

The Practical Teaching Design of Recommendation System Based on Collaborative Filtering Algorithm

Panpan Liu

Kunming University of Science and Technology
Key Laboratory of Pattern recognition And Intelligent
computing of Yunnan College
Kunming, China
lpp.wk@163.com

Zhengtao Yu

Kunming University of Science and Technology
Key Laboratory of Pattern recognition And Intelligent
computing of Yunnan College
Kunming, China
ztyu@hotmail.com

Cunli Mao

Kunming University of Science and Technology
Key Laboratory of Pattern recognition And Intelligent
computing of Yunnan College
Kunming, China
maocunli@163.com

Jianyi Guo*

Kunming University of Science and Technology
Key Laboratory of Pattern recognition And Intelligent
computing of Yunnan College
Kunming, China
gjade86@hotmail.com
*Corresponding author

Yantuan Xian

Kunming University of Science and Technology
Key Laboratory of Pattern recognition And Intelligent
computing of Yunnan College
Kunming, China
yantuan.xian@gmail.com

Huafeng Li

Kunming University of Science and Technology
Key Laboratory of Pattern recognition And Intelligent
computing of Yunnan College
Kunming, China
lhfiww@126.com

Abstract—In order to resolve the implementation about the decision support system in "Decision Analysis and Decision Support" course for postgraduates. We introduced the design of recommendation system into the teaching practice. This paper describes the principles of recommendation system based on the collaborative filtering algorithm and introduces the process of the system. Through the practice of a practical problem, students can realize a recommendation system and understand the process of the decision support system.

Keywords—Intelligent decision; Recommendation system; Collaborative filtering; Practical teaching

I. INTRODUCTION

The design of the decision support system is the ultimate goal of the "decision analysis and decision support" course for postgraduate students. The recommendation system is one of an intelligent system. Currently, in teaching of the course, there is no special practice teaching design about the intelligent decision support system. This paper introduces the design of recommendation system into the practice teaching and gives a teaching experimental scheme. Therefore, students can understand the process of the intelligent decision support system and their practice ability can be improved.

With the rapid growth of the Internet data, information overload is getting worse. Recommendation system is an effective method to solve the information overload problem.

The results of the recommendation system are in line with the individual needs of users and can help people make effective decision. At present, there are several methods to solve the recommending problem: the content-based recommendation algorithms; the Network Structure-based recommendation algorithm; the collaborative filtering recommendation algorithm, which is the most widely used in recommending systems [1,2], has good performance in the recommending systems of books, movies, music and others [3,4]. But for the extreme sparseness of users score data, the similarity measure method can not effectively calculate the target user's nearest neighbors [5], so recommendation quality can not be guaranteed. To the question of the data sparseness of posts recommended, on the basis of traditional collaborative filtering algorithm, a new collaborative filtering algorithm incorporating PageRank algorithm is proposed in this paper. In this teaching practice, students are required to design a recommendation system based on the new collaborative filtering algorithm and apply it to the BaiDu PostBar recommendation.

II. THE PRINCIPLES OF RECOMMENDATION

The process of the recommendation system based on collaborative filtering algorithm fusion of PageRank is as follows: Firstly, obtaining the data set from Baidu PostBar; Secondly, calculating the PageRank score of each user; Thirdly, calculating the similarity between the users with high PageRank score and the target users by Pearson correlations,

calculation the score of posts combining with each user's PageRank score; Finally, obtaining the recommendation results which have high score. The process of the recommendation system is shown in figure 1.

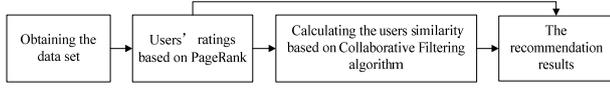


FIGURE 1. THE PROCESS OF THE RECOMMENDATION SYSTEM

A. Users' Ratings based on PageRank

Users' ratings are obtained through users' comment behavior. When the reply user is important, more people will reply the user and more time will be used to reply the user. So a PageRank algorithm fusing the users' comment behavior is used to score users participating in this post. The computation of users' ratings is shown as formula (1).

$$p(i) = \frac{d}{n} + (1-d) \times \sum_{u_j} r(u_i, u_j) \times p(j) \quad (1)$$

$p(i)$ represents the PageRank rating of user i ; d is the damping coefficient and its value is between 0 and 1, usually the value is 0.15; N represents the number of people participating in this post. u_i is user i and u_j is user j ; $r(u_i, u_j)$ represents the reply relations between user i and user j and it is defined as formula (2).

$$r(u_i, u_j) = \frac{M_{i,j}}{\sum_{u_k} M_{i,k} + \lambda} \quad (2)$$

In formula(2), $M_{ij} = \alpha \times M_{i,j}^r + \beta \times M_{i,j}^t$, $M_{i,j}^r$ represents the ratio of reply times that user j replies to user i to the total reply times of user j . $M_{i,j}^t$ represents the ratio of the reply time that user j replies to user i to the total reply time of user j . The reply time of user j is defined as: the difference between the last and the first time that user j reply to user i . The total reply time of user j is defined as: the difference between the last and the first time that user j participate in this post.

B. The Calculation of User Similarity

The Pearson correlations [6,7] is used to calculate users' similarity in this paper, as shown in formula (3).

$$sim(i, j) = \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in U} (r_{u,i} - \bar{r}_i)^2} \sqrt{\sum_{u \in U} (r_{u,j} - \bar{r}_j)^2}} \quad (3)$$

In the formula(3), $sim(i, j)$ represents the Similarity between user i and user j ; $r_{u,i}$ represents the post u 's rating from user i ; $r_{u,j}$ represents the post u 's rating from user j ; \bar{r}_i

represents the average rating of the posts in group U that user i gives; \bar{r}_j represents the average rating of the posts in group U that user j gives. The unread posts of the target user are scored, after the Similarity between the target user and the neighbor user is calculated.

C. The Collaborative Recommendation Fusing PageRank

The scoring formula of the traditional collaborative recommendation is shown in formula(4).

$$Score = \bar{r}_i + \frac{\sum sim(i, j) \times (r_{u,j} - \bar{r}_j)}{\sum sim(i, j)} \quad (4)$$

In the formula (4), $Score$ represents the score of one post; \bar{r}_i is the average score of the posts in the group U from the target user i ; \bar{r}_j means, the average score of the posts in the group U from the the neighbor user j ; $r_{u,j}$ is average score of the recommended posts from the neighbor user j ; $sim(i, j)$ is the similarity value between the target user i and user j .

The traditional collaborative filtering recommendation algorithm ignore the importance of the users. The importance of users is combined to the traditional collaborative filtering algorithm. The users' PageRank ratings is used to the formula(4). For a certain post, the recommendation value is high when the important users pay close attention to the post. When a post only have unimportant users concerning about, the recommendation value of this post is low. Addition and multiplication are used as the fusion method, due to the fast fusion speed and good real-time performance. The scoring formula of the collaborative filtering algorithm fusing PageRank is shown in formula (5).

$$Score = \bar{r}_i + p(i) + p(j) \times \frac{\sum sim(i, j) \times (r_{u,j} - \bar{r}_j)}{\sum sim(i, j)} \quad (5)$$

In the formula (5), $Score$ is the score of one post; \bar{r}_i is the average score of the posts in the group U from the target user i ; \bar{r}_j means average score of the posts in the group U from the neighbor user j ; $r_{u,j}$ is the average score of the recommended posts u from the neighbor user j ; $p(i)$ is the target user i 's PageRank rating; $p(j)$ is user j 's PageRank rating; $sim(i, j)$ is the similarity value between the target user i and neighbor user j . Then, the posts with high score will be recommended to the target user.

III. THE RECOMMENDATION RESULTS

A recommendation system is realized based on the algorithm proposed above with Java programming language and adopt the popular Web Development Framework named django. A document-oriented database named CouchDB is

used. Chrome browser is installed on the server and plug-in is imported. The experimental data set is acquired by the web crawler from the Baidu Post Bar.

The poster, respondents, post time, symbol of reply and reply time comprise the data records.5261 posts or reply records were obtained. When the data is ready, the system firstly do the Data Preprocessing and the extraction of user data set, including 572 users who delivered or replied the post. Then, scoring all users based on PageRank, a part of records are shown in Table 1.

TABLE I. THE USERS' RATING BASED ON PAGERANK

Order number	User name	Ratings based on PageRank
1	雨轩觉	8.564949
2	胖谷爱明器	7.0483512
3	美夕子	6.186294
4	拉小困	5.645916
5	拖棒棒主	5.386274
6	白面玫瑰十一	5.145831

The algorithm model proposed in this paper is applied to Baidu Post Bar recommendation system. Before the algorithm model applied, the posts is cluttered and it is difficult to find the popular and interesting posts.

After using the algorithm model, the results of the recommendation system are shown in Figure 2.

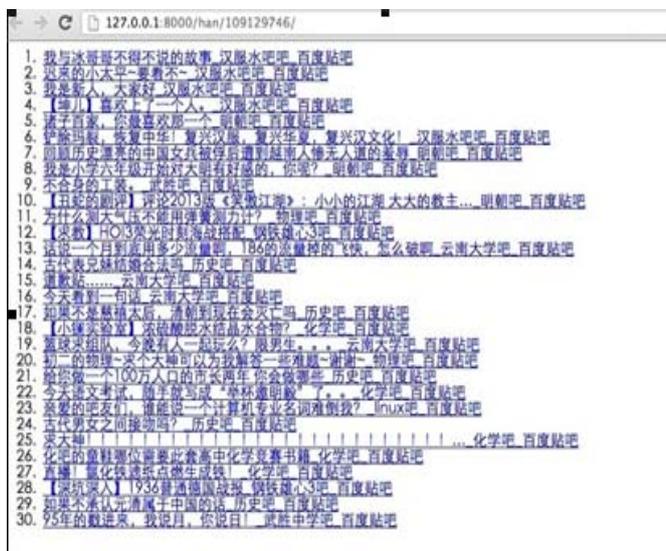


FIGURE II. THE RESULTS OF THE RECOMMENDATION SYSTEM

Due to the collaborative filtering recommendation fusing users' PageRank ratings, the information of posts is more concentrated, the redundant information is reduced, and the problem of information overload is solved.

IV. SUMMARY

The design of the recommendation system is introduced into the practical teaching which can help students deeply understand the decision support system and grasp how to design a intelligent decision support system.

ACKNOWLEDGMENT

This paper is supported by the project of postgraduate's Key course of Kunming University of Science and Technology.

REFERENCES

- [1] Shani G, Brafman R, Heckerman D. An MDP-based recommender system[J]. The Journal of Machine Learning Research, 6: 1265-1295, 2005.
- [2] Xiaoyuan Su, Taghi M. Khoshgoftaar. A survey of collaborative filtering techniques[J]. Advances in Artificial Intelligence, 1-19, 2009.
- [3] G. Karypis. Evaluation of item-based top-N recommendation algorithms[C] // Proceedings of the International Conference on Information and Knowledge Management, Atlanta, USA, 2001: 247 - 254.
- [4] M. R. McLaughlin, J.L. Herlocker. A collaborative filtering algorithm and evaluation metric that accurately model the user experience[C] // Proceedings of 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Sheffield, UK, 2004 : 329 - 336.
- [5] Ximeng Zhang, Shengyi Jiang. The Network Forum Personalized Recommendation Algorithm based on Collaborative Filtering Algorithm[J]. Computer Engineering, 38 (5) : 237-256, 2012.
- [6] Resnick P, Iacovou N, Suchak M, et al. GroupLens: an open architecture for collaborative filtering of netnews [C]// Proceedings of the 1994 ACM conference on Computer supported cooperative work. ACM, 1994: 175-186.
- [7] Shardanand U, Maes P. Social information filtering: algorithms for automating "word of mouth" [C]// Proceedings of the SIGCHI conference on Human factors in computing systems. ACM Press/Addison -Wesley Publishing Co., 1995: 210-217