A Kind of Precision Recommendation Method for Massive Public Digital Cultural Resources

A Preliminary Report

Xiaoshuang Zhai[1,2] ,Feng Jin[1,2*],Jian Wang[2*], Yi Yang[2],Qi Yao[2],Qiongwen Qiu[1],Guigang Zhang[2]
1. Department of automation, Beijing Institute of Technology, Beijing, China
2. Institute of Automation, Chinese Academy of Science, Beijing, China
{zhaixiaoshuang90, jinfeng226}@163.com; {guigang.zhang, jian.wang, yangyi}@ia.ac.cn
*Corresponding authors: jinfeng226@163.com;jian.wang@ia.ac.cn

***Abstract*—In this paper, we give a preliminary report regarding a precision recommendation method for massive public digital cultural resources, which help people find the resources they are really interested in. We classify the public digital cultural resources and propose a precision recommendation framework based on big data technology. Finally, we introduce a precision recommendation method for public digital cultural resources.**

Keywords: Public Digital Cultural Resources; Precision Recommendation; Big Data

I. INTRODUCTION

Nowadays, recommendation systems are widely used in many fields. However, in public digital culture field they are still developed in early stage. In order to effectively realize the sharing of public cultural resources, it needs to provide users in different regions with the resources that they are really interested in by using big data analysis and predictive analytics.

The collaborative filtering algorithms have the following common problems: data sparse problem, cold-start problem, and scalability problem. Although the problems can be solved respectively using different methods, the computing workload will be very large for massive public digital cultural resources and large-scale users. For addressing this problem, we propose a social graph-based collaborative filtering recommendation method that can solve the data sparse problem and the cold-start problem in a big data environment.

The rest of the paper is organized as follows. We introduce related work in Section 2. We present a precision recommendation framework in Section 3. In Section 4, we propose a precision recommendation method for massive public digital cultural resources. In the last section, we conclude and the future work.

II. RELATED WORK

Public digital cultural resources [1] are a widely range of concepts. Generally speaking, the public digital cultural resources are the digital resources of the public cultural facilities, e.g., public libraries, public museums, public archives, public galleries. In our research, the range of public cultural resources is extended to the resources on Interment, e.g., resources in social network, in blogs, and in forums. In this paper, the public digital cultural resources contain different types, e.g., literature, movie, music, drama.

Recommendation system [2] is an essential part in personalized service application that becomes very popular in recent years. The recommendation system is defined as a system that provides personalized recommendation or guides users to choose what they really need from massive resources. Recommendation systems have been already applied in many fields, particularly in electronic Commerce [3]. Amazon.com [4] can recommend all kinds of products, in Film and TV industry. Netflix [5] recommends different movies. In this paper, we introduce the recommendation system to the field of public digital cultural resources. Specifically, we develop an improved personalized recommendation, namely, precision recommendation for cultural resources.

Collaborative Filtering Recommendation Algorithms [6] are classified into Memory-based method and Mode-based method. The former method includes user-based and item-based recommendation algorithms. The algorithms are implemented in two steps: first, create similar user/item sets by computing the user/item similarity; then, predict the ratings of items in order to determine the items that should be recommended. The latter method adopts statistics methods or machine learning technologies to models, e.g., Bayesian Network-based collaborative filtering [7], SVD-based collaborative filtering [8], Linear Regression-based method [9], Markov Decision Processes-based method [10].

Although the effectiveness of Collaborative Filtering Recommendation Algorithms is proved in wide range of fields, there are still some questions [2] that researchers keep trying to resolve.

1) Data sparse prolem

The sparseness of data is measured by the proportion of the elements having no score in user-resource matrix. In some open datasets, the sparseness is up to 95%. A couple of methods were proposed for addressing data sparse problem. Breese et al. [7] proposed a method using a fixed default values to replace the ratings, Byeong et al. [11] proposed a method using clustering, and Sarwar et al. [12] presented an approach using dimension reduction.

978-1-5090-2179-6/16 $31.00 © 2016 IEEE
DOI 10.1109/BigMM.2016.46

2) Cold-start problem

There are two different categories of cold-start problems in collaborative filtering recommendation system: user cold-start problem and item cold-start problem. In these cases, we can't compute the similarity or recommend items because of the lack of ratings. Bell et al. [13] introduce the previous solutions for cold–start problems such as average and mode. Zhou et al. [14] proposed a method using matrix factorization similar to SVD. However, there is no such a method that can solve both the user cold-start problem and the item cold-start problem.

3) Scalability problem

Because the scale of users and resources are rapidly growing, the computational complexity of Collaborative Filtering Recommendation Algorithm is getting higher. In this case, the performance of recommendation system is reducing. The recommendation precision for massive public digital cultural resources can hardly be ensured. Many methods like clustering [11], dimension reduction [12], distributed structure are used to solve scalability problem.

III. A PRECISION RECOMMENDATION FRAMEWORK FOR MASSIVE PUBLIC DIGITALCULTURAL RESOURCES

According to requirements of recommendation systems, we propose a framework of a precision recommendation shown in **Error! Reference source not found.**. The framework is composed of six modules, namely, User interaction module, Data collection module, Data analysis module, Data storage module, Precision recommendation module, and Evaluation module.

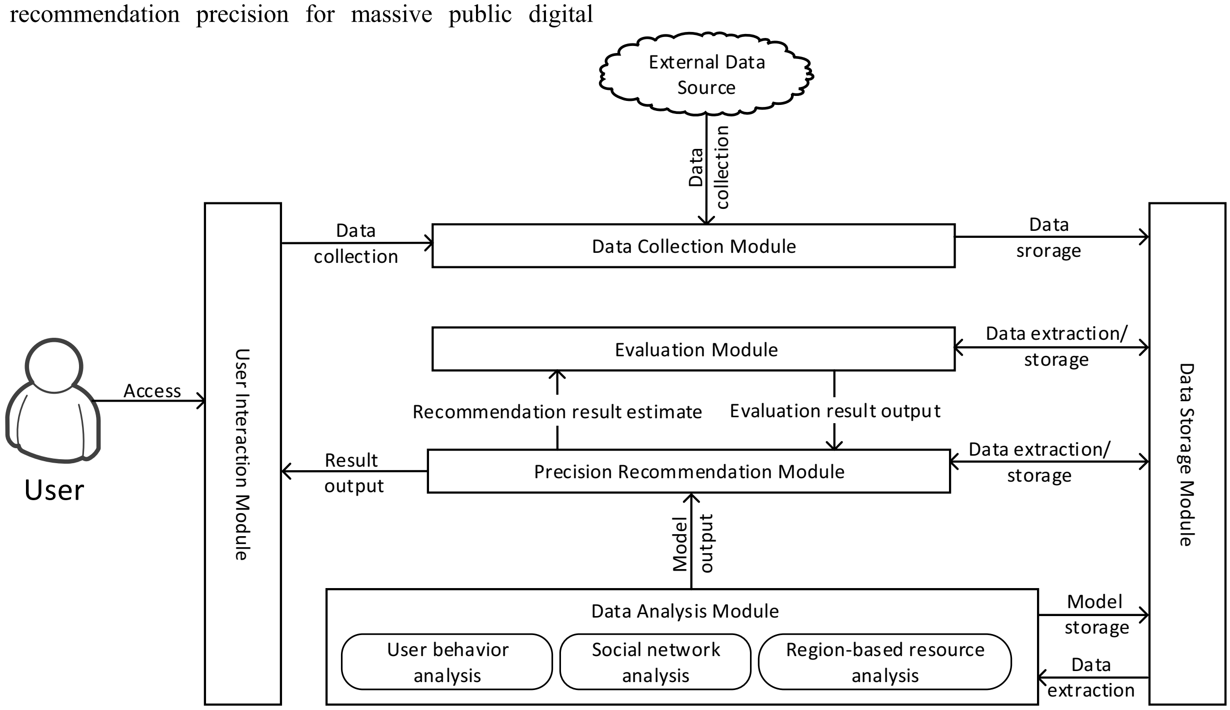

Figure 1.Precision recommendation framework

- **User interaction module**: providing users with UI for accessing, browsing, and displaying the recommendation results. Moreover, users can observe other resources they might be interested in.
- **Data collection module**: collecting the required data for precision recommendation including internal data and external data of the framework. All of these data reference public digital cultural resources such as blog, microblog, forum, web news and data from public cultural websites.
- **Data analysis module**: including three basic models: user behavior model, user relationship model in social network, and region-based hot resources model.
- **Data storage module**: storing different kinds of data, including basic data from data collection module, and application data from data analysis module.
- **Precision recommendation module**: helping for selecting the resources from the real-time data that the target user may be interested in for recommendation. For the requirements of complexity of public digital cultural resources, we design different recommendation engines showed in the next section.
- **Evaluation module**: evaluating the recommendation model with respect to rationality, effectiveness accuracy, novelty, and personalization according to the feedbacks of recommend results in order to optimize the recommendation algorithms.

The most important module of all of the six modules is precision recommendation module. **Error! Reference source not found.** shows the architecture of the module. The main workflow of precision recommendation module is described as follows:

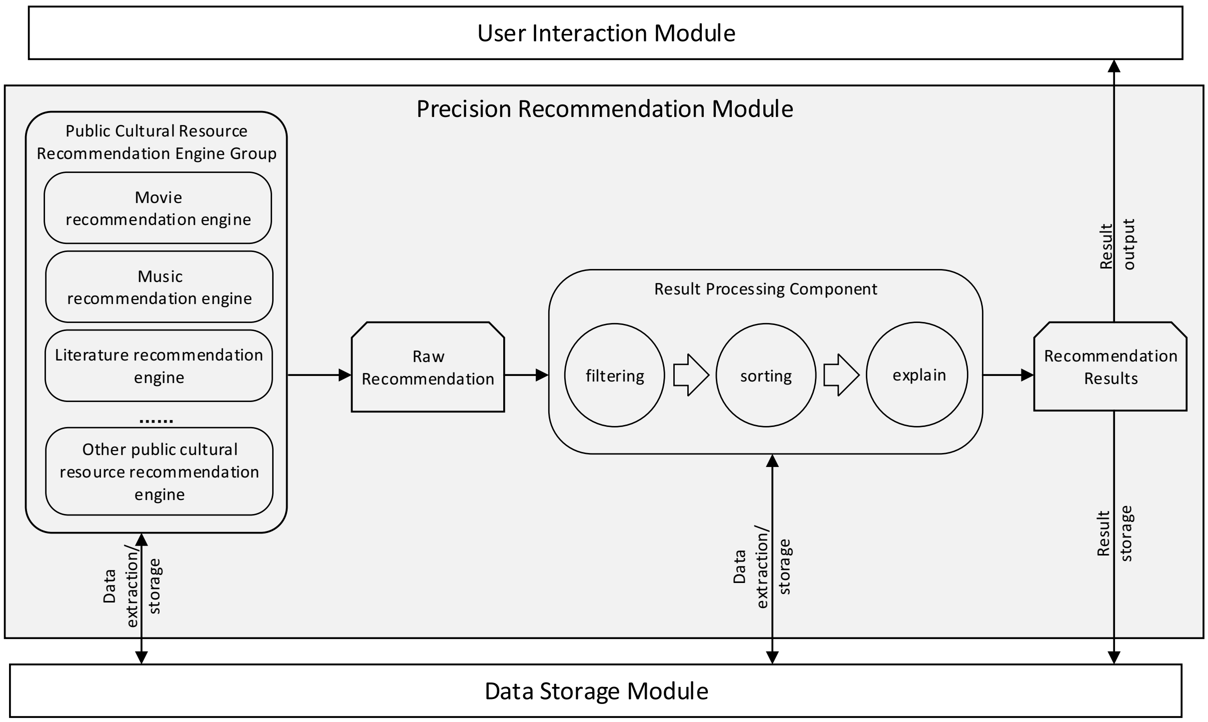

Figure 2.Architecture of Recommendation module

Step1: Select an appropriate recommendation engine and generate recommendation candidate set according to the type of cultural resources;

Step2: Filter the following resources from the candidate set

a) The resources that the user has visited ;

b) The resources that have been recommended to the user for many times, but have never been visited;

c) The resources that have very low overall scores;

Step3: Sort the candidate set in order to refine the recommendation results. This step mainly deals with the candidate set by considering the factors like novelty, surprise, personalization, and time factor.

Step4: Explain the final results why we recommend the resources to the user rather than others. This step is essential because it needs to present the trustworthy of the recommendations.

Step5: Sent the results and the explanation of the recommendation to the user interaction module. So far, the recommendation has completed.

IV. RECOMMENDATION METHOD ANALYSIS

A. Method introduction

In this paper, we present a social graph-based collaborative filtering recommendation method. The basic idea of this method is described as follows:

Step1: build the social graph of users.

Step2: find community by suitable community discovery methods.

Step3: search the close friends of the target user.

Step4: combine the results of step 3 with the collaborative filtering recommendation algorithms by computing the similarity of close friends and by creating recommendation results based on the close friends. **Error! Reference source not found.** shows the processing in detail.

Social network analysis plays a crucial role in this method. In this paper, we take a popular Chinese social media “Sina weibo [15]” as an example to present the community discovery [16] with the following steps.

Step1: build a social graph from users’ social network. There are two kinds of relationships between users: bilateral attention and unilateral attention. In order to reduce the complexity and have more tightness of social connections, we only establish bilateral relationship graph. Nodes represent different users, edges represent the relationships between users, and the weight of edge represents the user's intimacy.

Step2: calculate degree of nodes and select top-N nodes as the core nodes in order to find communities.

Step3: aggregate the rest nodes. We will determine the communities the nodes belong to in two steps. First, compare the distances between one of the rest nodes and the core nodes. While they have the same distance, the node belongs to the community who has the biggest intimacy. The step is iteratively carried out until all the rest nodes are assigned to a community.

Step4: Merge community. Communities created in the last step. The principle is to merge the communities that maximally share the nodes. If the numbers of shared nodes are equal, we will merge the communities that have the maximal intimacy.

B. Data sparse

The method solves the data sparse problem by establishing the user's social circle with community discovery and mostly reducing the dimensions of user-resources score matrix.

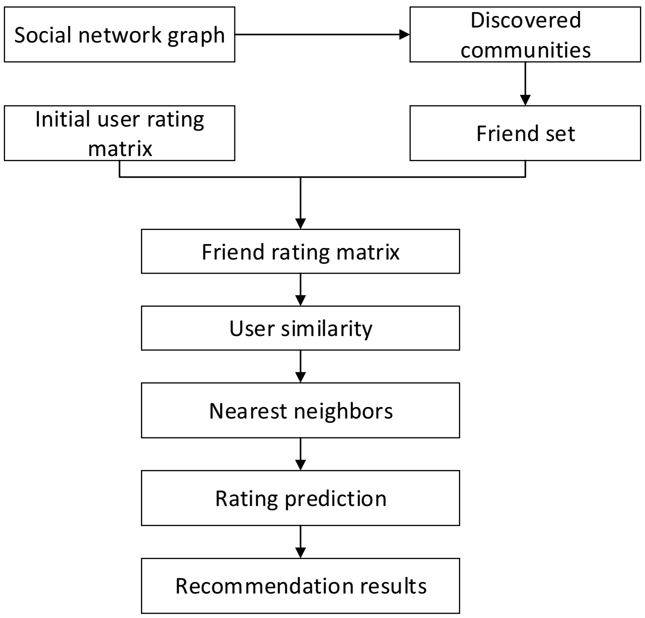

Figure 3.Process of recommendation method

C. Cold-start

1) user cold-start

A new user doesn't have any ratings so that it is impossible to compute the similarity. To solve this problem, we apply users' intimate degree instead of the similarity.

2) item cold-start

For the new cultural resources, on account of the absence of rating, we consider the value that combines the average value of the highest frequency rating of target user's rating history with the highest frequency rating of the user's close friends as the final rating.

V. CONCLUSION AND FUTURE WORK

In this paper, we propose a kind of precision recommendation method in order to recommend the resources that users are interested in from the large-scale public digital cultural resources. This method integrates the characteristics of social network into recommendation methods so that we can solve the data sparse problem and the cold-start problem of collaborative filtering recommendation algorithms. In addition, we design a precision recommendation framework for public digital cultural resources.

Our work introduced in this paper is an early-stage study of the precision recommendation of public digital cultural resources. In the future, we are going to perform experiments for evaluation of our method. Afterwards, we will expand this method from movie to other public digital cultural resources. We also plan to analyze the Recommendation method for the public digital cultural resources without ratings. Furthermore, we will optimize the calculation method of users' intimacy.

ACKNOWLEDGMENT

This research is supported by: (1) Research of Big Data Collection and Analysis for Public Culture Services (Project of National System Design of Public Cultural Service System 2015/2016) (2) the Support Program of the National '12th Five-Year-Plan of China' under Grant No. 2015BAK25B03 and 2015BAK25B04; (3) National Cultural Resource Sharing Project (key project of '13th Five-Year Plan of China'). At the same time, thanks for the help and encouragement of our colleagues and students in our lab.

REFERENCES

[1] http://www.ndcnc.gov.cn/. accessed 01-Dec-2015

[2] Dietmar Jannach, Markus Zanker, Alexander Felfernig, Gerhard Friedrich. An Introduction:Recommender Systems. Cambridge University Press.2010

[3] J. Ben Schafer, Joseph Konstan, and John Riedl. 1999. Recommender systems in e-commerce. In Proceedings of the 1st ACM conference on Electronic commerce (EC '99). ACM, New York, NY, USA, 158-166

[4] Greg Linden, Brent Smith, and Jeremy York. 2003. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. IEEE Internet Computing 7, 1 (January 2003), 76-80

[5] Carlos A. Gomez-Uribe and Neil Hunt. 2015. The Netflix Recommender System: Algorithms, Business Value, and Innovation. ACM Trans. Manage. Inf. Syst. 6, 4, Article 13 (December 2015), 19 pages

[6] Paul B. Kantor, Francesco Ricci, Lior Rokach, Bracha Shapira. Recommender Systems Handbook. Springer. 2010

[7] Breese,J.S., Heckerman,D, and Kadie,C, Empirical Analysis of Predictive Algorithms for Collaborative Filtering,Proceedings of the 14th Conference Uncertainty inArtificial Intelligence,PP.43-52,1998

[8] Polat H, Du W, SVD-based collaborative filtering with privacy Proceedings of Annual ACM Symposium on Applied Computing conferences(SAC'05),Santa Fe,2005, 791-795

[9] Agarwal D, Chen. C, Regression-based latent factor models, Proceedings of 15thACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD),PariS,2009:19–28.

[10] Browning,J and Miller D.J.,A Maximum EntropyApproach for Collaborative Filtering, Proceedings of the 2001 IEEE Signal Processing Society Workshop,PP.3−12,2001

[11] Byeong Man Kim,Qing Li,Jong-Wan Kim,and Jinsoo Kiln, A new collaborative recommender system addressing three problems, Proceedings of PRICAI 2004,P.95-504,2004

[12] Sarwar,B.M.,Karypis,G,Konstan,J.A.,and Riedl,J., Application of Dimensionality Reduction in Recommender Systems-A Case Study,ACM WebKDD 2000 WebMining for E-Commerce Workshop,2000

[13] Bell,R.,Koren,Y.,Volinsky,C., Modeling relationships at multiple scales to improve accuracy of large recommender systems.In:KDD'07:Proc.of the 13th ACM SIGKDD Int.cof

[14] Zhou K, Yang S H, Zhao H. Functional matrix factorizations for cold—start recommendat ion.Proceedings of the 34th internat ional ACM SIGIR conference on research and development in Information Retrieval(SIGIR), Beijing,2011:315—324

[15] Sina weibo. http://weibo.com/ ;accessed 01-Dec-2015

[16] Girvan M and Newman M E J. Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America, 2002, 99(12):782.