International Journal of u-and e-Service, Science and Technology
Vol.7, No.4 (2014), pp.53-62
http://dx.doi.org/10.14257/ijunnesst2014.7.4.6

Study on A Recommendation Algorithm of Crossing Ranking in E-commerce

Duan Xueying

*Department of Intelligence Engineering,
Jilin Police College, Jilin 132000, China
17222433@qq.com*

Abstract

The analysis of the existing recommendation system and the main task in the electronic commerce application and the existing problems of the basis, according to the new user "cold start" problem, to adopt a user in a number of different categories of electronic commerce website access multi-B2C behavior information recommendation. This paper presents a crossing ranking recommendation algorithm. Its accuracy can be far more than the random recommendation, at the same time keeping and diversity were recommended. All these ensure the algorithm has a good user experience. Experiments show that the algorithm is accurate and the algorithm is further enhanced.

Keywords*: Recommendation systems; Cluster analysis; Crossing ranking; E-commerce*

1. Introduction

Personalized recommendation system has made significant contributions in the past few years for the electronic commerce development and growth. For example, The comprehensive online stores of the world's largest mall, Amazon Mall recommendation system recommended for users may be interested in commodities, According to the recommended statistical brought over 45% additional sales[1-2].

Along with the development of e-commerce in the global development and growth, in the commodity business scope, corresponding with the Amazon this kind of comprehensive E-commerce website is the vertical electronic commerce website, hereinafter referred to as the vertical electric. Vertical electric refers to electronic commerce mode of deepening marketing in a certain industry and markets, by the website size limits, the site's goods are generally belong to the same category products. At this level, a plurality of vertical electric fields can be seen as the different categories of commodity composition of community. Including general electric and vertical electric business, the basic all e-commerce website, just using the recorded information in their website, the recommendation system as an inevitable part of their online marketing. At the same time, the known research only on a single dataset, and no attention and use the user focus on many types of merchandise data access to records, this development to a certain extent in the recommendation system in E-commerce [3]. Application of business demand often prompted more researchers will work in all aspects of energy into electronic commerce recommendation algorithm in. This chapter will be about how to use personalized recommendations for the corresponding research ranking website [4].

This paper considers the use of global information, behavior of users on the site of different types of vertical e-commerce, personalized recommendation under cold start in the

ISSN: 2005-4246 IJUNNEST

complete absence of a website information record. This part uses the research of new ideas, and research the analysis object of new, its ideas and results have important value for the study of the existing e-commerce recommendation system.

2. Data Analysis of Multi-B2C Behavior

In order to verify the effectiveness of ranking recommendation algorithm and a recommendation system gets information for recommendation accuracy, this paper will use the percent mall e-commerce consumer behavior data platform, from a number of different time range of typical vertical B2C website data for empirical analysis, and in accordance with the study on the problem of different to the original the data into two different groups of datasets [5]. Percent mall currently has the most professional recommendation engine platform, is the largest span electric consumption data platform. The main business services model for individual commodities to the electronic commerce website recommendation service. Percent store data on the platform, using the Global ID to capture and record the user behavior trajectory information in a plurality of electronic commerce web-site, these include the user clicks on a product, the collection and purchase behavior. Even if the same user registration different accounts in different business website, using GID information can also identify the user global ID. The entire dataset is provided in a single or a plurality of electric business website user behavior information. In this study, we use multiple business data on the information platform [6].

The experiment data are derived from the B2C web site: peel network focus on skin care and beauty products sales network, many kinds of the goods covered including skin care products, cosmetics, perfume and health products. Wheat bag is a domestic large-scale direct online website bag, bags of multi brands its sales of goods including computer bag and bag included. Red child is currently the nation's largest maternal and child products vertical electric [7-8].

The empirical analysis related to these sites, in the site scale and business scope are great differences.in addition, included the traditional form of B2C Internet marketing in the form, including popular form of network group purchase in recent years[9].

According to the different research, we will file the raw data into two different sets of data. This chapter is a part of the data for the user behavior information contained in the peel network, wheat bags and shoes base three site were selected in April 2012, Completing appropriate data preprocessing, and further divided into peel Network - wheat bags (referred to as G-M) and bread bag - the library name shoes (under referred to as M-S) of two independent datasets. This is an independent dataset that consists of two electric information in the composition is simple, it is known as the two electric dataset, it will be confirm the validity of the ranking recommendation algorithm. We also increased the electric quantity in October 2012, after screening, the wheat bags, red children, Yao point 100 and shows network user behavior information of 4 electric current months. And these four business sites are referred to as X1-MBB, X2-RB, X3-YD and X4-XT, this dataset called multi electric dataset. This information consists of second large groups of datasets to test algorithm, and to verify the effect of recommendation system gets information of recommendation accuracy [10-11].

3. Crossing Ranking Recommendation Algorithm of Multi-B2C Behavior

The key point of crossing ranking recommendation algorithm of multi-B2C behavior, using ranking user multi-B2C behavior, to visit each other in the commodity business, equivalent mapping for the target business on certain commodities in access and give different weights of the association. Figure 1 as an example, assume that the global existence of A and B two B2C sites, including A1, A2, A3, A4, B1, B2,..., B6 consists of 10 items, and 5 different users.

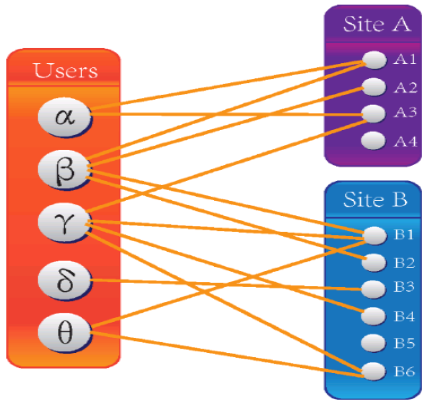

Figure 1. The Two Part Drawing of Ranking Recommended

3.1. The Recommendation Algorithm Based on Two Part Graph Resource Allocation

The user and the commodity as the abstract node, all algorithms exploit the information hidden in the corresponding relationship between users and goods in. Recently, some researchers will matter diffusion and heat conduction in the physical theory is introduced to research information personalized recommendation algorithm, put forward many recommendation algorithm based on network structure model, the method of the user and the commodity as the abstract node, all algorithms exploit the information hidden in the user and commodity choice relationship

The two part graph recommendation algorithm for resource allocation on the assumption of a composed of M users and N recommendation system based on, use a $m + n$ node, the two part graph, if the user I to select goods J, is established between I and j an edge, record $a_{ij} = 1(i = 1, 2 ..., m; j = 1, 2, ..., n)$, otherwise $a_{ij} = 0$. For any user i, recommendation algorithm aims to all users of the I have no choice over the goods are sorted according to the user I like degree, and the ranking of those goods are recommended to the user i. The idea of the algorithm is that, if the user i choose goods have some ability to recommend other commodities to the user i, the abstract ability can be regarded as a resource or energy in some related goods can be assigned. For a M users and N commodities general recommendation system. If w_{ij} commodity j is willing to assigned commodity i resource ratio, The general expression of w_{ij}:

$$w_{ij} = \frac{1}{k_j} \sum_{l=1}^{m} \frac{a_{il} a_{jl}}{k_l} \qquad (1)$$

Where, k_j represents the degree of commodity j, k_l indicates the degree of user i. For a given target user, he chose to use the product information to build over vector f, which sets the initial selection of goods over resources is 1, otherwise set to 0. This resource allocation configuration, carry different users personalized information. By the above formula, the final resource allocation vector can be expressed as:

$$f' = wf \qquad (2)$$

On this basis, The target user did not read all of the goods in accordance with the size of the corresponding element of vector f' sorting. If the value is greater, the user is more like, the top-ranking products will be recommended to the target user. Experimental results show that the recommendation algorithm based on two part graph resource allocation is better than the classic collaborative filtering algorithm in indicators of accuracy and diversity.

3.2. Recommendation List

The product obtained in the previous step based on the mapping table, multi-B2C crossing ranking recommendation algorithm system will online real-time response requirements of users. Due to the reasons for the online calculation, the actual requirements of calculation less meet the time complexity. Based on the above overview for the overall structure of recommendation system, the record module of collecting user information, the algorithm will get the history access user information And each user access over the goods in the target business website commodity mapping table weight addition, the weights of the first K recommendation to target users.

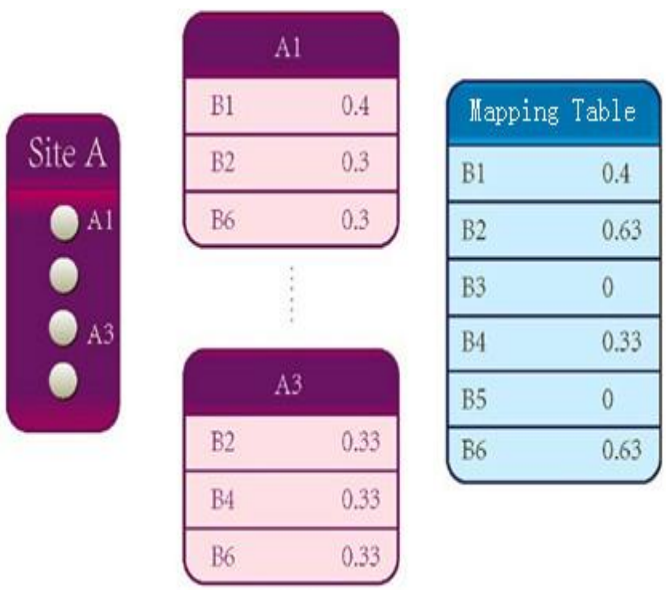

Figure 2. Recommended Mapping Table

For example, the Figure2 of the users, according to the record of his preference in A products including A1 and A3, A1 and A3 query recommendation mapping tables in the target site B weight, you can get a list of weights above. First assume that the user α visit the B website, the B corresponding to add weight of goods can be obtained for the α user recommendation mapping table, according to the actual situation of the recommendation list length, the system will recommend products are B2, B6, B1, B4.

4. Experimental Analysis and Results

4.1. Dataset Partition

Electricity providers in the existing two datasets and multiple datasets electricity supplier, to need to partition the training set and test set. In traditional recommendation system, the shield is according to a certain proportion of randomly selected test set information, and adopting this information to Assessment algorithm of quality.

However, in this chapter, multi-B2C crossing ranking recommendation algorithm still need to use electricity supplier other than the target behavior information. As the test set can be selected from the ranking of users, to adopt the following strategies: dataset partition will separate in the not ranking user electric A or electric B record behavior information of all classified in the training set; for the random division ranking users into two parts of 50%-50%, the part under the training set, the other as a part of the test set. This division results guarantee as part of ranking user training integration points, the other part is non ranking test set user, user types all ranking the user. Assuming the information of training set for recommendation system known information in the absence of any other knowledge, while the test set are carried to each user's information, The whole algorithm simulation test set for users to carry part of the information to predict the user to be interested in commodities. Electricity supplier for the two datasets for each of the training set - test set can be separately.

4.2. Correlation Algorithm

Compared with A crossing ranking recommendation algorithm, We also calculated two insensitive to the cold start recommendation algorithm. It is A Global ranking algorithm and random ranking recommendation algorithm. Global ranking algorithm recommended strategies according to the popularity of goods, the goods in accordance with the degree descending its arrangement. This algorithm has very low complexity, but it is a completely non personalized recommendation algorithm. At present, this recommendation algorithm of e-commerce sites cannot use the user information to achieve cold start. Random sort of recommendation algorithm from the internal target business website all goods produced randomly number of commodity, it is recommended to the target user. Obviously the disadvantages of the two algorithms, Unable to obtain the user information will not achieve the effective application, in a completely "cold start" environment cannot make recommendations satisfactory results.

4.3. The Experimental Results of One-to-One Crossing Ranking Recommended

In the two group was trained on two electric dataset and test set, Using the training set to generate a list of recommended settings recommended mapping table s=20, the largest number of m=5 threshold resource allocation. In the selection of distribution times threshold parameters, it has been confirmed experiments on G-M dataset, 82.90% of the goods can be in m=5 threshold F to obtain 20 commodity corresponding mapping table, and received 20 commodity non zero resources required for 1.67 times the average resource allocation. In M-S dataset, More than 93.21% of the product can be the threshold value F obtained 20 products within a corresponding mapping table, the resource of average distribution for 1.48 times

4.3.1. The Accuracy Rate and Recall Rate: In figure 3 and figure 4 Experimental results show that the crossing ranking recommendation algorithm has huge upgrade in two accuracy's indicators than outside the station information cold-start recommendation. On the M-S dataset with the K value change, Recall at $2.05*10^{-4} \sim 4.79*10^{-3}$ range, when recall is at $2.43*10^{-2} \sim 9.30*10^{-2}$ range, the accuracy obtained ascension of thousand. The accuracy improvement actually can be seen as the embodiment of the value of information outside the station.

According to the Global Ranking algorithm simply from the probability analysis, in the process of random test sets, probability selection from these hot commodities higher, which resulted in the off-line evaluation accuracy condition, Global Ranking of the currently used for "cold start" recommendation algorithm performs better. Even, we still can see obviously, not only the Crossing ranking in terms of accuracy than Random Ranking has greatly improved, but also focus on specific data, accuracy index close to or even more than the Global Ranking algorithm. Thus, the Crossing ranking algorithm can provide the recommended results of high precision cross category in this cross recommendation. Below we will in some inexact index of algorithm for further evaluation.

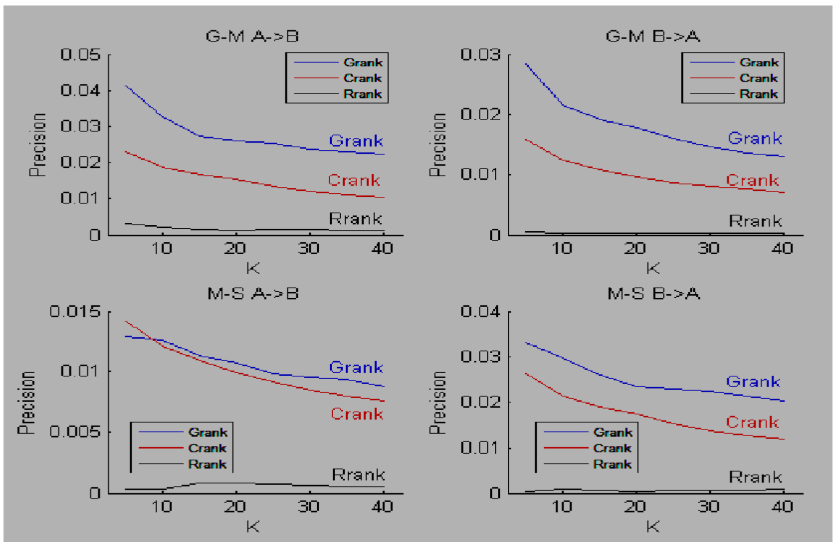

Figure 3. Comparison of Two Multi-B2C Dataset Accuracy

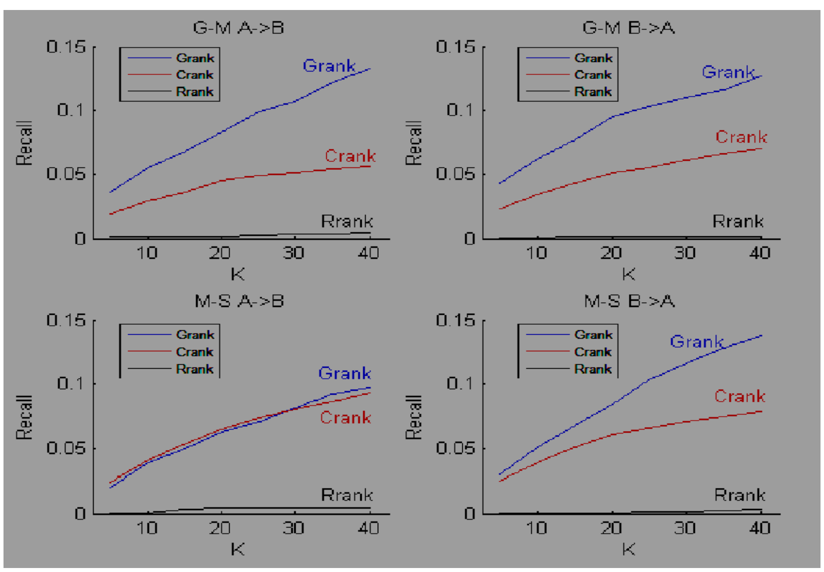

Figure 4. Comparison of Two Multi-B2C Dataset Recall

4.3.2. Average Degree: And similar and accurate evaluation algorithm, in terms of novelty using the same training set and testing set G-M and M-S, Calculation of the average index of 4 recommended each other in electric A and electric B. The abscissa is recommended list length K, K in the range from 5 to 40. In general, the recommended products of average degree of recommendation system smaller novel better, smaller average degree that recommendation system will not give users who are especially hot commodity, recommended popular products should have a greater role than the recommended hot commodity. If the system recommended a very popular item to the user, although accuracy can be very high, but the user may have bought or understanding to these projects through other channels, the user does not believe that such a recommendation system is valuable. The experimental results of Figure5 shows, between commodity tail characteristics Random Ranking random selection of goods is mostly non hot commodities, the average index is far smaller than other two cold start algorithm. While the Crossing ranking about Global Ranking 1/3, Crossing ranking recommendation algorithm in the new Ranking has more advantages than Global ranking.

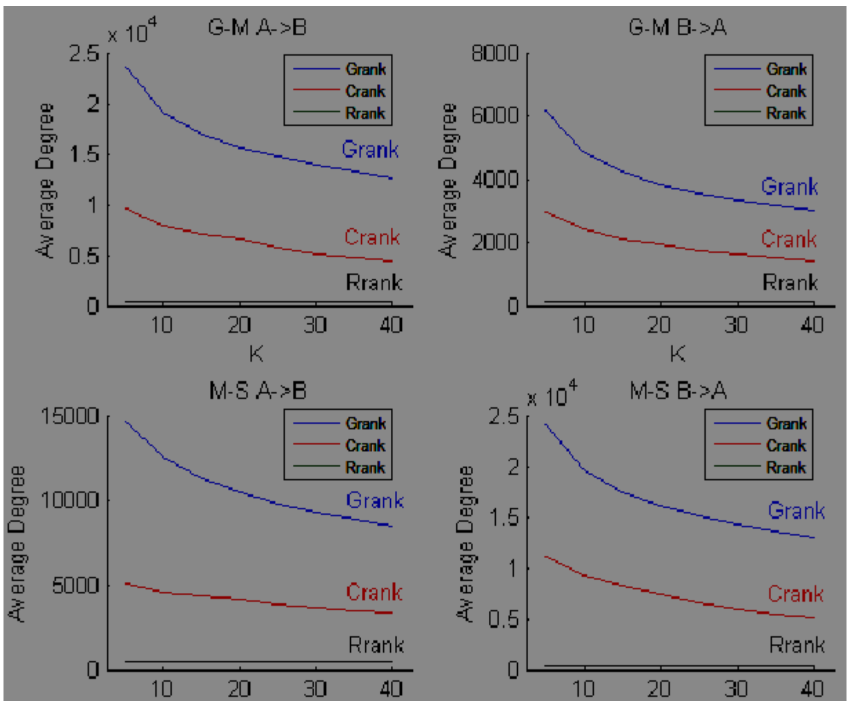

Figure 5. Comparison of Two Multi-B2C Dataset Average Degree

4.4. The Experimental Results of Many-to-One Crossing Ranking Recommended

The effectiveness of using multiple electric dataset experimental algorithm, and verify the effect of recommendation system gets information for recommendation accuracy. Along with the increase of system to obtain information quantity we can be more accurate and comprehensive understanding of user preferences, As we known information of users is introduced, the verification algorithm can give more accurate recommendations. The use of multiple electric dataset can influence the amount of information to verify accuracy of the Crank algorithm, and then the corresponding conclusions are given by comparing the related recommendation algorithm. Not only the different types of website of business analysis will affect the accuracy of algorithm, and the site operation type will also have a certain impact on the algorithm.

The Accuracy Rate and Recall Rate: In setting the size of the mapping table recommended S = 10, the maximum number of resource allocation threshold M = 5, the product of about 90% can be established in this manner recommended by the mapping table. Based on this foundation recommended mapping table, the four electric were cold start recommendation simulation. Using the recall rate and precision evaluation of the accuracy of Crank algorithm. At the same time compared with Grank algorithm and Rrank algorithm. In the limited space, select X3-YD and X4-XT two groups of simulation results are described in detail.

Many to one in Figure 6 can be seen in the experimental accuracy recommended. First of all, from 1for1, 2for1 to the case of 3for1, the actual number of different types of electricity providers that we were increased from 1 to 3, gradually improve the accuracy of Crank algorithm. For example in the recall rate were 0.06, 0.09 and 0.14, eventually more than Grank algorithm in the same circumstances. This shows that with the increase in the amount of information recommended accuracy will be getting higher and higher. Secondly, the Grank algorithm for the three user recommendation accuracy is not consistent, and gradually decline. This trend shows the severe cross user (both behavior in multiple business users compared to nonusers) cross, for the popular goods aren't particularly interested in, which is reflected in reduced accuracy index. Virtually all the reality of the users can not only browse a single or a few sites, Users in a certain extent are severe cross users or potential users. With the increasing amount of information recommendation system access, the Grank algorithm this seemingly will get better, recommendation accuracy will not meet the real needs of users. We believe that the recommended system under real conditions can get beyond popular recommendation accuracy. Of course, the Crank algorithm will still maintain good novelty and diversity.

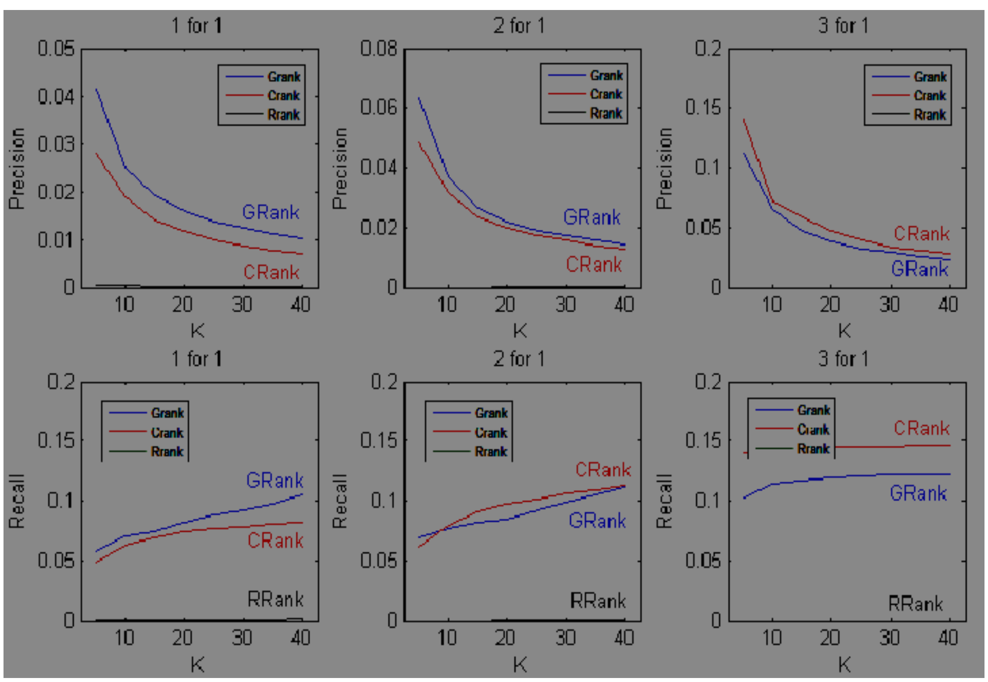

Figure 6. Comparison of X3-YD Accuracy Rate and Recall Rate

In Figure 7, with X4-XT as the target site to repeat the experiment. And there is no difference on the conclusion, but in the accuracy index Crank algorithm achieved greater advantages in the 1for1 simulation results have far exceeded Grank algorithmic. The above results further show that Crank algorithm successfully solves the problem of cross recommended a plurality of vertical type electric different types of goods. And much higher than the Rrank algorithm in the accuracy, Influenced by the data set characteristics and site factors, even more than the Grank approximation algorithm, At the same time, the users can increase the amount of information to carry the recommendation results more accurate algorithm.

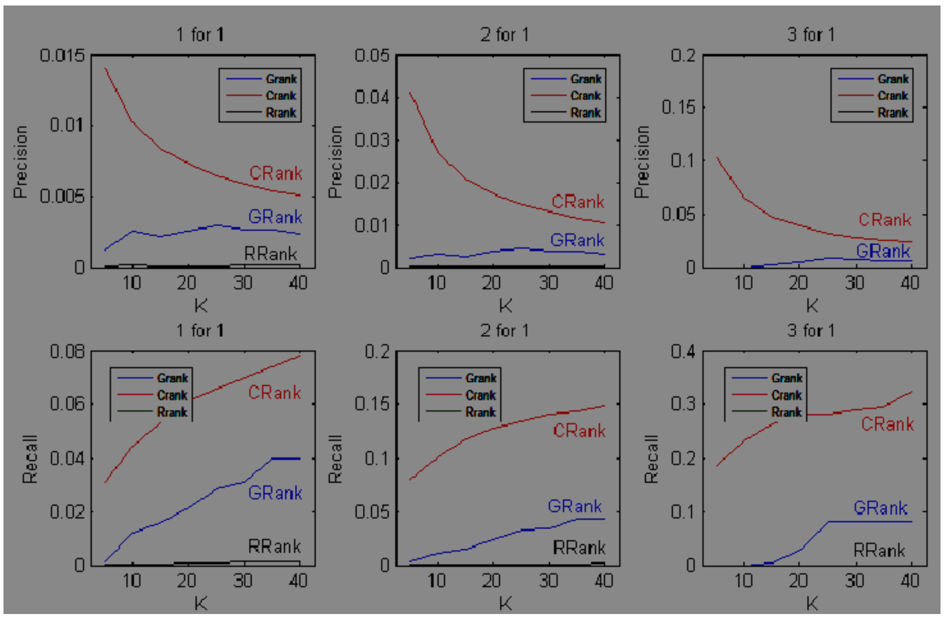

Figure 7. Comparison of X4-XT Accuracy Rate and Recall Rate

5. Conclusion

Cold start is the most difficult problem in information recommendation. Although there is no known information of the user, website can gives to the user according to popularity crossing ranking recommended, or simply randomly generated some recommendation results, but the former can't fully reflect the personality of the thought, the accuracy is very poor. In this paper, the user visits to cross behavior information by some different types of electronic commerce website. For a target website, no information users may have a considerable amount of access history at the other sites. We can use this information to create a considerable accuracy and personalized recommendation. In addition, to introduce more foreign station's data, the accuracy of this algorithm can be further improved. Through the weight considering different behavior, it increases the user attribute information and commodities labeling information. We can design a more accurate method.

References

[1] L. Zhu, F. Chung and S. Wang, "Generalized Fuzzy C-Means Clustering Algorithm With Improved Fuzzy Partitions", IEEE Transactions, vol. 39, **(2009)**, pp. 578-591.

[2] J. MacQueen, "Some methods for classification and analysis of multivariate observations", The 5th Berkeley Symposium on Mathematical Statistics and Probability, **(1967)**, pp. 281-297.

[3] S. Deerwester, S. T. Dumals and G. W. Furnas, *et al*, “Indexing by latent semantic analysis”, Joumal of the American Society for Information Science, vol. 41, no. 6, **(1990),** pp. 391-407.
[4] G. Linden, B. Smith and J. York, “Amazon.com Recommendations Item-to-Item Collaborative Filtering”, IEEE Internet Computing, , vol. 7, no. 1, **(2003)**, pp. 76-80.
[5] Z. Huang, D. D. Zeng and H. Chen, “Analyzing consumer-product graphs: Empirical findings and applications in recommender systems”, Management Science, vol. 53, no. 7, **(2007),** pp. 1146-1164.
[6] H. Ebel, M. Lutz-Ingo and S. Bornholdt, “Scale-free topology of e-mail networks”, Phys. Rev. E, vol. 66, **(2002)**, pp. 03510.
[7] Y. Zhang, M. Medo and J. Ren, “Recommendation model based on opinion diffusion”, Europhys Lett, vol. 80, **(2007)**, pp. 68003.
[8] T. Zhou, Z. Fu and B. Wang, “Epidemic dynamics on complex networks”, arXiv:cond - mat/ 0405258, vol. 16, no. 5, **(2006)**, pp. 452-457.
[9] T. Zhou, J. Ren and M. Medo, *et al*, “Bipartite network projection and personal recommendation”, Phys Rev E, vol. 76, **(2007)**, pp. 046115.
[10] S. Sen and J. Vig, “Connecting users to items through tags”, In Proceedings of 18th international conference on World wide web, WWW’09, **(2009)**, pp. 671-680, New York, NY, USA.
[11] Y. Zhang, M. Blattner and Y. Yu, “Heat conduction process on community networks as a recommendation model”, Phys Rev Lett, **(2007)**, pp. 99-120.

Authors

Duan Xueying, she is a lecturer at college of Jilin Police. She is in the research of Internet monitoring.

