An Agricultural Information Recommendation Model Based on Collaborative Filtering

Shuilong Zou

School of Information, Nanchang Institute of science & Technology, Nanchang, 330108, China

Abstract. With the rapid development of internet, all kinds of information increases exponentially. It is different for users to find their requisite information because of the widely distributed information of internet. In ordered to provide timely, conveniently and efficiently to farmers the interested agricultural information and knowledge, this paper proposes an agricultural information recommendation system model based on collaborative filtering. The model uses terms clustering to construct user evaluation matrix of no-absence, and standardizes rating before evaluation prediction in accordance with farmer behaviour difference. Experiment shows that recommendation model has good effect, and realizes information recommendation at a certain extent.

Introduction

In the process of agricultural information service, and how to let farmers get their concerned information on internet is an important work. Due to limitation of the current farmer's cultural quality, they have not understood computer technology very well, especially, they have lack the ability of information filtering and knowledge judgments. Agricultural information collaborative recommendation system can predict level of interest of objective users through generating the nearest neighbour set, find information users' interests, and provide actively and timely user with required information, so as to provide personalized agricultural information for users.

Recommendation System

Recommendation system is defined as process that it provides commodity information and advices for users in electronic business system, helps customer decide to purchase goods, and simulates salesman to implement purchase [1]. In short, it recommends items to users in accordance with different customer requirement.

Based on implementation technology, recommendation system is classified by recommendation system based on content filtering, the recommendation system based on collaborative filtering and the recommendation system based on data mining.

The Recommendation System Based on Content [2]

The recommendation system based on content filtering is that it implements recommendation operation in accordance with item contents and similarity of user interests. Its basic idea is recommendation process; firstly, using every user interest model to construct data structure and describe their preferences, secondly, implementing feature extraction to every items content and forms the eigenvectors; thirdly, when one user is recommended, user interest model of user will be compared with feature matrix of all items, and get similarity of items, finally, implementing recommendation through similarity。

The Recommendation System Based on Collaborative Filtering [3]

The recommendation system based on collaborative filtering is the most successful recommendation technology application, which looks for the similar user of the specify user, synthesizes the one item evaluation of similar user in users set, and forms rating prediction of the specify user to the item. The process of recommendation system based on collaborative filtering is divided into three parts: the user interest file; calculating the similarity of current user and other user, and generating the nearest neighbour set.

The Recommendation System Based on Data Mining [4]

With incensement of user and item number, a lot of user trading data, user register data and user rating data can be collected. Meanwhile, the server also saves lots of log data, shopping basket information, etc. In the face of these huge amounts of data, in ordered to improve the quality and efficiency of recommendation system, all kinds of data mining technologies (such as, association rules, sequential patterns mining, clustering, classification, etc) have been widely applied in the recommendation system.

   The recommendation system based on data mining mainly involves three key technologies: original data processing of customs, the establishment and expression of pattern library, application of data miming algorithm.

The Mixed Recommendation System [5]

Owing to the recommendation system based on content filtering and recommendation system based on collaborative filtering have their own disadvantages, some researchers propose hybrid recommendation model. The hybrid recommendation model cannot saves advantages of the above two recommendation systems, but also it is better than the recommendation system based on content filtering and recommendation system based on collaborative filtering in the recommendation effect.

The Agricultural Recommendation System Based on Collaborative Filtering.

This recommendation system is subsystem of integration information service platform of Hunan province. The system aims to evaluate information of platform based on user, calculate interest similarity of current user and other user, generate the nearest neighbour set, and so as to recommend information to users. The overall recommendation framework is shown in Fig1.
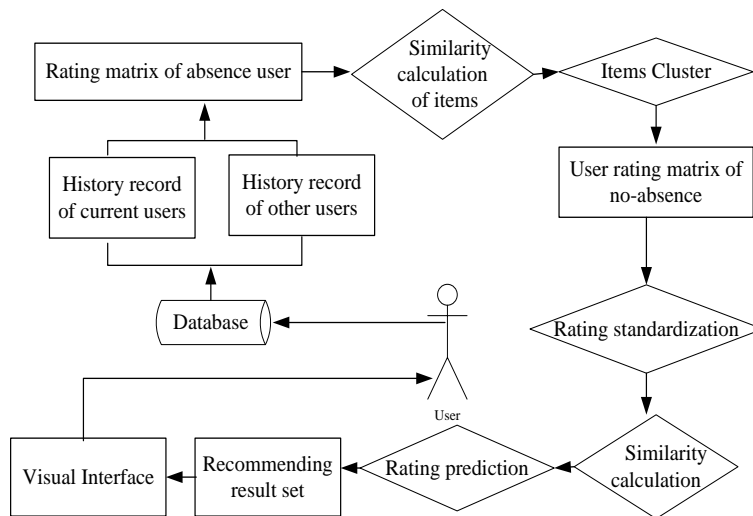

Fig.1:Agricultural Recommendation model based on collaborative filtering

The No-missing Evaluation Matrix.

   In ordinary collaborative filtering recommendation model, the user evaluation matrix must firstly be constructed in accordance with historical information; Then calculating the user interest similarity and generating neighbour set; Finally, information which the user might be interested be pushed on the page. The specific process of the module is shown in Fig.2.
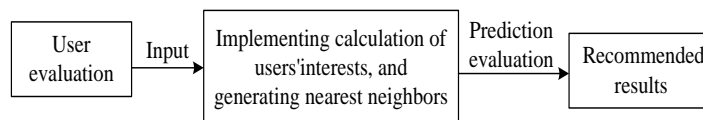

Fig.2: The flow figure of collaborative filtering

   According to sparseness problems of the collaborative filtering recommendation system, there are two solutions:

i) Using one method to implement sparse matrix, so as to reduce sparse degree of the matrix.

ii) Setting unevaluated item to be a fixed default value, such as average value. The two methods are both some disadvantages, they cannot sufficiently reflect personalized interests and preference of user, and solve data sparseness of collaborative filtering algorithm [6].

Therefore, in ordered to effectively solve the sparseness of user rating data, similarity of agricultural formation should firstly be determined before calculating the user similarity, then implementing the cluster and find the nearest neighbour of the same as clustering information, finally, predicting rating of the objective information in accordance with rating of the nearest neighbour information, and constructing a rating matrix. The flow is shown in Fig.3.
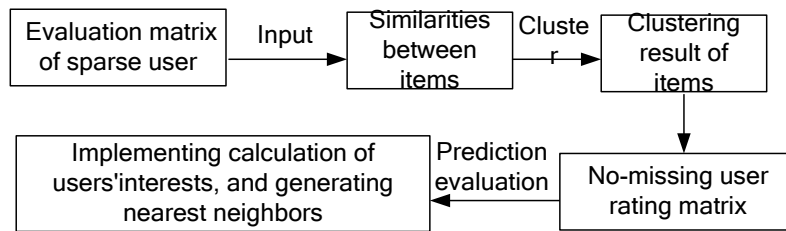

Fig.3: Flow of collaborative filtering module

Let the item union of user i and user j rating to be U_{ij}, rating item of the user i to be K_i, that is:

$$U_{ij} = K_i + K_j \qquad . \qquad (1)$$

Unevaluated item of the user i is N_i, that is:

$$N_i = U_{ij} - K_j. \qquad (2)$$

Rating of N_i can be predicted by the similar rating of user i and user j. For objective item $r \in N_i$, prediction rating of item r can be represented by P_{ir}.

i) Calculating similarity of items. For calculating the similarity of item n and item r, firstly, calculating the user rating of item n and item r, then calculating the similarity $\sin(r,n)$ of item n and item r by similarity measure.

ii) Implementing items clustering. There are many clustering methods, such as k-means algorithm, Clara algorithm, etc. This paper measures Euclidean sum of the squares with component connection method *etween–groups Linkage*. After finishing clustering, Finding the nearest neighbour of objective item r, and take item set of the most close similarity as neighbour of item r, it is represented as $H_r = \{K_1, K_2, ..., K_n\}$, $r \notin H_r$. And the similarity of K_1 is the most close, and so on.

iii) After getting H_r, we use prediction method, which is proposed by Sarwar [7], that is:

$$P_{ir} = \frac{\sum_{n \in H_r} \sin_{r,n} \times R_{i,r}}{\sum_{n \in H_r} |\sin_{r,n}|} \qquad (3)$$

Where $\sin_{r,n}$ represents item r, $R_{i,r}$ represents rating of item n of user i.

Find the Nearest Neighbour Set.

After getting no missing user evaluation matrix, this model calculates interest similarity of objective and other users with similarity measure method, selects top n user of the closest similarity as the nearest neighbour set objective user. At present, there are some similarity methods, such as Person correlation coefficient, cosine correlation and modified cosine correlation, we calculate the interest similarity of user with Person correlation coefficient.

The User Rating Standardization [8]

Because user's rating is associated with the user's own personality, rating of conservative user may be lower than the deserved rating of item, while optimistic user might give the higher rating than actual rating. Therefore, rating of user should firstly be standardized. The standardizing formula is as follow:

$$P_{k,m}(X_{a,b}) = X_{a,b} - (\overline{X_a} - \overline{X_k}) - (\overline{X_b} + \overline{X_m}) \quad \cdots\cdots\cdots\cdots \quad (4)$$

$P_{k,m}(X_{a,b})$ is standardized function, $X_{a,b}$ is rating that $User_k$ rates $Item_b$, $\overline{X}_a$ is rating average value of the $User_a$, $\overline{X}_k$ rating average value of the $User_k$, $\overline{X}_b$ rating average value of the $Item_b$, and $\overline{X}_m$ rating average value of the $Item_k$

Generating Recommendation Result.

After getting the nearest neighbour set, then predicting the user's interest of objective user, and generating recommendation result information. Generally, according to object of recommendation, there are various kinds of recommendations methods. In this paper, the agricultural information recommendation model recommends potential item set of for objective user with Top-k.

The Test of Recommendation Model

In ordered to test the recommendation method of the agricultural information recommendation model and verify the superiority of the agricultural information model, we firstly collect original evaluation data and implement preprocess and implement the corresponding experimental test, so as to obtain a sparse evaluation matrix. The initialized matrix of user evaluation uses 5 items and 10 users. The detail information of the evaluation matrix is shown in table 1.

Table.1: Original matrix of user evaluation

	Item1	Item2	Item3	Item4	Item5
User1	-	4	-	4	2
User2	3	-	4	5	-
User3	-		-	-	3
User4	1	1	-	-	5
User5	3	-	2	-	1
User6	-	-	-	-	3
User7	4	-	-	-	-
User8	-	4	2	-	-
User9	5	-	-	-	-
User10	-	-	4	-	2

Calculating the similarity of items with the Pearson correlation coefficient, and then getting cluster tree of items with clustering. The result is shown in Fig.4.
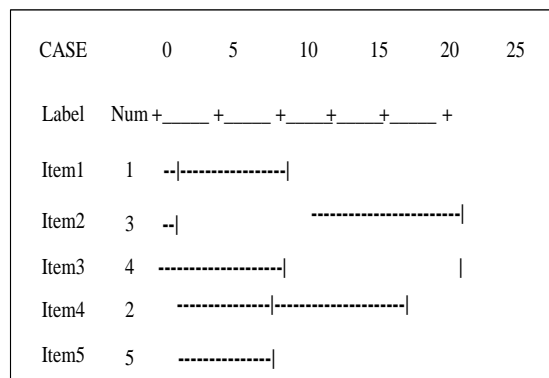


```
CASE          0     5    10    15    20    25

Label   Num +_____+_____+____+____+_____+

Item1    1    --|-----------------|

Item2    3    --|                    -----------------------|

Item3    4    ------------------|                |

Item4    2    ---------------|--------------------|

Item5    5    ---------------|
```

Fig.4: Clustering result

Therefore, the clustering result can be concluded:

$$H_{item1} = \{K_{item3}, K_{item2}, K_{item5}\}$$
$$H_{item2} = \{K_{item4}, K_{item3}, K_{item1}\}$$
$$H_{item3} = \{K_{item1}, K_{item2}, K_{item5}\} \tag{5}$$
$$H_{item4} = \{K_{item4}, K_{item2}, K_{item5}\}$$
$$H_{item5} = \{K_{item2}, K_{item4}, K_{item1}\}$$

According to above formula (5), we can get the missing user evaluation matrix, as shown in table 2.

Then, the similarity of users can be obtained: $U_1 = \{U_6, U_3, U_4\}$; $U_2 = \{U_4, U_6, U_7\}$, $U_3 = \{U_6, U_1, U_4\}$, $U_4 = \{U_2, U_6, U_1\}$, $U_5 = \{U_9, U_8, U_1\}$, $U_6 = \{U_4, U_2, U_1\}$, $U_7 = \{U_9, U_2, U_4\}$, $U_8 = \{U_5, U_2, U_4\}$, $U_9 = \{U_7, U_2, U_4\}$, $U_{10} = \{U_1, U_6, U_4\}$. After implementing standardized operation, getting standardized evaluation matrix, and implementing evaluation prediction for any items of any user, then generating the final recommendation result in accordance with TOP-N Principle.

Table.2: No-missing user evaluation matrix

	Item1	Item2	Item3	Item4	Item5
User1	5	4	3	4	2
User2	3	3	4	5	5
User3	5	5	5	5	3
User4	1	1	3	4	5
User5	3	0	2	2	1
User6	1	1	2	2	3
User7	4	3	1	0	2
User8	0	4	2	1	1
User9	5	4	1	0	3
User10	1	0	4	1	2

Conclusions

This paper takes processing of optimizing recommendation system based on collaborative filtering, implements item clustering in accordance with similarity of items, and predicts missing rating. It solves the problem of sparseness of user evaluation matrix a certain degree. In addition, it implements standardized processing before prediction, and overcomes the evaluation difference because of user personality difference, so as to make rating more referenced.

Acknowledgements

References

[1] Resnick. Varian Recommender System.Communications of the ACM, vol,40,3(1997):56-58.

[2] GE. R.X. Study on Collaborative Filtering Recommendation System Based on Content Clustering, Master' s Degree Paper of Shandong Normal University,2008.

[3] T.W, PENG.Y.Q. Review of Collaborative Filtering Technology Improvement Based on Bussiness Application, Computer Engineering & Science. vol,30,10(2008):61-63.

[4] WU.Y.W, HUANG.L. Automatic Recommendation of Hybrid algorithm for Learning Resource. Computer Applications and Software. vol,25, 12(2008):202-203.

[5] DING LING.G.Y. Implement study and Realization on Recommendation System Based on Hybrid algorithm. Master's Degree Paper of Shanghai Tongji University, 2008.

[6] ZHOU. L.J, XU M.S, ZHANG, Y.Y. Course Recommendation Model Based on Collaborative Filtering . Application Research of Computers, vol,27,4(2010):1315-1318.

[7] DEB K.Muti-objective Genetic Algorithms Problem Difficulties and Construction of Test Problems,Evolutionary Compution, vol,7(1997): 205-230.

[8] Jun Wang Arjen P.de Vriesl, Marcel J.T. Reinders Unifying Userbased and Item based Collaborative Filtering Approaches by Similiary Fusion[C]// Proc 29 th ACM SIGIR Conference on Information Retrieval, 2006.