Analysis of the knowledge map of user portraits abroad

Xiaoyu Fan

School of Economics and Management, Xidian University, Shaanxi Xi'an 710071, China

15229221817@163.com

Abstract

User portrait, CiteSpace, Visualization.

Keywords

This paper makes a visual analysis of the literatures on user portrait in foreign countries and analyzes the research status and future research frontiers of the foreign users' portraits. This research provides references for the research in this field. Based on the literature on user portrait in Web of Science database, this paper analyzes the research progress of user portrait from the perspective of time, country, author, citation and research hotspot, using the CiteSpace tool visualization analysis. The study found that the number of articles about the user's portrait showed a general trend of growth, indicating that the field was increasingly concerned by researchers. From the perspective of author cooperation, there are more cooperation between authors, but there is less cooperation among groups. Countries such as the United States, France and the United Kingdom are countries with high research influence in this field. China is a country with a large contribution to this field. From the perspective of research hotspot, the retrieval and recommendation of user portrait and information are closely related, and the construction of self-adaptive user portrait has gradually become a hotspot, and ontology, social media and recommendation system will be the direction of future research in this field.

1. Introduction

After the Internet was entering the age of big data, the number of data on the network grew, and it was hard to find the information to satisfy your needs in a huge amount of data, and to be able to capture the user's needs in a precise way, by collecting information about the user, data analysis and mining of the user's information to discover the preferences of the users, and to provide a personalized service. Thus, the concept of "User portrait" was born. The concept of User portrait was first introduced in 1997 by a. i. Kokkinaki of the university of Cyprus. He believed that the User portrait was based on various attributes of users to describe the User's methods and to construct the picture of credit card users by credit card number, transaction date, business type, location, expense amount, credit line and expiration date to detect fraudulent behaviors that could occur during the transaction[1].In addition, there are other scholars' definition of userportrait: Tang Jie of Tsinghua university believes that the construction of user portraits is the process of obtaining the value of various attributes required by the user model [2].Bhavani Raskutti believes that the user profilr contains a summary of the attributes affecting the user's preferences [3].Miha Grčar think user portrait is to provide the user preferences based on your browsing history.Giuseppe Amatobelieves that the user portrait is a structured way to represent the user's needs[4],The author thinks that the user portrait is a comprehensive and detailed extraction of the user's information from the mass of information, and the user's information label is extracted from the aspects of demographic attributes, interest characteristics and behavioral attributes, so as to build a user model for a real user. This paper takes the Web of Science database as the data source and takes CiteSpace as the tool to make a visual analysis of the author, country, cited literature, cited author, research hotspot and other aspects in this field to understand the research progress in this field, so as to provide references for subsequent studies.

2. Data preparation and method introduction.

2.1 Data preparation

Based on the web of science database as the data source, the subject word = "user portrait", the document type is Article, the time span is all the years, the retrieval date is April 25, 2017, the language is English, and 552 records are retrieved. The field includes the author (AU), the title of the literature (TI), the publication name (SO), the author's key words (DE), the abstract (AB), the fund (FU), the reference (CR), the citation frequency (TC), etc.

2.2 Method introduction

The map of scientific knowledge is the measurement unit of scientific knowledge, which shows the development process and structure relation of scientific knowledge in a visual way, reveals the dynamic development law of the field, and provides reference for the subject research. [5].The tools currently used to map knowledge are CiteSpace, Ucinet, HistCite, etc. This article is based on the CiteSpace invented by Dr. Chen chao - mei..The unique innovation of CiteSpaceII is that the scientific knowledge graph drawn by CiteSpaceII can show the trend and trend of a certain discipline or knowledge domain in a certain period of time, and form a number of research frontier fields. [6]

3. Time and space distribution
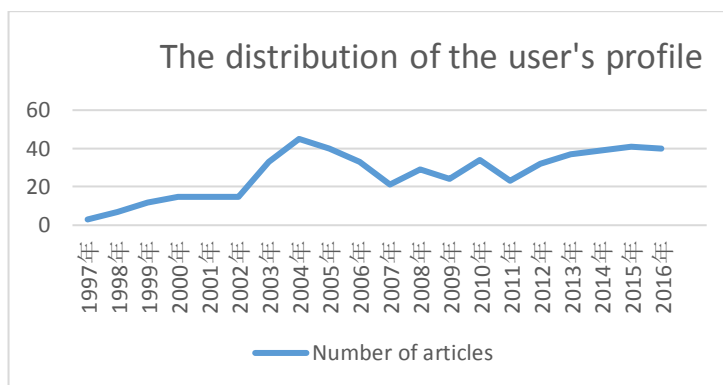
3.1 Time distribution

Figure 1 The number of articles on user portraits

From 1997 to 2016, the number of research papers on user portrait has changed, as shown in figure 1. The term "User Portrait" first appeared in 1997. According to the trend of the number of articles, the study of User portraits can be divided into three stages.

The first stage was from 1997 to 2004, when the number of articles increased, especially between 2002 and 2004, with a large increase, which peaked in 2004. The user portrait is an important method to provide personalized service, and the development of personalized service has a great influence on the research of user portrait. In the middle of 2005, the research of personalized service was gradually transferred from the concept research to the realization of technology, which promoted the development of user portrait. Therefore, the number of studies on user portraits increased greatly from 2002 to 2004, and reached its peak in 2004.

In the second stage, between 2004 and 2011, the number of articles declined and there was a fluctuation. The study of user portraits was bottleneck, and the researchers applied the user portrait to information filtering, retrieval and recommendation process on the basis of in-depth understanding of the concept of user portrait. However, with the change of users' personalized requirements, traditional methods found it difficult to find the law of user interest change, and to build a dynamic self-adaptive user portrait was the key and difficult point in this field. The third stage is that after 2011, the number of articles has gradually increased steadily, indicating that the user portrait has become a research hotspot and favored by more scholars. It's a development that's consistent with professor Rogers of the university of Mexico, and the development of new things, and the development of new things is usually the beginning of the development of the "S" or the development

of new things, and then it progresses slowly, and then it's going to accelerate, and then it's going to go into a stable development phase, and then it's going to go back to the stable development phase, but it doesn't mean that it's not getting any more attention, but it's the opposite, and it's actually the new growth point of the paper, and once it breaks through the field and the literature is going to go into a new kind of development -- a period of rapid growth[7].As you can see, the user's portrait is getting noticed by researchers.
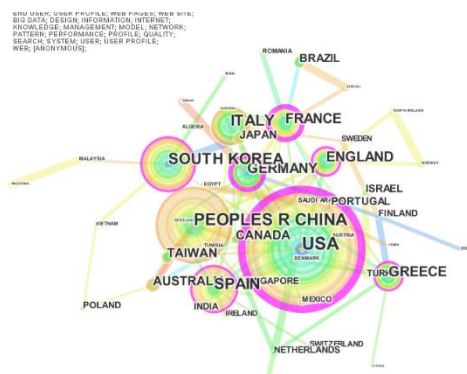
3.2 Regional distribution

Figure 2 National cooperation on user portraits

Table 1 In the top ten countries of centrality

Serial number	Countries	Centricity	Number of articles
1	USA	0.51	89
2	France	0.27	27
3	South Korea	0.20	41
4	Germany	0.20	26
5	England	0.16	27
6	Greece	0.14	28
7	Spain	0.11	36
8	Italy	0.08	39
9	Canada	0.08	17
10	Peoples R China	0.07	63

CiteSpace was used to analyze the countries that studied user portraits, as shown in figure 2, and listed the top ten countries in the ranking list, as shown in table 1. The central centricity of the nodes in the network is a measure of the importance of nodes, and they use a purple circle in the visual graph. The nodes of the United States, France, South Korea, Germany, Britain, Greece and Spain are all marked with purple circles, and their centrality is greater than 0.1, indicating that these nodes are the most important nodes in the whole network, and have a strong influence on the network structure. In this field, the intermediary role is relatively obvious, and from the cited literature, eight of the top ten cited articles are from the United States, so the U.S. is in a dominant position in this field. From the size of each country node in figure 2, the US has the largest node radius, indicating that it has the largest number of posts, second only to China. Analysis of the author in the field of cooperation, in the number of the top ten authors, three from the United States, four are from China, is the United States and China's great contribution in this field.

3.3 Analysis of authors cooperation

It use CiteSpace to conduct collaborative analysis. Set the node type as Author, and get a network diagram of the Author with 306 nodes and 172 connections, as shown in figure 3. From figure 3, it is found that the network density of the whole author is relatively low. Although there are many

cooperation between the author and the author, the cooperation between them has not reached the long-term cooperation. 90% of the literature published by the sub-network is 1 to 2, and the research power is relatively dispersed.
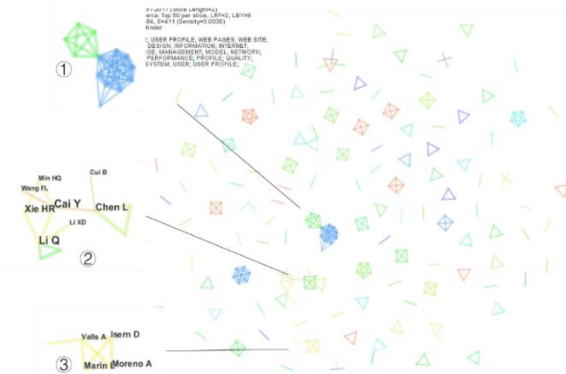

Figure 3 Graph of authors cooperation

According to the results of the analysis of the annual article in section 2.1, the research trend of the user portrait is S shape. Before 2011, the researchers continuously explored and tested the user portraits. After 2011, the researchers gradually accepted and started the research on the user portrait, but the two most important factors that formed the research group were the stable research subjects and the academic leaders who were able to grasp the academic frontier, and the research on user portraits did not have the conditions to form a scientific research group. Therefore, there has not been a strong cohesive research group on user portraits.

3.4 Analysis of institutional

In terms of the number of published articles, there are 398 units in the study of user images in foreign countries. In the top three, Ben Gurion Univ Negev 10, City Univ Hong Kong 8, and Natl tech Univ Athens 7, indicating that these three institutions have strong scientific research capabilities. In terms of cooperation between institutions, the research groups in this field are relatively distributed, and the cooperation among different institutions is less, and the only cooperation is obviously regional. From different time periods, Fdn&Res Technol Hellas related results, published as early as 1997, but its influence is not big, after 2000, the number of results it without sustained growth, and the city university of Hong Kong since 2005, the number continues to rise; The Chinese academy of sciences only published relevant literature in 2014, but as of 2016, the number of posts in the Chinese academy of sciences has ranked sixth, and the results of scientific research are quite remarkable, which belongs to the rising star..

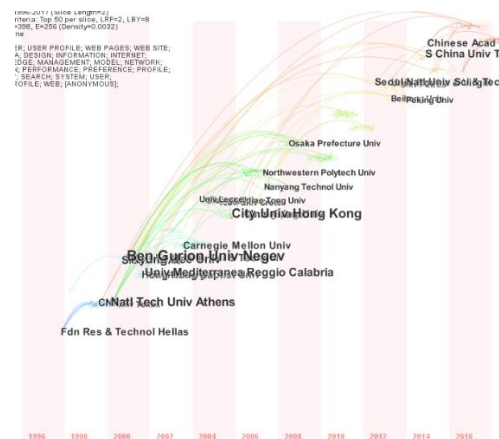Figure 4 An institutional collaboration map of user profiles.

Figure 5 The time sequence map of user portraits

3.5 Co-citation analysis
3.5.1 Co-citation analysis of authors

Analyses were Cited authors use CiteSpace, node type is set to Cited Author, get Cited Author collaboration network diagram, as shown in figure 6, lists the portrait about user study of highly Cited authors, as shown in table 2. The research areas of these authors include artificial intelligence, data mining, recommendation system, personalization, etc. At Cornell university, computer science and technology experts Gerald theoden and his team have developed out of the intelligent information retrieval system, famous Vector Space Model is put forward by scholars from all walks of life are widely used.. Thorsten Joachims is also a professor at Cornell university, in 2003, he put forward the model of support vector machine (SVM) based on sequence learning, vector space model and Ranking SVM model is widely studied and applied to various fields. Gedas Adomavicius, from the university of Minnesota, is dedicated to the study of personalized technology and recommendation systems. Professor Bamshad Mobasher, of the university of Chicago in Chicago, has been studying personalized studies since the 1990s. He has a comprehensive and unique perspective on the characteristics, history, current situation and future trends of individuation. His field of research involves artificial intelligence and data mining, which he is regarded as one of the leading scholars in Web mining, Web personalization and recommendation systems.
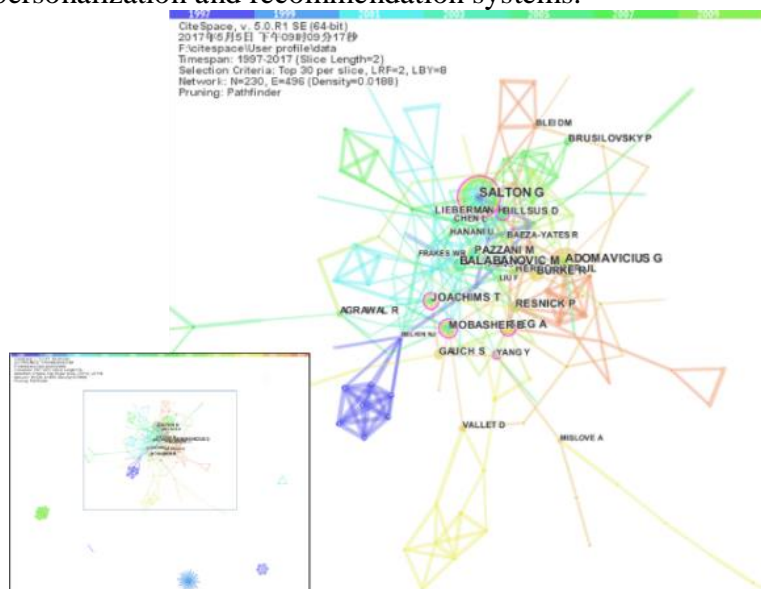

Figure 6 The co-occurrence graph of the authors

Table 2 The core author information list

Name of core author	Citation frequency	centricity	Name of core author	Citation frequency	centricity
Gerard Salton	52	0.04	Robin Burke	24	0.03
Gedas Adomavicius	43	0.15	Michael Pazzani	23	0.09
Marko Balabanovic	30	0.05	Ahu Sieg	21	0.04
Bamshad Mobasher	25	0.08	Paul Resnick	21	0.02
Thorsten Joachims	25	0.06	Daniel Billsus	20	0.01

In addition to the relatively large network of cooperation, we can find several small groups around it, which may be caused by a small number of authors' research involving other fields. More than 10 people like Duong T, Bailey BP and Arroyo E form A small network, and their research involves wireless network system [8],In addition to the relatively large network of cooperation, we can find several small groups around it, which may be caused by a small number of authors' research involving

other fields. More than 10 people like Duong T, Bailey BP and Arroyo E form A small network, and their research involves wireless network system [8]. In another small network, Bardossy A is A research scholar in geography. It can be seen that that us's portrait involves a wide range of field. From the density and color of the node distribution in the graph, the number of nodes was small and the network formed was small between 1997 and 1999. During the period from 2002 to 2011, more and more people began to study user portraits (the number of nodes increased and formed a huge and dense network). After 2012, the number of nodes was also larger, but the distribution was more dispersed, indicating that the user portrait had different branches

3.5.2 Co‑citation analysis of papers

In order to make the broader map, and the node type choice Cited reference and term, namely choosing citations and keywords at the same time, network, draw their co-occurrence graph nodes in number increase, contains more information. The threshold value is set to (2,2,20), (4,3,20) and (3,3,20), and a network map consisting of 520 nodes and 964 wires is obtained, as shown in figure 7. The citation of the thesis partly reflects the researchers' recognition of the paper's viewpoints, the higher the citation frequency, the greater the academic influence.

According to the key literatures in this field, the nodes with purple outer ring are the nodes with large middle centrality, and the key literature is the node on the network connection path, which is a transitional document, and it is the frontier of research in the period.[6].U.Cetintemel first studied the adaptive change in user portraits. In order to be able to accurately understand the user's interests, this paper proposes a supervision and user feedback based users interested in updating method, the user portrait expressed as multiple vector, the number and weight vector will change according to user's access behavior, the literature is static user portrait portraits to dynamic user key literature transition [9].
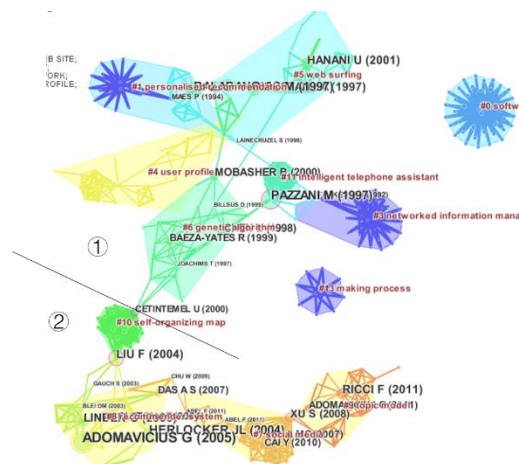

Figure 7 A map of Co-citation

Using this document as a cut-off point, the maximum network is divided into two parts. Comparing with network 1 and network 2, it is found that the network density in network 1 is lower than that of network no. 2, and the nodes are relatively scattered. Moreover, it is found that in network no. 1, clustering topics include multiple topics such as genetic algorithms, network information management, intelligent telephone assistant, etc., and the research content is not concentrated. In the second network, the nodes are closely distributed, which can be clearly found that the research topics are mainly divided into three: recommendation system, social media and thematic model, and according to the color distribution of the network, it has appeared in recent years. Gediminas Adomavicius summarizes the three recommended methods commonly used in recommendatory systems: content-based, collaborative, and hybrid recommendations, and analyzes the limitations of these three approaches.[10],As the basic literature of the field of recommendation, this article is cited as the highest citation frequency, and the author Gediminas Adomavicius is also a highly cited author in this field, which can be seen as an important reference value in the field of user portrait. S Xu and others put forward a model of personalized search, the user's interest and the theme of the web page

by social tagging, both associated with folksonomy, then match the retrieval results are obtained [11],The citation frequency of this document is only 4 times, but it plays an important role in the network, combining the theme model and social media, which is the connection point of clustering #7 and clustering #9. The join points of clustering #7 and clustering #8 combine the recommendation system with social media. Wei Chu proposes a machine learning method based on the characteristics, build a yahoo user's portrait, and on the basis of the characteristics of user's portrait, bilinear regression model is developed to predict, for an existing user and new user provides accurate personalized recommendation [12].In addition, the document of higher node centricity rate in the node and the article by Michael Pazzani in 1997 describes the use of a simple bayesian classifier for modifying the contents of a user's portrait, designing a Syskill & Webert system, which is a typical client-personalized service system, more and more cited [13].

There are also several smaller and darker networks around the network, indicating that the literature on user portraits will also involve other fields, and the color depth represents the further time of these documents. Through the analysis of the literature and information, it is found that the important sources of knowledge in the field of user portraits provide important reference and guiding significance for the study of user portraits.

3.6 Research hotspot analysis

The key word is extracted from the article and can summarize the words of the topic. High frequency keywords usually represent a research hotspot in a field. Use CiteSpace to draw keywords co-occurrence map. Node type selection for the Keyword, as shown in figure 8, 683 a network node in figure 8, article 1479 of attachment between nodes and node radius big point frequency is higher, to a certain extent, represents the research focus in the field, table 3 lists the portrait about user part of the high frequency words.

As can be seen from table 3, information retrieval and personalization have the highest frequency, which are often linked together, namely personalized information retrieval. It is a key step to provide personalized service for users to acquire personalized features, namely to build user portraits.In the image retrieval, we can build user portraits according to the information of the image color, so as to improve the efficiency of image retrieval [14],In video retrieval, we can recommend the video information that conforms to the user's interest by constructing the user's portrait[15].In the cross-language retrieval system, the construction of user portraits enables users to acquire relevant information of other language countries[16].
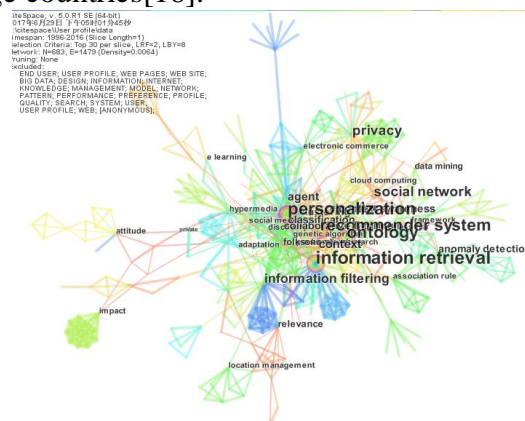

Figure 8 Keywords co-occurrence map

Figure 9 The detection graph of bursts words

Table 3 The list of high frequency words

High-frequency words	frequency	Rate of bursts	centricity
Information retrieval	48		0.17
Personalization	38		0.27
Ontology	35	5.01	0.10
Recommender system	34	6.15	0.14
Social network	16	4.23	0.06
privacy	15		0.04
Information filtering	13		0.09
context	11		0.04

Ontology, recommender system and social network are the emergent words in the field of user portrait, which represents the research frontier of this field to a certain extent..The ontology is a philosophical concept, and Alvarado Matl first combines the ontology with the user portrait. Through the position ontology, it uses the experience and knowledge of the members of the organization to construct its portrait. This method is used for the decision-making problems of position matching [17].Personalized recommendation is a hot research topic in recent years. As an important means of providing personalized service, user portrait plays an important role in the recommendation system [18][19].The emergence of social network has promoted the development of user portrait. The construction of user portrait not only starts with interest, but also takes into account the user's relationship. Recommendations from content recommendation also extend to people recommend [20].The user's interest is changing, and the static user portrait has been unable to meet the changing needs of users. Users' interests may be affected by factors such as the environment, so it is important to grasp the real needs of users at that time. Adaptive technology to automatically adjust the recommended as a result, building a dynamic user, to provide users with accurate content, improving accuracy recommended[21].

4. Summary

Based on the core database of Web of science, this paper makes a visual analysis of the research progress of user portraits with the help of CiteSpace visualization software. From the perspective of authors and institutional cooperation, there are many cooperation between authors, but there is less cooperation among groups, and there is no large-scale cooperation group. From the perspective of national distribution, the United States, France and South Korea have relatively high academic influence in the field. China has a low centrality but has the highest volume of publications, which is the country that contributes a lot to this field. From the perspective of co-citation analysis, the study of user portraits, in addition to user information retrieval and recommendation, also covers the field of communication and geography. In recent years, researchers have studied the construction and updating of user portraits. With the development of social network, the construction of user portrait is no longer only around users' interests, but also the social relations of users are gradually being considered. With the customization of users' needs, self-adaptive user portrait will be more satisfying to users' personalized needs, and ontology, social network and recommendation system are the direction of future user portrait research.

References

[1] Kokkinaki A.I.On Atypical Database Transactions: Identification of Probable Frauds using Machine Learning for User Profiling[C]// IEEE Knowledge and Data Engineering Exchange Workshop. IEEE Computer Society, 1997:107.
[2] Tang J, Yao L, Zhang D, et al. A Combination Approach to Web User Profiling[J]. Acm Transactions on Knowledge Discovery from Data, 2010, 5(1):2.

[3] Raskutti B, Beitz A, Ward B. A Feature-based Approach to Recommending Selections based on Past Preferences[J]. User Modeling and User-Adapted Interaction, 1997, 7(3):179-218.

[4] Amato G, Straccia U. User Portrait Modeling and Applications to Digital Libraries[C]// European Conference on Research and Advanced Technology for Digital Libraries. Springer-Verlag, 1999:184-197.

[5] Chen Yue, Liu Zeyuan, Chen Jin, et al. The development of scientific knowledge map [J]. Studies in Science of Science | Stud Sci Sci, 2008, 26(3):449-460.

[6] Chen Chaomei,Chen Yue,Hou Jianhua. et al．CiteSpaceⅡ：The recognition and visualization of new trends and new developments in the scientific literature［J］．Journal of The China Society for Scientific and Technical Information,2009（3）：401－421

[7] Shao Bo, Song Jiwei. Research and analysis of personalized service of e-commerce at home and abroad. [J]. Journal of Intelligence, 2008, 27(7):78-80.

[8] Marin L, Isern D, Moreno A. Dynamic adaptation of numerical attributes in a user portrait[J]. Applied Intelligence, 2013, 39(2):421-437

[9] Etintemel U C, Franklin M J, Giles C L. Self-Adaptive User Portraits for Large-Scale Data Delivery[C]// International Conference on Data Engineering, 2000. Proceedings. IEEE, 2000:622-633.

[10] Adomavicius G, Tuzhilin A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions[J]. IEEE Transactions on Knowledge & Data Engineering, 2005, 17(6):734-749

[11] Xu S, Bao S, Fei B, et al. Exploring folksonomy for personalized search[C]// International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2008, Singapore, July. DBLP, 2008:155-162.

[12] Chu W, Park S T. Personalized recommendation on dynamic content using predictive bilinear models[C]// International Conference on World Wide Web. ACM, 2009:691-700.

[13] Pazzani M, Billsus D. Learning and Revising User Portraits: The Identification of Interesting Web Sites[J]. Machine Learning, 1997, 27(3):313-331.

[14] Zhang H, Zhu X, Zhao J, et al. Online image information retrieval strategy based on the color preference of the user portrait[C]// IEEE International Conference on Ubi-Media Computing. IEEE Xplore, 2008:196-200.

[15] Shin S Y, Kang O H. A Remote Video Study Evaluation System Using a User Portrait[M]// Computational Science and Its Applications – ICCSA 2005. Springer Berlin Heidelberg, 2005:284-294.

[16] Safi H, Jaoua M, Belguith L H. PIRAT: A Personalized Information Retrieval System in Arabic Texts Based on a Hybrid Representation of a User Portrait[C]// International Conference on Applications of Natural Language to Information Systems. Springer, Cham, 2016:326-334.

[17] Alvarado M, Romero-Salcedo M, Sheremetov L. A corporative memory based on the user portraits[J]. Expert Systems with Applications, 2004, 26(1):87-94.

[18] Li L, Zheng L, Yang F, et al. Modeling and broadening temporal user interest in personalized news recommendation[J]. Expert Systems with Applications, 2014, 41(7):3168-3177

[19] Goh J Y, Taniar D. Mobile Data Mining by Location Dependencies[C]// Intelligent Data Engineering and Automated Learning - Ideal 2004, International Conference, Exeter, Uk, August 25-27, 2004, Proceedings. DBLP, 2004:225-231

[20] Raad E, Chbeir R, Dipanda A. Discovering relationship types between users using portraits and shared photos in a social network[M]. Kluwer Academic Publishers, 2013

[21] Mazhari S, Fakhrahmad S M, Sadeghbeygi H. A user-portrait-based friendship recommendation solution in social networks[J]. Journal of Information Science, 2015, 41(3):284-295