Personalized Recommendation Algorithm based on SVM

Bing Wu[1]
[1]School of Management,
Wuhan University of Technology
Wuhan 430070, China
wubingscheduling@whut.edu.cn

Luo Qi[2,3] and XiongFeng[1]
[2] School of Electrical Engineering, Wuhan Institute of technology, Wuhan 430070, China
[3] Department of Information Engineering, Wuhan University of Science and Technology Zhongnan Branch, Wuhan, 430223, China
Whluo2008@gamil.com

***Abstract*—with the development of the E-commerce, personalized product and service have become a developing trend gradually .To meet the personalized needs of customers in E-commerce, a new personalized recommendation algorithm based on support vector machine was proposed in the paper. First, user profile was organized hierarchically into field information and atomic information needs, considering similar information needs in the group users. Support vector machine was adopted for collaborative recommendation in classification mode, and then Vector Space Model was used for content-based recommendation according to atomic information needs. The algorithm had overcome the demerit of using collaborative or content-based recommendation solely, which improved the precision and recall in a large degree. It also fits for large scale group recommendation. The algorithm could also used in personalized recommendation service system based on E-commerce.**

I. INTRODUCTION

With the development of the E-commerce, personalized product and service have become a developing trend gradually [1].The management pattern of E-commerce may greatly save the cost in the physical environment and bring conveniences to customers. People pay more and more attention to E-commerce day by day. Therefore, more and more enterprises have set up their own E-commence websites to sell commodities or issue information service. But the application of these websites is difficult to attract customer' initiative participation. Only 2%-4% visitors purchase the commodities on E-commence websites [2]. The investigation indicates that personalized recommendation system that selecting and purchasing commodities is imperfect. The validity and accuracy of providing commodities are low. If E-commence websites want to attract more visitors to customers improve the loyalty degree of customers and strengthen the cross sale ability of websites, the idea of personalized design should be needed. It means that commodities and information service should be provided according to customers' needs. The key of personalized design is how to recommend commodities based on their interests.

At present, many scholars have carried on a great deal of researches on personalized recommendation algorithms, such as collaborative recommendation algorithm and content-based recommendation algorithm [3]. Although collaborative recommendation algorithm is able to mine out potential interests for users, it has some disadvantages such as sparseness, cold start and special users. Similarly, the content-based recommendation algorithm also has some problems such as incomplete information that mined out, limited content of the recommendation, and lack of user feedback [4].

According to this, a new personalized recommendation algorithm based on support vector machine is proposed in the paper. It has improved the traditional algorithm. First, user profile is organized hierarchically into field information and atomic information needs, considering similar information needs in the group users. Support vector machine is adopted for collaborative recommendation in classification mode to ensure the recall, and then Vector Space Model is used for content-based recommend according to atomic information needs, which ensures the precision. The algorithm is also used in personalized recommendation service system based on E-commence. The results manifest that it can support E-commerce better.

II. USER PROFILE

Supposed that some information fields' space S_d constructs global information space $S_{global} = S_{d1} \bigcup S_{d2} \bigcup \cdots S_{dn}$, space of personal information need is the subset of space of some field information $DIS = DIS_{i1} \bigcup DIS_{i2} \bigcup \cdots \bigcup DIS_{in}$. DIS is the subset of GIS, the relations among them as follows:

$$PIS \subseteq DIS \subseteq GIS$$

In fact, in large scale group users, users' information need has a big superposition. $PIS_{i1} \bigcap PIS_{i2} \bigcap \cdots PIS_{in} \neq \varnothing$ Usually, a user may interest in some fields. Organizing users' information need hierarchically is not only satisfied with users' actual need, but also the system can provide recommendation service conveniently. Thus, the paper uses this method to organize users' information need, the structure is Fig .1

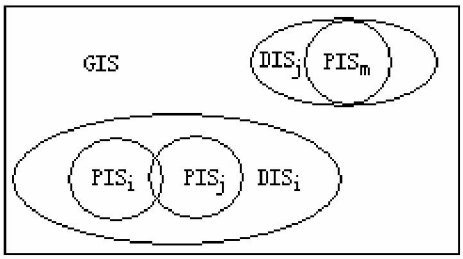

Fig. 1. Relationship of user information

Space of users' information need could be seen as a U= {UID, PIS, Period, Freq}

$$\text{UID=inter UID} \,||\, \text{Exter UID}$$

$$\text{PIS=GIS} \,||\, \text{DIS} \,||\, \text{PIS}$$

$$\text{DIS=DIS}_1 \,||\, \text{DIS}_2 \,||\, DIS_3 \cdots ||\, DIS_k$$

$$PIN = PIS_1 \,||\, PIS_2 \cdots ||\, PIS_m$$

$$PIS_i = < I_i, w_i >$$

$$Period = StartTime \,||\, EndTime$$

$$\text{Freq=Hour} \,||\, \text{Day} \,||\, \text{Week} \,||\, \text{Month}$$

$< I_i, w_i >$ Separately represents item and weighting of users' actual need. StrartTime and End time separately represent the time of system service

Form the structure, a model of users' information need is a tree. Root node is user UID, the middle node is corresponding to the catalog of field information. Leafage node saves actual need of current field information PIS_i. PIS_i Is not separable, which is called atomic information need. The tree is Fig. 2

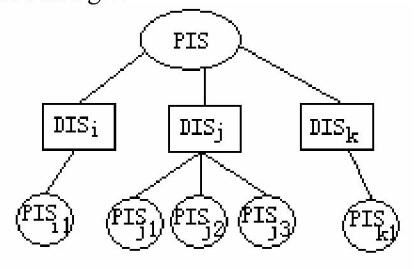

Fig. 2. Hierarchical information need

.

III. Personalized Recommendation Algorithm Based on Support Vector Machine

A. Recommendation Strategies

Field need information is used by classification recommendation, and then users' actual atom information need is used by content-based recommendation. There are two phases:

(1)First filtering. Users' field information need is classified by SVM, which ensure the recall. This strategy considers user recessive need.

(2)Second filtering. Basing on first filtering, Second filtering realizes user atom information need recommendation by VSM, which ensure the precision.

B. Field Information Needs classification

Support vector machine for classification can be seen as the application of perceptron [5]. When classification problem is linearly separable, a hyperplane that makes two categories of separable data more close to the plane is set up usually the plane is called optimal separation hyperplane. Regarding to nonlinear problem, original data is mapped from a low dimensional space to the new data sets of a higher dimensional space (feature space) through a nonlinear mapping (nuclear function).New data sets are linearly separable in feature space, thus the classification in higher dimensional space is completed.

Basing on it, we can put the question of nonlinear separable and linear separable into a unified formula.

Regarding to the sample space $\Omega = \{(x_i, y_i) \,||\, i = 1, 2, \cdots, N\} \subset R^n \times \{-1, 1\}$ and the function $\{\Theta(x_i), \Theta(x_j)\} = K(x_i, x_j)$ Standard support vector machine can represent as follows:

$$\min Q(w, \varepsilon) = \frac{1}{2} \,||\, w \,||^2 + C \sum_{i=2}^{l} \varepsilon_i \qquad \textbf{(1)}$$

$$s.t. y_i [\{w \bullet \Theta(x_i) + b] - 1 + \varepsilon_i \geq 0, \quad \varepsilon_i \geq 0, i = 1, \cdots, l \qquad \textbf{(2)}$$

(a). When $K(x_i, x_j)$ is linear transform, especially when $K(x_i, x_j)$ is linear invariant mapping, and C=0, $\forall \varepsilon_i = 0$, (1) (2) is correspond to linear separable condition.

(b). $K(x_i, x_j)$ is nonlinear mapping that transforms Ω to a higher dimensional space H, and $K(x_i, x_j)$ satisfies Mercer theorem. If $K(x_i, x_j)$ is Kernel Function, (1) (2) is correspond to nonlinear separable condition. C>0 is a constant, which control the punishment degree of

misclassification. Loss function $\sum_{i=1}^{l}\varepsilon_i$ is an upper bound of misclassification. In fact, $\sum_{i=1}^{l}\varepsilon_i$ can represent as

$$F_\sigma(\varepsilon)\sum_{i=1}^{l}\varepsilon_i^\sigma \quad (3)$$

σ =1 is corresponding to support vector machine of one time loss function, σ =2 is correspond to support vector machine of twice time loss function.

(1)- (2) can be summarized to solve the problem of quadratic programming.

$$Min \quad W(a) = -\sum_{i=1}^{l} a_i + \frac{1}{2}\sum_{i,j=1}^{l} a_i a_j y_i y_j K(x_i \bullet x_j) \quad (4)$$

$$s.t: 0 \le a_i \le C, i = 1, \cdots, l; \sum_{i=1}^{l} a_i y_i = 0; \quad (5)$$

That is corresponding to decision functions of the broadest and optimal classification

$$f(x) = \mathrm{sgn}(\sum_{S.V.} y_i a_i K(x_i \bullet x) + b) \quad (6)$$

In fact, a classifier is a judge function, which makes variable D in definition domain divide to inconsistent range subspaces $C = \{C_1, C_2, C_3, \cdots, C_n\}$ though certain principles, $C_i \bigcap C_j = \varnothing$ the input variables is the most approach to another defined category according to some testing degree σ .

$$f_\sigma = (x_0) = \arg\max_{c_{ieC}}(c_i, x_0) \quad (7)$$

C. *Atom Information Needs Recommendation*

Salton method is used to compute the similarity. In Salton method, each text is represented as an n dimensional vectors W= ($w_1, w_2, w_3, \cdots, w_n$), component w_i is corresponding to the weigh of feature in this text. The measurement of w_i as follows:

$$w_i = \frac{tf_i \times \log(N/n_i)}{\sqrt{\sum_j (tf_j \times \log(N/n))^2}} \quad (8)$$

While, tf_i is occurrence times of feature in assigned texts, N is the total number of text in set of texts, n_i is occurrence times of feature.

The similarity s is computed according to the cosine value between user need vectors and every text.

$$S = COS(w_u, w_d) = \frac{\sum_{i=1}^{n} w_{ui} \bullet w_{di}}{\sqrt{\sum_{i=1}^{n}(w_{ui})^2 \bullet \sum_{i=1}^{n}(w_{di})^2}} \quad (9)$$

For recommendation results, real information needs and potential users information needs should be considered. So, the recommendation results R are divided into positive region sets R^+ and negative region sets R^- , which is corresponding to user real needs and potential needs, and thus system recommendation decision as follows:

$$R = \begin{cases} R^+, s \ge s_0^+ \\ R^-, s_0^- \le s \le s_0^+ \\ discard, s < s_0^- \end{cases} \quad (10)$$

s_0^+, s_0^1 are represented separately as recommendation of positive region and negative region.

IV. CONCLUSION

In summary, a personalized recommendation algorithm based on support vector machine is proposed in the paper. The algorithm can also used in personalized recommendation service system based on E-commence. The results manifest that the algorithm is effective through testing in personalized recommendation service system based on E-commence.

.

ACKNOWLEDGMENT

Supported by National Natural Science Foundation of P.R. China (Grant No.60574070 and 70271033).

REFERENCES

[1] Chen Deli, "Design of Personalization E-commerce Model Based on Data Mining", Journal of Information, Vol.8,2006.

[2] Liu Wei, "Research on Information Recommendation in E - commerce Systems", Information Science, Vol.2,2006.

[3] Wang Xia, "Research on application of item-based collaborative filtering in recommender systems", Computer Engineering and Design, Vol.7,2007.

[4] Liu Pingfeng, "Self-adaptive recommendation policy in recommender systems", Computer Engineering and Applications, Vol.4,2007.

[5] Gan Liangzhi, "Sparse least squares support vector machine", Journal of Zhejiang University(Engineering Science), Vol.2,2007.