2017 IEEE Third International Conference on Multimedia Big Data

Integrated Recommendation for Public Cultural Service

Yi Yang[1], Guigang Zhang[1], Jian Wang[1*], Shufeng Ye[1]
[1] Institute of Automation
Chinese Academy of Sciences
Beijing, China
{yangyi, guigang.zhang, jian.wang, yeshufeng2015}@ia.ac.cn
*Corresponding author: Jian Wang

Jianqiao Hu[1,2]
[2] Department of Automation
Beijing Institute of Technology
Beijing, China
songshansls@sina.com

***Abstract*—Recent years recommender systems have been used in the public cultural services to provide personalized services. The single recommendation methods only cover the limited areas of interest of users. To address this issue, this paper proposes an integrated recommender system for cross-domain heterogeneous recommendations of the public cultural resources. The proposed system works on the recommendations for resource, knowledge, semantic tags, and user interest. Using the proposed system, users can explore interesting information in different point of views. Additionally, the proposed system visualizes the recommendation results in suitable ways.**

Keywords- big data, recommender system, cross-domain, hybrid, recommendation, heterogeneous, tagging system, knowledge graph.

I. INTRODUCTION

Recently the Chinese online public cultural service has grown rapidly. To effectively process and analyze the big data generated from the service, we proposed the public cultural big data analysis platform in previous work [1]. Because users desire to learn rich knowledge from the cultural resources, we discover the hidden knowledge in public cultural big data. We proposed the public cultural knowledge graph platform [2] to manage and analyze the extracted knowledge. We built a recommender system [3] that provided collaborative filtering (CF). However, the system has open issues e.g., the cold-start problem, the sparsity problem, and the diversity problem.

In this paper, we propose an integrated recommender system that extends the above system [3] with the benefit of the public cultural knowledge graph platform early developed. The proposed system provides cross-domain heterogeneous hybrid recommendations in order to improve the CF-based methods and provide knowledge recommendation. The system can visually explores the semantic of resources and user interests.

The rest of the paper is organized as follows. Section 2 introduces the related work. Section 3 presents the integrated public cultural recommender system. Section 4 introduces the architecture of the system. Section 5 shows an example. Finally, conclusion and future work are given in Section 6.

II. RELATED WORK

Information technologies have been applied in many public cultural services. The project European Digital Library (EDL) [1] integrated and shared the data of public cultural institutions of European countries. The project OCLC[2] provided services for libraries to manage and share catalog data. Zhang et al. [1] built a big data platform for managing and sharing the Chinese public cultural resources. Knowledge can be extracted from big data. Google proposed Knowledge Graph [4] to extend the traditional search engine for knowledge search. Dong et al. [5] proposed Knowledge Vault for the web scale knowledge base by integrating web extractions based on the Google Knowledge Graph. Pujara et al. [6] proposed approaches for knowledge graph identification by combining the ontological information and the distributional information. Yang et al. [7] built a public cultural knowledge graph platform that can extract and explore cultural knowledge by analyzing a knowledge graph.

Recommender system [8] is used for personalized services in many fields, e.g., electronic commerce [9] [10], and TV industry [11]. The widely used recommendation method is collaborative filtering [12] that finds potential interesting items by analyzing the co-occurrence of user behaviors. There are methods based on the CF, e.g., SVD-based CF [13], Linear Regression-based CF [14], and Markov Decision Processes-based CF [15]. Zhai et al. [2] proposed a recommender system with an applied adaptive CF for public cultural resources. Gao et al. [16] proposed a cross-domain recommendation for Cyber-Physical Systems. The article [17] presented a cross-domain citation recommendation for the research publications related to a specified patent. Kumar et al. [18] showed a latent sematic-based method for cross-domain recommendations using topic modeling of vocabulary clusters of distinct domains.

III. INTEGRATED RECOMMENDER SYSTEM

In this paper, we present an integrated recommender system for public cultural resources (see Figure 1). Our system is an important part of the Public Cultural Big Data Analysis Platform [1] so that it needs cooperate with components of the platform including the public cultural resource database, the user profile database, and the public cultural knowledge graph platform. The public cultural resource database manages the public cultural resources and their metadata. The user profile database manages user (behavior) data. The public cultural knowledge graph platform builds a knowledge graph and the related mining methods. The platform collects data from Chinese Wikis, Chinese culture websites, and public cultural

Sponsors: The Support Program of the National '12th Five-Year-Plan of China' under Grant No. 2015BAK25B04.

[1]EDL. http://www.edlproject.eu/

[2]OCLC. https://www.oclc.org/

institutions for building and updating a knowledge graph. The platform provides services for knowledge graph analytics.

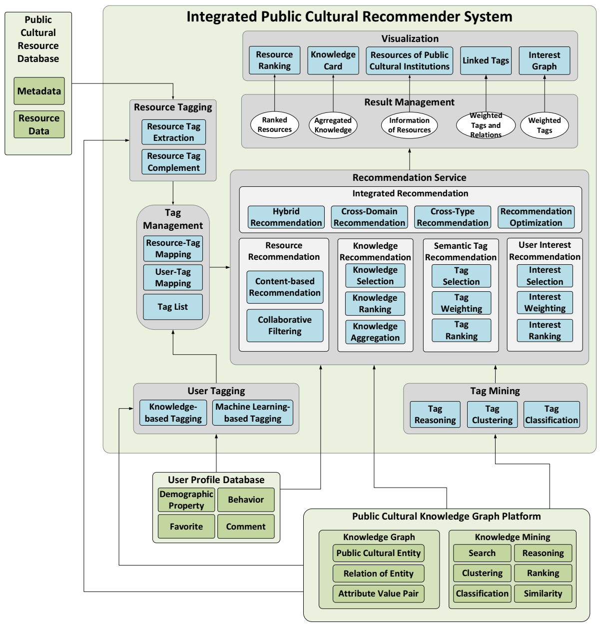

Figure 1 Workflow of the Integrated Recommender System

A. *Resource Tagging*

Our system obtains metadata of the resources in the public cultural resource database. The system extracts tags using NLP and create new tags by the public cultural knowledge graph. The resource tags are managed as resource-tag mappings.

B. *User Tagging*

Our system extracts tags from the user profile database for modeling user interest using machine learning-based methods and knowledge-based methods. The machine learning methods focus on statistics of the tags. The system maps the tags into entities of the knowledge graph for identifying new tags. The attributes of the entity and the linked entities may be new tags. The results of the user tagging are user-tag mappings.

C. *Tag Management*

The integrated recommender system manages the resource-tag mappings and the user-tag mappings in the form of key-value pair. The system summarizes the all the tags in a tag list.

D. *Recommendation Service*

The integrated recommender system provides cross-domain heterogeneous recommendation services.

1) Resource Recommendation: This service provides two types of recommendations: content-based recommendation and CF. The content-based methods match user interest tags and item tags. The results are presented as the resource list.

2) Knowledge Recommendation: Users desire to learn the related knowledge of the resource when viewing the cultural resources. The proposed system gives two types of knowledge with the help of the cultural knowledge graph: descriptive knowledge and collection of public cultural institutions. The descriptive knowledge shows attributes and textual description of the resource. The collection of public cultural institutions presents the related articles of the institutions. Our system also considers the location of users, e.g., recommend the articles of the local museums. The system can search the semantic tags of the given resource over the knowledge graph.

3) Semantic Tag Recommendation: Semantic of resources are described by tags. This service helps recommend resources from the point of view of semantic of the content. For the target tags, the service recommends the semantically related tags that present the points users may like. Tag Mining Unit provides basic analysis for the service. A resource tag may come from multiple metadata with different importance: $S_{keyword}>S_{name}>S_{description}$. For a tag t, the weight is defined as

$$W_t=S_{keyword}\cdot E_{keyword}(t)+S_{name}\cdot E_{name}(t)+S_{description}\cdot F_{description}(t)$$

where $S_{keyword}$ presents the impact factor of the keyword data, $E_{keyword}(t)$ decides the existence of the tag t in the keyword data: 0 means t does not exist in the keyword, 1 is the opposite; S_{name} means the impact factor of the resource name, $Ename(t)$ decides the existence of t in the resource name; $S_{description}$ means the impact factor of the descriptive content, and $F_{description}(t)$ is a semantic analysis function that analyzes and normalizes the importance of t to the description; $S_{keyword}+S_{name}+S_{description}=1$.

4) User Interest Recommendation: The service uses tags to present user interests because the tags mean the semantic of the interests. The weight of a tag t is defined as

$$W_t=S_{demo}\cdot E_{demo}(t)+S_{behavior}\cdot B_{behavior}(t)+S_{favorite}\cdot F_{favorite}(t)+S_{comment}\cdot C_{comment}(t)$$

where S_{demo} is the impact factor of the demographic property, $Edemo(t)$ decides the existence t in the demographic property; $S_{behavior}$ is the impact factor of the user behavior data, $B_{behavior}(t)$ is the behavior weighting function considering multiple factors, $S_{favorite}$ represents impact factor of the user favorite, $F_{favorite}(t)$ is the occurrence frequency of t in the tags of the favorite resources; $S_{comment}$ means the impact factor of the comment data, $C_{comment}(t)$ is a semantic analysis function that analyzes and normalizes the importance of t according to the comments; $S_{demo}+S_{behavior}+S_{favorite}+S_{comment}=1$. The service ranks and selects the tags according to the weights.

5) Integrated Recommendation: The service provides the cross-domain recommendation integrating recommendation methods that respectively work well in single domains.

E. *Tag Mining*

Tag Mining provides basic analysis methods for tags. The methods include rule-based methods, clustering methods, and classification methods. The rule-based methods depend on the semantic relations between tags that can be analyzed by the cultural knowledge graph. The clustering and classification methods are based on machine learning and deep learning.

F. *Visualization*

Our system visualizes the results. Resources are listed based on the rank. Knowledge is organized as the knowledge card. The linked tags are visualized as a node-link graph and the user interests are presented as a tag cloud.

IV. ARCHITECTURE

The architecture of the integrated recommendation system mainly consists of three layers: framework layer, analysis layer, and visualization layer (see Figure 2). In technical, we apply Apache Hadoop ecosystem for building our system.

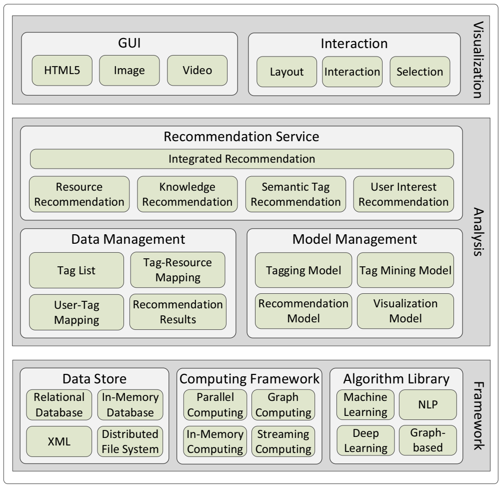

Figure 2 Architecture of Integrated Recommender System

A. Framework Layer

The framework layer provides underlying services of the system for data store, computing, and algorithms.

1) Data Store: The tags are stored in relational database cluster, i.e., MySQL Cluster in the system. The system applies Hadoop HDFS as the distributed file system to support the batch processing of massive data. The recommendation results are stored in the form of XML. The in-memory database Redis is used for accelerating data access. Other data stores are provided by the public cultural big data platform and the public cultural knowledge graph platform.

2) Computing Framework: We deploy multiple computing frameworks in the system for different demands. The Hadoop MapReduce are used as the parallel computing framework for batch processing of massive data. We use Spark GraphX as the graph computing engine for graph-based algorithms. The system applies Apache Spark as the in-memory computing framework for improving knowledge query. Spark Streaming is used for the streaming computing for real-time computing.

3) Algorithm Library: We deploy algorithm libraries for data processing and analysis. The libraries involve machine learning, deep learning, Graph processing, and NLP.

B. Analysis Layer

This layer provides modules for recommendation methods and the associated data and model management.

1) Data Management: Management of the tag list, tag-resource mapping, user-tag mapping, and recommendation results. The system provides interfaces for data access and data configuration.

2) Model Management: The tagging models assign tags to resources and users. The tag mining models identify tags and their relations. The recommendation models work on the recommendation services. The visualization models visually present the recommendation results.

3) Recommendation Service: The module provides recommendation methods for the resources, the knowledge recommendation, the semantic tag recommendation, the user interest recommendation, and the integrated recommendation.

C. Visualization Layer

This layer focuses on presenting the resulting information of the recommendations. We develop the browser-based GUI using HTML5. Meanwhile, the GUI also presents images and videos. The recommendation services proposed in the paper have individual areas for visualizing the results.

V. APPLICATION

In this section we present an example scenario to depict the integrated public cultural recommender system. When a user views a video clip of the famous Chinese history novel "Romance of the Three Kingdoms" on a cultural application, our system provides cross-domain recommendations with six main views on the screen (see Figure 3).

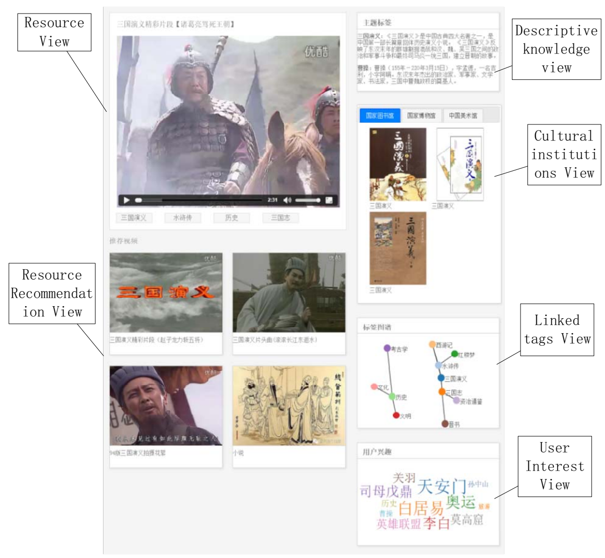

Figure 3 Example Scenario

The resource view demonstrates the video being viewed. The tags of the video clip include the content description of the video clip, the important roles of the clip. According to the tags, the integrated recommender provides five recommendations. The resource recommendation view shows the thumbnails of the resources the user may like. This information is obtained by the resource recommendation service of our system.

The descriptive knowledge view shows description and knowledge card for the relevant knowledge for depicting the background and the important roles of the clip. The cultural institution view shows the collection of three institutions related to the video clip. In this view, each tab presents an institution. The first tab part shows resource links of the National Library of China for the related books. If the user has an account of the library, she can directly read the book online. The second tab shows the antiques related to the story of the video clip. The antiques are collected in the National Museum of China. The third part of the view shows the famous paintings of the important roles in the video clip. The paintings are collected in the National Art Museum of China. When the user is interested in the collection of the museums, she can go to the institutions to view the real collections. In this way, our system additionally recommends the public cultural institutions. It implements a cross-domain recommendation.

The linked tags view shows the tags related to the video clip and the semantic relations of the tags. We visualize the tags using node-link graph. Nodes represent the tags and links represent the relations of the tags. The more relevant the tags are, the shorter the link between the tags is.

The user interest view shows the user's interests by using tag cloud layout. This layout presents interests as tags. The size of the tag depends on the significance of the interest to the user. The important interest is represented with a large tag. In this way, the interests are explicitly demonstrated, so that the user may explore more interesting resources.

The integrated recommendation shows different types of resulting data. Starting with a resource, the system suggests the videos, images, and texts, i.e., heterogeneous recommendation. From another perspective, starting with a TV show clip, the system recommends books, background knowledge, work of art, and antiques, i.e., cross-domain recommendation.

VI. Conclusion

In this paper, we introduced an integrated recommender system for the cross-domain heterogeneous recommendation of public cultural resources. The system can partially enhance the diversity of recommendation. The proposed system improves the CF-based methods by tagging methods. The proposed system provides the knowledge recommendation for depicting the content of the resource and showing the collection provided by the public cultural institutions for recommending the public cultural institutions. The proposed system also applies tags for exploring meaningful resources in the point of view of semantic and analyzing user interests. The proposed system visualizes the recommendation results for easy understanding. Using the proposed integrated recommender system, users can obtain the cross-domain heterogeneous recommendations. In the future, we will perform experiments to evaluate our system. We will analyze the recommendation patterns of our methods and apply the identified patterns to optimize the embedded recommendation methods.

Acknowledgment

We thank the National Key Technology Support Program of the National '12th Five-Year-Plan of China' under Grant No. 2015BAK25B04 for partially supporting our work.

References

[1] G. Zhang, Y. Yang, X. Zhai, W. Huang, and J. Wang, "Public Cultural Big Data Analysis Platform," 2016 IEEE Second International Conference on Multimedia Big Data (BigMM), Taipei, 2016, pp. 398-403.

[2] Yi Yang, Guigang Zhang, Jian Wang, Shufeng Ye, Jianqiao Hu, "Public Cultural Knowledge Graph Platform," IEEE International Conference on Semantic Computing (ICSC 2017) and The 1st International Workshop on Big Data for Intelligent Services and Applications(BDISA), San Diego, California, USA, 2017.

[3] X. Zhai, F. Jin, J. Wang, Y. Yang, Q. Yao, Q. Qiu and G. Zhang, "A Kind of Precision Recommendation Method for Massive Public Digital Cultural Resources: A Preliminary Report," 2016 IEEE Second International Conference on Multimedia Big Data (BigMM), Taipei, 2016, pp. 56-59.

[4] A. Singhal, "Introducing the knowledge graph: Things, not strings," The official Google blog, 2012: http://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html accessed on 01-Nov-2016.

[5] X. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmann, S. Sun, and Wei Zhang, "Knowledge vault: a web-scale approach to probabilistic knowledge fusion," In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '14). ACM, New York, NY, USA, pp. 601-610.

[6] J Pujara, H. Miao, L. Getoor and W. W. Cohen, "Ontology-aware partitioning for knowledge graph identification," In Proceedings of the 2013 workshop on Automated knowledge base construction (AKBC '13). ACM, New York, NY, USA, pp. 19-24.

[7] Y. Yang, G. Zhang, J. Wang, S. Ye, and J. Hu, "Public Cultural Knowledge Graph Platform," The International Workshop on Big Data for Intelligent Services and Applications (BDISA 2017) in conjunction with 2017 IEEE International Conference on Semantic Computing (ICSC 2017), in press.

[8] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich. "An Introduction:Recommender Systems," Cambridge University Press.2010.

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

[10] G. Linden, B. Smith, and J. York. "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," IEEE Internet Computing 7, 1 (January 2003), 76-80

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

[12] P. B. Kantor, F. Ricci, L. Rokach, and B. Shapira, "Recommender Systems Handbook," Springer. 2010

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

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

[15] J. Browning and D. J. Miller, "A maximum entropy approach for collaborative filtering," Neural Networks for Signal Processing XI: Proceedings of the 2001 IEEE Signal Processing Society Workshop (IEEE Cat. No.01TH8584), North Falmouth, MA, 2001, pp. 3-12.

[16] S. Gao, H. Luo, D. Chen, S. Li, P. Gallinar, Z. Ma, and J. Guo, "A Cross-Domain Recommendation Model for Cyber-Physical Systems," in IEEE Transactions on Emerging Topics in Computing, vol. 1, no. 2, pp. 384-393, Dec. 2013.

[17] S. Tantanasiriwong and C. Haruechaiyasak, "Cross-domain citation recommendation based on Co-Citation Selection," 2014 International Conference on Electrical Engineering / Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Nakhon Ratchasima, 2014, pp. 1-4.

[18] A. Kumar, N. Kumar, M. Hussain, S. Chaudhury, and S. Agarwal, "Semantic clustering-based cross-domain recommendation," 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), Orlando, FL, 2014, pp. 137-141.