ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 2 (2012) No. 4, pp. 50-57

Application of vague set in recommender systems

Cui Chunsheng[1], Zhenchun Zang[1], Liu Feng[2], Qu Ying[3]
[1] Organization Department, Henan University of Economics and Law Zhengzhou, Henan University of Economics and Law, China
[2] Department of Information Management, the Central Institute for Correctional Police, China
liufeng153@sina.com
[3] School of Economics and Management HeiBei University of Science and Technology Shijiazhuang, China

Abstract: In the paper, vague set theory is introduced into the study of recommender systems to solve its core problem which is similarity. The existence of uncertainty of customer behavior in the course of e-commerce provides a theoretical basis for the introduction of vague set. Recommendation of goods relies on the degree of similarity between customers or goods, while the calculation of similarity is a mature area in the research of vague set. First, Different customer types are identified according to the general shopping way in e-commerce. Then based on the customer classification, statistical methods are used to define the vague value of the commodity. This method makes a perfect combination e-commerce recommendation system and Vague set and provides new idea for the study of e-commerce recommendation system.

Keywords: Recommender Systems, Similarity, Vague Set, Vague Value

1. Introduction

Recommender systems or recommendation engines, resulted from a specific type of information filtering system technique, attempt to recommend informational items (films, music, books, etc.) which are likely to cater to the users. Up to now, many recommendation algorithms have been proposed, such as collaborative filtering (Adomaviciu & Tuzhilin, 2005 ; Goldberg, et al, 1992; Schafer, et al., 2007), content-based analysis (Pazzani & Billsus, 2007), spectral analysis (Goldberg, 2001), latent semantic models (Hofmann, 2004), heat conduction (Zhang, et al., 2007; Zhou, et al.,2010), opinion diffusion (Zhang, et al., 2007; Zhou, et al.,2007), and so on.

One of the most important issues in recommender systems research is similarity. Obviously, it is a perfect way for recommendation quality to descript actual situation similarity from the different angle. Nowadays, methods in recommender systems include: cosine similarity, related similar, revision cosine similarity and so on. There are three disadvantages of these approaches. Firstly, since the rating information is the fundamental ingredient for calculating these methods, they cannot be applied to the systems without explicit ratings. Secondly, their performances may decrease when data gets sparse. An evidence is shown in Ref (Shang, et al., 2009), where they found that similarity based on the relevance information, namely whether a user has voted an object, can output a better recommendation than that based on the exact rating correlations on sparse data. Thirdly, their computational complexities are relative high, especially for the huge-size and dense data sets.

So, many scholars are also attempting to use new methods for similarity (Cui, et al., 2009), but effects are poor. In our opinion, changing question direction will be helpful to enhance similarity accuracy, and promote recommendation quality.

Dekang Lin (Lin, 1998) presents an information-theoretic definition of similarity that is applicable as long as there is a probabilistic model. The definition can be used to measure the similarity in a number of different domains. Meanwhile, there are lots of achievements about similarity in the field of vague sets. Maybe the achievements will help to research recommender systems (Cui & Wu, 1995).

2. Vague value of recommender systems

The issue of content-based recommendation is the similarity of items. With the result of products similarity, recommender systems can recommend right item to target user. So, the basic of this algorithm is products classification by their characteristics description. While, products classification relies on user historical behaviour. Therefore, from the user behaviour, with Vague set describing product characteristic is foundation of recommendation system.

There are two ways to get user behavioural data of recommender systems, one is explicit navigation, and another is implicit navigation. The former need user provides rating, appraisal or other operations. It is too unfriendly to antipathy for user. The latter is popular nowadays for its feasible to user. But it is more difficult than others in technology. For example, Stevens obtained user behaviour through three kinds of data which are read or ignored, saved or deleted, replied or not replied. Nichols [16]gave the implicit feedback behaviour tabulation, like purchase, visit, repeat, use, printing or save, deletion, quotation, reply, labelling, research or reading, skip, connection, inquiry and so on. Browsing, collection, shopping cart and purchases are used to obtain user behaviour in Literature (Tang & Fan, 2009). And they believe weight rank of four behaviours is Browsing < Collection < shopping cart< Purchase.

Therefore, the five factors affecting the users behaviour which are described by the vague theory and the products described by vague value recommender system are adopted, the large amount of research result of vague set theory will be studied in the recommender system accordingly.

2.1 Question definition

In order to describe the topic reasonably, the related definitions below are given.

Definition2. Browsing process: The process of user in B2C included turning website, browsing and shopping.

Definition3. Product choice: user gain satisfaction product from many products.

Definition4. Shopping behaviors: For some goal, user selects and purchases product in a B2C commerce website and submits order. Regarding the business, such behavior may also regard as commercial transaction (Ji, 2004).

Usually, there are 8 kinds of browsing process electronic commerce for certain product, such as the second column of Table 1.

Table1 Browsing process and inference table

serial	Browsing process	Behavior tabulation	reasonable inference
situation1	Purchase without browsing	purchase	The browsing time may neglect, enters pay page directly, purchased the product
situation2	Analysis and purchase after browsing	browsing→choosing→purchase	With normal browsing time, and purchased the product.
situation3	Unnoticed or unknown	none	May be interested in the product, but finally not purchased due to without knowing any information about the product,
situation4	Uninteresting in browsing	browsing	Do not review the detail information of the product due to lack of interest
situation5	Not purchase with Browsing shortly	browsing	Browse with disoperation, or the information is the opposite of what expected, and then leave.
situatio6	Not purchase with Browsing	browsing→choosing	With normal browsing time, also be interested in the product, but finally didn't purchase.
situation7	Not purchase but download or shopping cart	browsing→choosing	Be interested in the product,but finally didn't purchase due to time issue or payment capability

situation8	others	none	With some accident, i.e. power failure, payment failure

2.2 Question classification

We can obtain reasonable inference like the fourth column of Table 1 by the second column of Table 1. Furthermore, we may obtain:

①user purchased the product in situation 1and situation 2; ②Because of high uncertainty of situation 3, advertisement is necessary to reduce user uncertain degree;③user is uninterested in the product in situation 4; ④By definition 3, browsing process only consider user behavior once, therefore user may be regarded as uninterested in the product in situation 5; ⑤user is potential customer in situation6, situation 7 and situation 8 for the choice behavior.

Based on above analyzes, there are three kinds of attitude according to user's behaviors of browsing, choice and purchase.

①advocate recommendation (including situation 1 and situation 2); ②not recommendation (including situation 3, situation4 and situation 5); ③attitude of recommendation is unclear (including situation 6 to situation 8).

2.3 product vague value and relevant definition

Vague set is popular in group decision-making. In certain extent, vague value comes from group decision-making statistical data. Here, with statistical method, we introduce the method of obtaining product vague value through user group behaviors.

Definition 5. Browsing time: users browsing time T_{Br} satisfied the constraints： $T_{Br} \geq \alpha$， here, α is the min browsing time. If $T_{Br} < \alpha$, the user be looked as without browsing.

Definition 6. There are n users and m products in recommender systems, defined x_{uik} as the attitude k of user c_u to product s_i . Where, $u = 1,2,\ldots,n;\ i = 1,2,\ldots,m;\ k = 1,2,3$ ， there are three kinds attitude which are advocate recommendation $(k=1)$, not recommendation $(k=2)$ and otherwise $(k=3)$, here, satisfy $\sum_{k=1}^{3} x_{uik} = 1$, then

$$x_{ui1}=\begin{cases}1,\text{advocate recommendation}\\0,\text{otherwise}\end{cases}\qquad x_{ui2}=\begin{cases}1,\text{not recommendation}\\1,T_{Br}<\alpha,\text{and undownload}\\0,\text{otherwise}\end{cases}$$

Definition 7. (Vague value of product s_i **)** Vague value of product described by *truth-membership* function t_{x_i} and *false-membership* function f_{x_i}

$$t_{x_i}:U\rightarrow[0,1]\,,\quad f_{x_i}:U\rightarrow[0,1]\quad t_{x_i}=\frac{1}{n}\sum_{u=1}^{n}x_{ui1},\ f_{x_i}=\frac{1}{n}\sum_{u=1}^{n}x_{ui2}\ .$$

$[t_{x_i},1-f_{x_i}]$ is product s_i Vague value. $\pi_{x_i}=1-t_{x_i}-f_{x_i}=\frac{1}{n}\sum_{u=1}^{n}x_{ui3}$ is the uncertain degree of product. π_{x_i} is very important in recommender system, because π_{x_i} expressed those users who are interested in the product but not purchase. In electronic commerce, the people can be called potential users. Uncertain degree indicates the rate of implicit users. The role of Recommender systems is to transfer π_{x_i} to t_{x_i} , so the quality of recommender depends on the degree of transferring from π_{x_i} to t_{x_i} .

Definition 8. the Vague set of one kind product can be defined as $A=\sum_{i=1}^{m}\frac{[t_{x_i},1-f_{x_i}]}{x_i}$ .Obviously，product Vague set is a set of Vague value.

3. Similarity

The research field of vague set mainly studies the similarities of Vague set value. There are three trains of thought based on manifestations and main features of the existing formula.

The first kind is score function based on vague value; the second kind studies similarity measurement method based on distance measurement. For example: to evaluate the similarities between x and y by making use of distance antithesis, the distance of x and y and the ratio of complementary distance; the third kind is based on redistribution base of uncertain degree. The comparison is studied by measuring similarities which prorate uncertain degree to truth-membership function.

The paper solely explores the application of the Vague set similarities in the E-commerce intelligence recommendation system, so the simplest score function method is adopted.

Definition 9. (Chen, 1995) Let $x=[t_x,1-f_x], t_x, f_x \in [0,1]$ be a vague value, and $t_x+f_x \le 1$. Then, the score of x can be evaluated by the score function S is $S(x)=t_x-f_x$. Score function reflects the definite evidences in existence, supports and opposes strength contrast, so it is called superior function.

Then, the similarity of vague value is: $sim_1(x,y)=1-\frac{|S(x)-S(y)|}{2}=1-\frac{|t_x-f_x-(t_y-f_y)|}{2}$

On the other hand, Wang weiping think uncertain of Vague should be overlaid but offset. So, the similarity of vague value is:

$$sim(x,y)=1-\lambda_1|t_x-t_y|-\lambda_2|f_x-f_y|-\lambda_3|S(x)-S(y)|-\lambda_4(\pi_x-\pi_y)$$

Here, $\lambda_1=\lambda_2, \lambda_1+\lambda_3=\frac{1}{2}, \lambda_4 \le \frac{1}{2}$, Let $\lambda_1=\lambda_2=\lambda_3=\lambda_4=\frac{1}{4}$, then $sim_2(x,y)=1-\frac{1}{4}\left[|t_x-t_y|+|f_x-f_y|+|S(x)-S(y)|+\pi_x+\pi_y\right]$.

4. Application

In order to confirm the validity, we tracked in six products about birthday paper which website is http://gcnh.taobao.com/ shown in Table 2. Birthday paper is fresh products in domestic market. So, young people who are ability to pay keen on it.

In hundreds of thousands of web stores, the website is not a well-known one. After 15 days' tracking, the basic data obtained are shown in Ref (Cui, et al., 2010).

Through page information and normal browsing speed analysis, takes $\alpha=24$. With definition 5 and definition 6, we obtain the Vague values of six products are $s_1=[0.491,0.877]$, $s_2=[0.442,0.739]$, $s_3=[0.030,0.547]$, $s_4=[0.174,0.742]$, $s_5=[0.168,0.388]$, $s_6=[0.177,0.297]$.So, we get the different similarity.

5. Conclusion

Products in recommender systems have been expressed by vague value. So we can work on two aspects furthermore. On one hand, the definition of product vague value comes from users' attitude which is the product popular in network. Therefore the product vague value indicated in the non-personalized product recommendation. On the other hand, the product vague value is based on user recommendation attitude. Those results can be used in similarity between products, and also can be used to describe recommendation relationship between products.

References

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extension. *IEEE Trans. on Knowledge and Data Engineering*, 17 (6), 734-749.

Goldberg, D., & Nichols, D., & Oki, BM, & Terry D. (1992). Using collaboraive filtering to weave an information tapestry. *Communications of the ACM*, 35 (12), 61-70.

Ben Schafer, J., & Frankowski, D., & Herlocker, J., & Shilad, Sen. (2007). Collaborative filtering recommender systems. *Lecture Notes in Computer Science*, 4321, 291-324.

Pazzani, J. M., & Billsus, D. (2007). Content-based recommendation systems. *Lecture Notes in Computer Science*, 4321, 325-341.

Goldberg, K., & Roeder, T., & Gupta, D., & Perkins, C. (2001). Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, 4(2), 133-151.

Hofmann, T. (2004). Latent semantic models for collaborative filtering. *ACM Transactions on Information Systems*, 22(1), 89-115.

Zhang, Y.C., & Blattner, M. (2007). Heat conduction process on community networks as a recommendation model. *Physics Review Letter*, 99(15), 4301-4305.

Zhou, T., & Kuscsik, Z., & Liu, J. G., & Medo, M., & Wakeling, J. R., & Zhang, Y. C. (2010). Diversity versus accuracy: Solving the apparent dilemma facing recommender systems. *Proceeding of Natural Academic Science*, 4511-4517.

Zhang, Y. C., & Medo M., & Ren, J., & Zhou, T., & Li, T., & Yang, F. (2007). Recommendation model based on opinion diffusion. *Europhysics Letters*, 80(6), 3-7.

Zhou, T., & Ren, J., & Medo, M., & Zhang, Y. C. (2007). Bipartite network projection and personal recommendation. *Physics Review letter*, 76(4), 115-121.

Shang, M. S., & Lü, L. Y., & Zeng, W., & Zhang, Y. C., & Zhou, T. (2009). Relevance is more significant than correlation: Information filtering on sparse data. *Euro-physics Letters*, 88(6), 68008.

Cui, C. S., & Wu, Q. Z., & Zhai, X. K. (2010). Similarity Measure of Vague Value Based on Depentment Function. *2010 Chinese Control and Decision Conference (2010 CCDC)*, Xuzhou, 632 – 635.

Lin, D. K. (1998). An Information-Theoretic Definition of Similarity. In: *Proceedings of the 15th International Conference on Machine Learning*, Madison.

Cui, C. S., & Wu, Q. Z. (2010). Research on Electric Commerce Recommender Systems Based on Vague Set. *Nanjing: First IEEE International Conference on Information Science and Engineering*, 5382 – 5385.

Gau, W. L., & Buehrer, D. J. (1993). Vague Sets. *IEEE Trans SMC* , 23(2), 610- 614.

Nichols, D. M. (1997). Implicit rating and filtering. *Budapaest: Budapaest: ERCIM*.

Tang, X. B., & Fan, J. (2009). Recommendation Based on Customer Clustering. *Journal of intelligence*, 6, 143-146.

Ji, J. Z. (2004). *Bayesian network and its application in web-based intelligent*. Beijing University of Technology, Beijing.

Chen, S. (1995). Measures of Similarity Between Vague Sets. *Fuzzy Sets and Systems*, 74(2), 217-223.