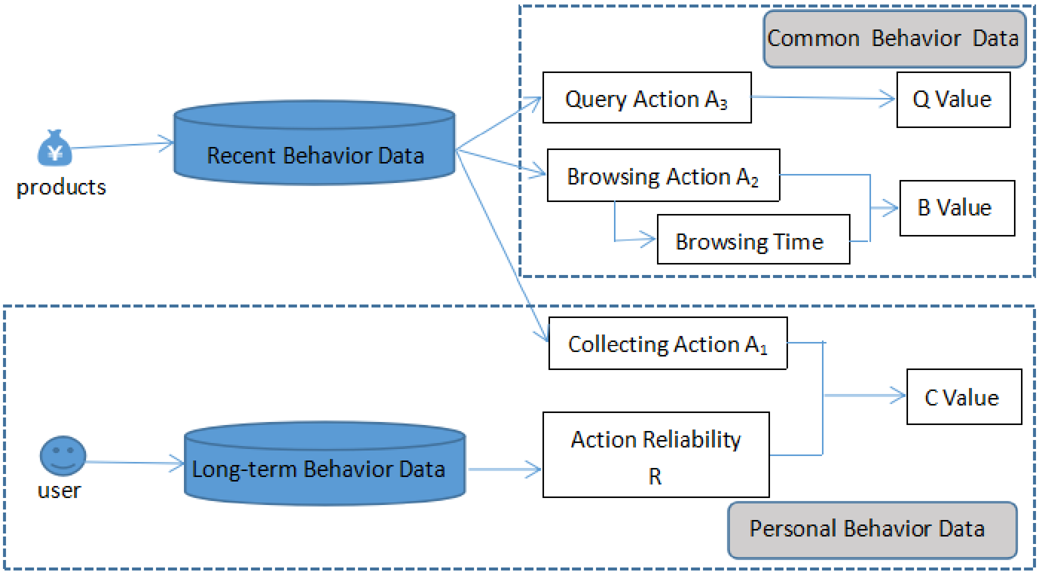

Fig.1 Architecture of PPM

3.3 Procedure of hotness model.

Table 1. The procedure of hotness model

Step 1: Get the query behavior matrix A_3 based on the online behavior database, calculate the sum of query times for different product.
Step 2: Extract the browsing behavior and browsing time matrix A_2 and T. Calculate the B value according to the Eq. (4).
Step 3: 1) Extract the collecting behavior matrix A_1;
2) I f a user collect, he has to log in. So we can have his history behavior data. Then calculate the action reliability R according to Eq.(1);
3) Calculate the C value by Eq. (3).
Step 4: Based on Eq. (5), calculate different products' hotness H.

4. Simulation Analysis

In order to test our model we create a virtual online shopping platform about smart phone. After opening to the users for a period of time, we use the data to calculate products hotness based on the HM. And, we also calculate the hotness only considering browsing times. The results are showed as below in Fig.2.

Fig.2 Products Hotness

In Fig.2, the solid line shows the hotness estimated by HM. The dotted line describes the hotness only considering browsing times. From the picture we can see that the hotness only considering browsing times is basically higher than that predicted hotness by HM. Because browsing is a frequent behavior .Its action reliability is not so high. Collecting does not happen so often .So, only considering browsing behavior is not enough and reliable.

In order to verify our model's accuracy, we sort the different phones according to the hotness value calculated by HM and browsing times. Then we ask our users to sort phones. This is the actual condition .The result is shown in the Table 2.

Table 2. Sort List

HM	User	Only Browsing Times
iphone 7	iphone 7	iphone SE
iphone SE	iphone SE	iphone 7
iphone 6s	iphone 6s	iphone 6s
honor 8	honor 8	Huawei nova
Huawei nova	Huawei nova	honor 8
Xiaomi 5s plus	Xiaomi 5s plus	Xiaomi 5s plus
OPPO R9s	Hongmi NOte 4	vivo X9 plus
Hongmi NOte 4	OPPO R9s	OPPO R9s
vivo X9 plus	vivo X9 plus	Hongmi NOte 4
galaxy s7	galaxy s7	galaxy s7

In Table 2, comparing three different sort list, we can see that the sort list from HM is closer to list given by user.than that from browsing times. So, our model is more precise than only considering browsing times.

5. Conclusion

User's online behavior can reflect the user's actual interest. Aiming at the relationship between the user's online behavior and the user's interest,we create the HM to calculate products' hotness. The users' typical behavior are query, browsing and collecting.Different behaviors have different calculation indexes. These three typical behaviors' index are frequency, browsing time and action reliability,respectively.Finally,we have to normalize different values to calculate the hotness.

The experiment proofs that our model can predict the relatively accurate products hotness. There can be many applications for the hotness.For example,it can be applied in precise marketing.And also ,this can supervise the warehouse stock.It can tell the merchants which products need replenishment.So,this will help reduce the inventory cost and improve users’ shopping experience by faster delivery.

However,our model has deficiencies .Such as,it cannot recognize some malicious data.There is still lots of work needing to be done to polish this model.

References

[1] Jian-Ting L I, Guo Y, Tang Z J. User interest degree calculating based on analysis users' browsing behaviors [J]. Computer Engineering & Design, 2012, 33(3):968-972.

[2] Xia Y, Liu Y. An Improved Method to Calculate User's Interest Degree and Amend User's Interest [J]. Journal of Modern Information, 2014.

[3] Xu Bo, Zhang Jiekui, Zhou JUn. A user interest model based on the analysis of user behaviors [J].Journal of Intelligence, 2009(185):166-169.

[4] Cheng S, Liu Y. Time-Aware and Grey Incidence Theory Based User Interest Modeling for Document Recommendation [J]. Cybernetics & Information Technologies, 2015, 15(2):36-52.

[5] Song T, Zhao H, Zheng S, et al. Construction of User Interest Model Based on Queries and Concept[J]. Journal of Jilin University (Information Science Edition), 2015.

[6] Sethi S, Dixit A. Design of personalised search system based on user interest and query structuring[C]// International Conference on Computing for Sustainable Global Development. IEEE, 2015.

[7]Tamine B L, Boughanem M, Zemirli W N. Inferring the user’s interests using the search history[C]// The Workshop on Information Retrieval. 2014.

[8] Puwei Wang,Anup K Kalia, Munindar P Singh.A Collaborative Approach to Predicting Service Price for QoS-Aware Service Selection[C], IEEE International Conference on Web Services,2015.

[9] Improved collaborative filtering algorithm based on tags. Computer Engineering and Applications, 2016, 52(8): 56-61.

[10] Feng H, Tian J, Wang H J, et al. Personalized recommendations based on time-weighted overlapping community detection [J]. Information & Management, 2015, 52(7):789-800.

[11] Cormode G, Shkapenyuk V, Srivastava D, et al. Forward Decay: A Practical Time Decay Model for Streaming Systems [J]. 2009:138-149.

[12]Xianbing Wang.Research on Method of Evaluating Confidence of Social Annotations [D]. Huazhong University of Science & Technology,2012.

[13]Rao, Kannasani Srinivasa, Krishnamurthy, M, Kannan, A. Extracting the User’s Interests by Using Web Log Data Based on Web Usage Mining[J]. Journal of Computational and Theoretical Nanoscience, 2015.

[14]Shi Y. Research on Precision Marketing Model under the Internet and Big Data Background[C]// International Conference on Social Science, Education and Humanities Research. 2016.

[15]Hendrickson B M. System and method for improving customer wait time, customer service, and marketing efficiency in the restaurant, retail, hospitality, travel, and entertainment industries [J]. 2016.