Based on user demand depth-driven personalized recommendation algorithm

Qianqian Liu [a], Mingchun Zheng [b]

School of Shandong Normal University, Shandong 250000, China

[a]13589016820@163.com, bzhmc163@163.com

Abstract

As the rapid development of the Internet, Online Shopping slowly became a kind of contemporary consumption habits of people. However, the huge amount of data on the Internet results in information overload and other issues. Given Recommended has considerable social value and economic value, personalized recommendation attracts more and more scholars to study, and a large number of recommendation algorithms emerge. The current recommendation algorithms are: (1) recommendation algorithm based content [1], (2) recommendation algorithm based on bipartite graphs [2], (3) collaborative filtering recommendation algorithm [3]. This paper presents a based on user demand depth-driven personalized recommendation algorithm. In this algorithm, we classify the user by user behaviors. The data sparse [4] and cold start problems [4] could be solved validly. At the same time, we combined industry knowledge with horizontal recommend and vertical recommend in order to improve accuracy of recommend algorithm.

Keywords

User behavior analysis, personalized recommendation, Demand depth-driven.

1. Introduction

As the Internet going into our lives, a lot of information comes into our field of vision. On this network platform, getting information becomes more and more easy, as well as the accuracy and speed of getting information face enormous challenges. Faced with the problem of information overload, individualized recommendation system shows great vitality. Algorithm not only creates social value cannot be ignored, but also can generate enormous economic benefits. According to the actual data, in 2006, the emergence of personalized recommendation system rises nearly 30% of turnover for Amazon [5]. It produces huge commercial value. However, due to the cold-start problem and sparseness of scoring matrix and other issues cause the current recommendation algorithm is also facing recommended accuracy difference and recommend lag [5]. Therefore, this paper put forward the based on user demand depth-driven personalized recommendation algorithm. The algorithm combines industry domain knowledge with personalized recommendation algorithm, and it is no need studying complex social relationships. It could classify all user according to user behavior as well as the time when this behavior occurred. The behavior includes the following acts: browse, collect, adding shopping cart and purchase. And then by requirements phase- identify functions and need-deviation function to determine the needs of users migrate [6] and demand evolution [6]. Combined with collaborative filtering algorithm to achieve the appropriate requirements phase personalized precise recommendations.

2. Consumer Behavior Motivation And Its Application

2.1 Principle of Hierarchy Of Needs

Abraham Maslow—American social psychologist, personality theorist—had raised people's "hierarchy of needs". It said that human needs could be divided into five levels, namely the physiological needs, security needs, social needs, esteem needs and self-actualization needs. When the low-level needs are met, the needs will change. Combined the classical management theory with the current background, we can find that in lots of industries, users demands often change as their living background and living environment change. The change includes three forms. Frist is horizontal demand migration. The second is vertical demand evolution. The last is a burst of social factors results the user's current needs change [6]. The last change could not be manipulated by human. However, through data analysis, the horizontal demand migration [6] and vertical demand evolution [6] could master the relevant laws. Demand migration is a form of change that as a result of the change of user attributes, the user needs to be re-clustered. Then according to the new classification results, users' current needs should be re-forecasted. Demand evolution is the current demand phase has been unable to meet the current needs of the user. Recommended products are required to move to a higher level in order to meet users' current needs. Once the user's classification is ensured, traditional recommendation algorithm has none feedback mechanism to detection accuracy of recommend product list. Therefore we are unable to determine whether the current recommendation could meet the needs of users.

2.2 User Clustering

Before analyzing the user, we should obtain evaluation of relevant information firstly. The evaluation of information could be divided into two types- the display evaluation and implicit evaluation. The traditional collaborative filtering algorithm calculates the similarity between users by the display evaluation of relevant information. Faced with the new user, due to having no data about the new user, algorithms cannot achieve the user clustering accurately [8]. However, in this paper, the algorithm calculates the similarity between users by the implicit evaluation of relevant information. Generally, when users want to purchase a good, they will pay more attention to the goods that meet their own living standard and consumer habits. Although they maybe pay attention the "luxury" for which beyond their ability to pay occasionally, compared to products that meet their needs, paying attention to "luxury" is a small probability event. So, we put forward to a new method to calculate the similarity between users by record the user's behavior during the past period of time window. The behavior includes four types: browse, collection, plus shopping cart and purchase. This method could be a very good solution to the problem of cold-start problem [5]. In the absence of scoring records we could calculates the similarity between users by others behaviors.

According to the user's different actions play different roles on user classification, we should give different weights to different actions. Formula is as follows:

$$\mathrm{Waction} = W_b B(u_{ij}) + W_c C(u_{ij}) + W_a A(u_{ij}) + W_p P(u_{ij}) \tag{1}$$

In the above formula, $\mathrm{Waction}$ represents the behavior weights of the users Ui and Uj. W_b, W_c, W_a, W_p represent different behaviors weights of different actions. $B(u_{ij})$, $C(u_{ij})$, $A(u_{ij})$, $P(u_{ij})$ represent coincidence rate of the same action of users Ui and Uj.

We should consider the weight of the different actions as well as time interval of actions [9]. The longer the same action time interval, the importance of the data to measure the similarity of users will show the characteristic of exponential decay. Formula is as follows:

$$W_{Time} = e^{-\alpha \frac{\sum_{i,j=1}^{n} |Ti - Tj|}{n}} \tag{2}$$

In the above formula, W_{Time} represents the time weight of the same behavior of users Ui and Uj. α is the adjustment parameter, which is used to adjust the time attenuation rate. Ti and Tj represent respective time of users Ui and Uj, when they take the same action. In order to make sure the calculation more accurately, we adopt $\sum_{i,j=1}^{n} |Ti - Tj|$ to represent the time interval of taking same action of users Ui and Uj.

In summary, the similarity of users can be determined by the time and behavior. Formula is as follows:

$$sim(Ui,Uj) = Waction \times W_{Time} \tag{3}$$

Then setting an appropriate threshold value k and combined with Top-N algorithm, we could achieve the user cluster analysis.

2.3 User Needs Analysis

According to the theory of consumer behavior, we took extensive social survey and found that consumers have huge potential demand in the horizontal and vertical.

In the horizontal, they want to pursue the production that is slightly higher than their current level of consumption, and they are more willing to pay for such commodities. So we construct a bipartite graph that is composed of users and products which are purchased by the users. According to the target user's browsing record and the browsing times as well as the corresponding behavior weight of the similar users to carry on the resources allocation, generate the final recommendation list[7].In the vertical, they will buy their products associated with the purchased of products. Therefore, we will use domain knowledge [10] to carry out weighted association recommended. As a result, we based on the above findings; we will achieve horizontal recommended and vertical recommended simultaneously. The vertical recommended set clear directions as well as the horizontal recommended pick out the specific commodity.

Horizontal Recommended

The bipartite graph is represented as G (M, U, E), where E is the edge set of the two figure, the project node M is expressed as $m_1, m_2, m_3......m_t$, and the user node U is expressed as $u_1, u_2, u_3......u_j$ .As shown below:

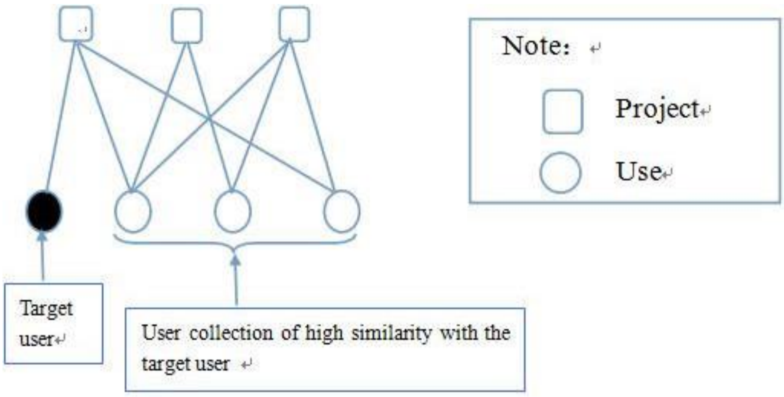

Fig. 1 Horizontal Recommended

The weight of the target user is weighted according to the views of the corresponding products and the retention time of the page. That is , the more the number of browsing the higher the value, the longer the page stay time, the higher the value[11]. The weight of the similar user is weighted according to the product of the similarity between similar users and target users and the weight of behavior. Formula is as follows:

$$Ine_{pq}\begin{cases} 0 & m_t u_j \notin E \\ sim(u_i,u_j)\times W_\beta & m_t u_j \in E \ \& \ \mathrm{U}j \text{ isn't belong to the target user} \\ f(l) & m_t u_j \in E \ \& \ \mathrm{U}j \text{ is belong to the target user} \end{cases} \tag{4}$$

In the above formula, Ine_{pq} represents value of row p column q of adjacency matrix of $t\times j$. $sim(u_i,u_j)$ represents the similarity between similar users and target users. W_β represents the corresponding weights of different behaviors. l represents target user's browsing times for projects. For target user, the weight is a function of the browsing times. In accordance with weight distribution of above formula (4), the proportion of initial target user project resources allocated to other projects is calculated as follows:

$$InLine(pq) = \sum_{n=1}^{j} \frac{\mathrm{Ine}_{pn}\mathrm{Ine}_{qn}}{k(m_p)k(u_n)} \tag{5}$$

In the above formula, $k(m_p)$ and $k(u_n)$ represent the degrees of the project nodes and the degrees of user nodes respectively. Namely, $k(m_p)$ represents number of user nodes connected to the each project node. $k(u_n)$ represents number of project nodes connected to the each use node. W_{pq} represents resource transformation matrix[7].

η represents the initial resource distribution. η^* represents the final result of the distribution of resources. Thus, the final formula is:

$$\eta^* = W_{pq} \times \eta \tag{6}$$

According to (6), the final result of the allocation of resources could achieve the goal of the horizontal recommended.

Vertical Recommended

The vertical recommendation of the target users are mainly based on the industry and the industry stage to which users belong, the user's purchase records in time to generate the recommended candidate set in order to avoid the current recommendation algorithm to repeat the recommend effectively. Frist, we should establish a set of industry product. $PS = \{ps_{11}, ps_{12}, ps_{21}, ps_{22}, ps_{23}.\ldots\ldots\}$. Combined with a large number of users to buy records, taking the method of data mining to mine association between the different products in order to detect the user to buy a certain stage of a product in a timely manner for the user to recommend the industry related to other products, and stimulate consumption. Through the data survey, this paper constructs a simple industry chain to illustrate the algorithm.

As shown in figure 2:

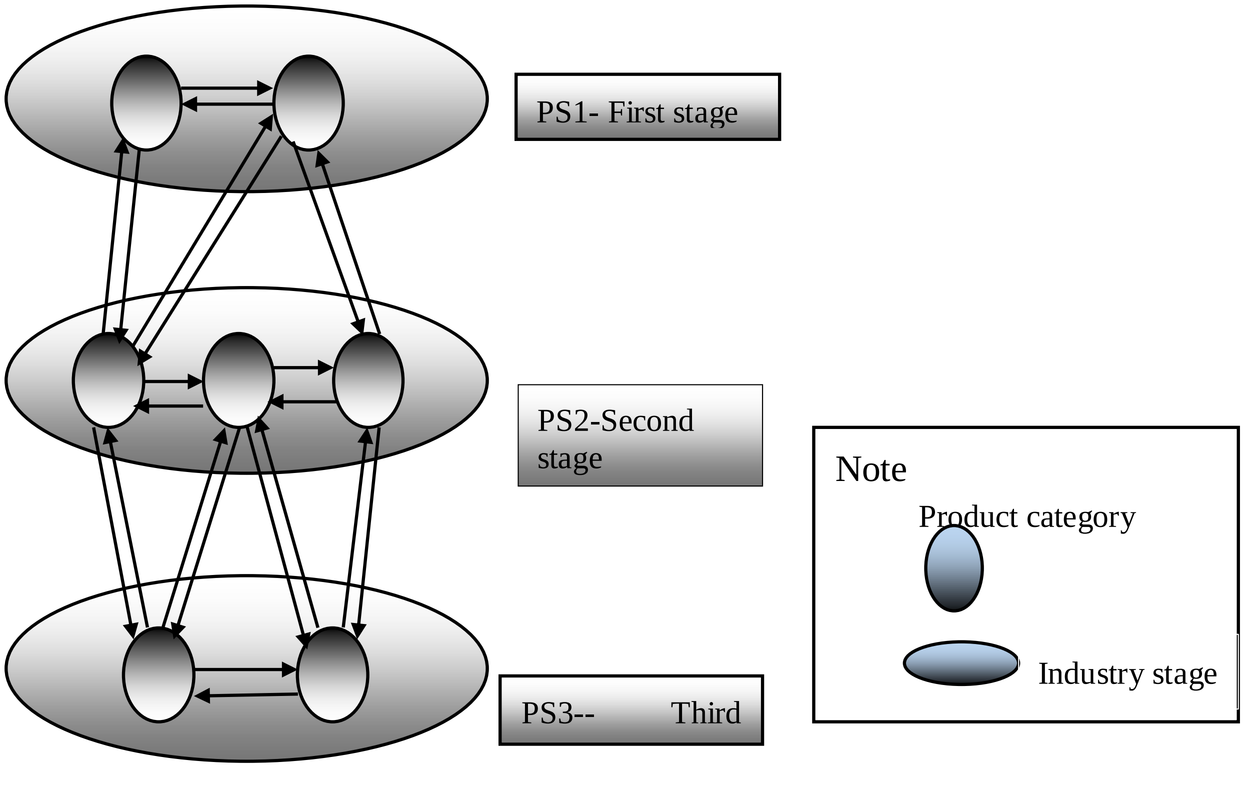

Figure 2 Simple industry product associations

Figure 2 shows a simple industry chain. Because taking into account the purchase of the stage between the purchases of cross phenomenon, there is a two-way relationship between the different products.

Even if the user has been detected in a higher stage, algorithm still will continue recommend the last stage of the purchase of the products do not do recommend, but weaken the recommended weight. During vertical recommendation, because the needs of users mainstream is the current stage, the recommend weight of no consumption of products at the same stage with user is higher than the recommend weight of no consumption of products at the different stage with user. Algorithm uses the weight adjustment factor τ and ϑ to achieve weight regulation.

$$OutLine(st) = \begin{cases} Oute_{st} + \nabla\tau & \text{t and s at the same stage} \\ Outest - \nabla\vartheta & \text{t and s at the same stage} \end{cases} \tag{7}$$

In the above formula, $Oute_{st}$ represents the associated weights of the having purchased product and having not purchased products.s represents the product that user has purchased lastly in the industry chain.t represents the product set that user has not purchased. τ and ϑ as the weight adjustment factor to achieve weight regulation.

2.4 Feedback Mechanism

As the demand for mobility and the evolution of the needs of users demand occurs, the current recommend set may lag behind the needs of users causing invalid recommended. Therefore, to timely detect changes in user demand, in order to guide the user consumption. In this paper, According to the cybernetic principles, the paper put forward to the demand deviation function --- Error to adjust the recommendation set timely by calculating the deviations between the actual purchased set with recommendation lists of user. If the deviation is too large, then lower the accuracy of the current recommendation, we need to update the recommendation list timely. Demand deviation function calculated as follows:

$$Error = \frac{Recommend \cap Purchase}{Recommend} \tag{8}$$

In the above formula, Recommend represents current recommend set---Recommend = $\{r_1, r_2, r_3, r_4 \ldots\ldots\}$. $Purchase$ represents the actual purchased set. Algorithm sets a determination condition---$\alpha\theta$.θ represents specific probability estimates. α is the adjustment factor. To determine the conditions were adjusted accordingly based on the nature of different industries. Demand deviation function needs to run once every period of time window, once discovered deviation is too large, the recommendation list to promptly adjust in response to changes in user demand, which recommend more accurate.

3. Detailed Model

The main mechanism of the model is by the addition of industry knowledge in the traditional collaborative filtering algorithm to predict the user's next record according to the needs of the user's behavior, thus achieving accurate personalized recommendations. The model as follows:

The main operating mechanism of the model is to add industry knowledge in traditional collaborative filtering algorithm, and according to the user's behavior record to predict the next step of the user's needs, in order to achieve accurate personalized recommendation. The specific process is as follows:

cluster analysis of similar users:

Step 1 First of all to take a period of time window of U_i behavior record, through the behavior record of U_i to tell the current needs of the industry and the current demand phase of the industry of U_i.

Step 2 According to the results of previous step; calculate the similarity between U_i and the users who have the same behavior record as U_i by (3).

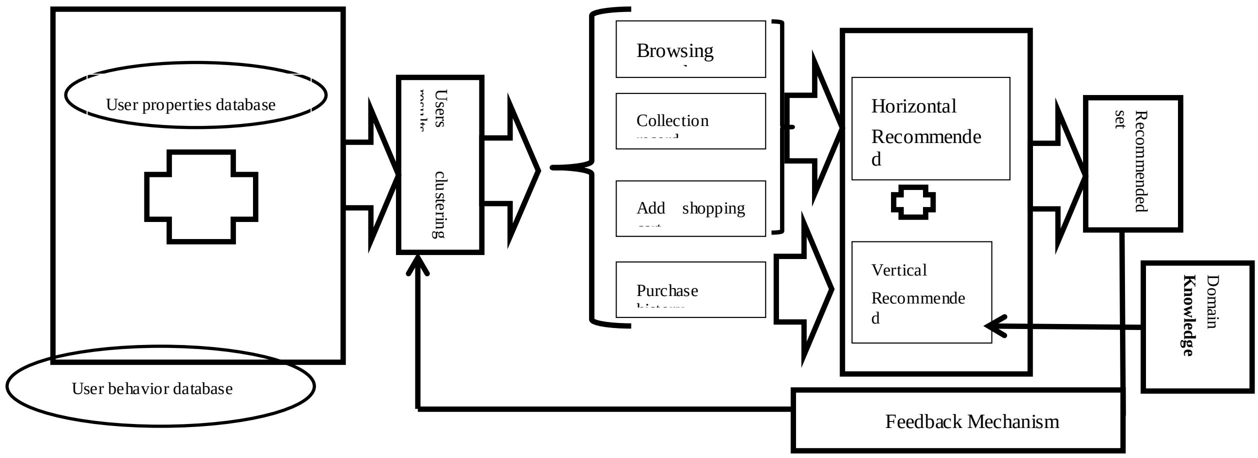

Figure 3 Based on user demand depth-driven personalized recommend model

Step 3 Pick out the user set to meet the threshold value y , according to the priority order of similarity to determine which class Ui belongs to.

horizontal and vertical parallel prediction algorithm

Step 1 According to the user's current browsing record using 5 type of product recommendation.

Step 2 Once Ui has been detected to buy a product within the industry, Ui would have purchased the product of similar products in the industry recommended list to remove all from list to be recommended to update the list.

Step 3 According to (7) the weighted correlation of products recommended and the law of the industry stage for Ui to recommend a higher correlation of items.

recommendation accuracy supervision

Step 1 If the coincidence rate between the recommended list of the target users and the actual purchase of the product list is satisfied with the parameter $\alpha\theta$, it is considered that the current recommendation list meets the needs of the current user.

Step 2 Once the deviation of the two sets is detected too large, it can be recognized as a decrease in the accuracy of recommendation. First, there may be change in the level of consumption of Ui, and Ui needs to be re-clustering. Second, there may be change in the current user demand stage, according to the user's browsing and purchased records for the user's current stage of re-positioning.

Step 3 The demand deviation function Error needs to run once every time interval, which is convenient to discover the change of the target user's requirement in time.

4. Conclusion

By using the user's implicit rating to classify the users, this is a good way to solve the data of the cold start and data sparse problem. Combined with industry knowledge, parallel recommendation algorithm based on horizontal and vertical recommendation could be achieved more accurate recommended results. And the feedback mechanism is started up in time after the recommendation in order to find the change of user's requirement. Comprehensive use of the above mechanisms makes the recommended results more in line with the mainstream needs of the target users.

Based on user demand depth-driven personalized recommendation algorithm not only has great social value, but also has great economic significance. The algorithm could be effective for data filtering and improve the site's user stickiness and business transactions. But the complexity of the algorithm remains to be improved. The mechanism for industry Determination and stage decision is still to be improved. This is our next step should be further discussed.

Acknowledgements

This work is support by the Natural Science Foundation of China(NO.614002266),Natural Science Foundation of Shandong Province, China(ZR2012MF013),Social Science Foundation of China(14BTQ049).We thank you our shepherd Mingchun Zheng and the anonymous reviewers for comments.

Reference

[1] Katrien Verbert, Nikos Manouselis, Xavier Ochoa, Matin wolpers. Context-Aware Recommender Systems for Learning: A Survey and Future Challenges [J].IEEE Transactions on Learning Technologies. 2012.

[2] Yahui Pei, Xinliang Liu. Recommendation method based on timing bipartite graph of user feedback [J]. Journal of Henan University (Natural Science). 2015.

[3] Hongwei Ma, Guangwei Zhang, Peng Li. Review on Collaborative Filtering Recommendation Algorithm [J]. Journal of Chinese Computer Systems. 2009.

[4] Duo Liu, Yu-Xun Lu,Yong-Sheng Zhang,Ling-Yun Guo. A new item recommend algorithm of sparse data set based on user behavior analyzing[C].International Conference on Signal Processing (ICSP). 2014.

[5] Bo Yang, Pengfei Zhao. Review of recommendation algorithms [J].Journal of Shanxi University (Natural Science Edition). 2011,34(3)：337~350.

[6] Haitao Xie, Xiangwu Meng. A personalized information service model adapted to the evolution of user needs [J]. Acta Electronica Sinica.2011.

[7] Xinmeng Zhang, Shengyi Jiang. A personalized recommendation algorithm based on weighted two parts graph [J].Journal of Computer Applications. 2012, 03:654-657+678.

[8] Ren Huang, Tingting Meng. Review of personalized recommendation algorithms [J]. Management and technology of small and medium sized enterprises. 2015, 03:271-273.

[9] Guangfu Sun, Le Wu, Qi Liu, Chen Zhu, Enhong Chen. Collaborative filtering and filtering recommendation algorithm based on time series behavior [J]. Journal of Software. 2013, 24(11):2721-2733.

[10] Quan Chen, Lingling Zhang, Yong Shi. Research on the personalized recommendation model based on domain knowledge and its application [J]. Chinese Journal of Management.2012, 10:1505-1509.

[11] Yezheng Liu. Network consumer behavior: a theoretical approach and its application [M]. Beijing: Science Press, China.2011