Research on User Portrait Technology of Scientific Researcher for University Scientific Research Management

Shuqin Li
North China University of Technology, Beijing 100144, China

Abstract
Aiming at the problem of single angle of description and unintuitive presentation of information for university scientific researchers, we integrate multi-source heterogeneous data, such as information of scientific researchers, projects, papers, monographs, awards, patents and standards to draw user portraits of scientific researchers in university and accurately portray the "thousands of people and thousands of faces", comprehensively understand the information of scientific researchers, and objectively evaluate their research level, in order to improve the efficiency and level of scientific research management, and provide references for precise scientific research services and scientific decision-making.

Keywords
Scientific Research Management; User Profile; Scientific Decision-making; Data Visualization.

1. Introduction
Scientific research management is an important part of scientific research work in university. It runs through the entire process of scientific research activities in university and plays the role of integration, guidance, encouragement, service and supervision for scientific research in university. Scientific researchers carry out scientific research activities and undertake a large number of scientific research projects in university, which will generate massive amounts of heterogeneous scientific research data. For a long time, traditional data statistical analysis methods have been used to process and organize relevant data, which is not only a squander of time and energy, but also fails to effectively dig out the scientific research value hidden behind the data, resulting in a huge waste of scientific research resources [1].

Scientific researchers are the main body of scientific and technological activities, and their relevant information plays a pivotal role in scientific research, project review, achievement transformation, decision-making consultation. With the rapid development of mobile Internet, big data and artificial intelligence, the relevant information of university researchers presents characteristics such as dynamic, massive, and multi-source heterogeneity. The effective collection, management and analysis of scientific researcher information will help grasp the status quo of scientific research, accurately describe the characteristics of scientific researchers, and then fundamentally change the management and decision-making mode of traditional scientific and technological work.

Aiming at the bottlenecks and problems in scientific research management, this article uses user portrait technology to collect and integrate multi-source heterogeneous data, and take scientific researcher, papers, monographs, projects, awards, patents and standards as the entry point to draw user portraits of university scientific researchers, and visually display university scientific research management data. So as to provide scientific reference for scientific research management decision-making, and improve the efficiency and level of scientific research management.
2. User Portraits of Scientific Researchers

2.1. User Portrait Definition

User portrait is user information labeling, by collecting data in various dimensions such as users' social attributes, consumption habits, preference characteristics. And then portraying the user or product characteristic attributes, analyzing and counting these characteristics, and mining potential value information. The purpose is to abstract the full picture of the user's information, and user portrait is an information collection composed of user information obtained from massive data [2-3].

The first person who proposed the concept of user portrait was Cooper, the father of interaction design. He believed that "user portrait can represent a real user, and a user model established by using the user's real data." Regarding the constituent elements of user portrait, T. Lafouge et al.[4] think that user characteristic information includes stable factors related to users (such as users' basic personal information, behavior information, and habit information) and variable factors (such as search environment and search target) two parts. Guimaraes et al. [5] classified the user's basic literacy, education level, social relationship, work status, location, time information and other information as the constituent elements of the user portrait. Chinese scholars have also proposed different component dimensions of user portrait from different perspectives and fields. For example, Zeng Jianxun [6] proposed to construct user portrait tags based on the user's professional background, interest preferences, and special tasks. Liu Haiou et al. [7] and Chen Jing [8] added the user's contextual attributes to the study, dividing the user portrait into four dimensions: natural attributes, social attributes, interest attributes, and ability attributes.

User portrait technology has been widely used in areas such as personalized recommendation, e-commerce, and social networking. In the field of scientific research management, due to the existence of data island, data empowerment has not been truly realized. For example, massive scientific research data is scattered on the Internet, such as CNKI, Google Scholar, Baidu Academic, EI, and Web of Science. These websites contain a large number of researcher papers and research social data. In order to achieve accurate evaluation of scientific researchers, it is necessary to integrate various scientific research data and break the data island.

2.2. User Portraits of University Researchers

Researchers are the source of scientific research creativity in university. The characteristics of user portraits of research researchers must revolve around the entire process of scientific research activities. Scientific research activities mainly include scientific research management activities and specific scientific research work. Scientific research management activities are generally led and organized by scientific researcher with a high academic level, including guidance, communication, evaluation and incentives for scientific researchers, and are the fundamental guarantee for the implementation of scientific research. Specific research work includes knowledge accumulation, scientific research project establishment, scientific research exploration, achievement formation, and achievement release. In the process of scientific research, it is always accompanied by various needs of scientific Researchers, such as the need for obtaining papers, the need for new research on topics, the needs for academic exchanges, and so on. These subjective needs are reflected in the specific activities and behaviors of scientific researchers around scientific research. Collecting, sorting, researching and tracing the relevant behaviors and needs of the whole scientific research process into user portrait, which is conducive to scientific research managers to identify the needs of scientific research users, predict changes in subjective and objective factors, and realize personalized, professional and precise scientific research decision-making and scientific research management services.
The user portrait of scientific researcher is a labeled and visualized user model abstracted from the basic information, research habits, and research behavior of the researcher. User portrait technology applied to scientific research services in university, it has the following advantages.

(1) Applied to the whole life cycle service of scientific research activities. According to the characteristics of scientific researchers, grasp the status of scientific research process and provide full life cycle services. For example, provide professional project application guidance before project application; provide application guidelines, precautions and progress information in project application; provide refined project process management after project application.

(2) Applied to academic exchanges and cooperation. Using user profile technology to analyze the behavior of scientific research users to promote group cooperation, it can not only facilitate scientific research cooperation and exchange activities, but also promote potential cooperation and exchanges in different regions, units and professional fields.

(3) Applied to convenient scientific research management and evaluation, including talent assessment and talent evaluation. The user portrait of scientific researcher objectively evaluates the scientific research ability and reflects the scientific research experience and academic ability. It is possible to provide targeted academic guidance and encouragement to individuals with little experience and weak abilities, and encourage individuals with rich experience and strong abilities to make breakthroughs and enterprising. When scientific researchers are introduced, the scientific research potential of scientific researchers can be easily and efficiently evaluated.

2.3. The Process of Constructing User Portraits of Scientific Researchers

There are various sources of relevant data for scientific researchers in university. Different data describes different attributes and presents data characteristics such as dynamic, massive, and multi-source heterogeneity. The integration of multiple data sources expands the data collection and can improve the accuracy and comprehensiveness of User Portrait. Based on multi-source heterogeneous data, the attributes of scientific researchers are analyzed to form complete user portrait of scientific researchers. The construction process is shown in Figure. 1.

![Figure. 1 The Construction Process of User Portraits of Scientific Researchers](image)

2.3.1. Data collection

There are four main types of scientific researcher data obtained from various information sources, including basic attribute data of personnel, achievement data of scientific research, academic expertise data, and cooperation data of scientific research. The basic attribute data of personnel refers to the basic attributes of scientific researchers, including name, gender, age, job title, and educational background. The achievement data of scientific research are the crystallization of the hard work of each scientific researcher, revealing their scientific research ability and academic influence in various disciplines, including journal papers, conference...
papers, academic monographs, patents, etc. Academic expertise data represents the research direction of scientific researchers. Cooperation data of scientific research refers to the data generated by the cooperation of scientific research members. The content and collection methods of various data are shown in Table 1.

### Table 1. Data Source of Scientific Researcher Portrait

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Data Sources</th>
<th>Main Information</th>
<th>Collection Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic attribute data of personnel</td>
<td>Baidu Encyclopedia, Personal Homepage</td>
<td>Name, Gender, Age, Job Title, Educational Background</td>
<td>Python</td>
</tr>
<tr>
<td>Achievement data of scientific research</td>
<td>CNKI, EI, Web of Science and other electronic literature libraries, National Fund website, and State Intellectual Property Office</td>
<td>Paper Information</td>
<td>Database export</td>
</tr>
<tr>
<td>Academic expertise data</td>
<td></td>
<td>Academic Conference Report</td>
<td>Python</td>
</tr>
<tr>
<td>Cooperation data of scientific research</td>
<td></td>
<td>Research Project Information</td>
<td>Manual collection</td>
</tr>
</tbody>
</table>

#### 2.3.2. Data Fusion and Integration

Data fusion is a high-level knowledge organization. Through the process of heterogeneous data integration, disambiguation, processing, reasoning verification, and updating of data from different data sources, it realizes the integration of data and unifies multi-source heterogeneous source data into standard structured data, which is convenient for user portrait. First, we need to format and extract heterogeneous information, mainly for named entity recognition for information in text or pictures. Named entities in this paper mainly are name entity, such as personal names, organization names, place names, and time expressions, such as date and time. Secondly, the data is cleaned, missing values are filled in with empty values or special values, and duplicate data is deleted.

Data integration can integrate data from multiple data sources to solve the problem of data redundancy caused by different data sources having different names for the same attribute. According to the data type and source of the scientific researcher, the data are stored in the achievement database of scientific research, academic expertise database, cooperation database of scientific research and basic attribute database of member, and unified fields are used to describe the scientific researcher's information. The specific fields are shown in Table 2.

### Table 2. Researcher-related databases and information fields

<table>
<thead>
<tr>
<th>Database Name</th>
<th>Information Type</th>
<th>Main Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database of basic attributes of researchers</td>
<td>Personal Information</td>
<td>Name, Gender, Age, Job Title, Educational Background</td>
</tr>
<tr>
<td>Achievement database of scientific research</td>
<td>Achievement Information of Scientific Research</td>
<td>Journal papers, Conference papers, Invention patents</td>
</tr>
<tr>
<td>Academic expertise database</td>
<td>Academic Expertise Information</td>
<td>Research direction, Research results</td>
</tr>
<tr>
<td>Cooperation database of scientific research</td>
<td>Partnership Information</td>
<td>Collaborators, Cooperative projects, Cooperative achievements</td>
</tr>
</tbody>
</table>

#### 2.3.3. Visualization of User Portrait

The visualization of user portrait can present user characteristics intuitively and clearly. The visualization methods include tag cloud (word cloud), various statistical graphics (histograms, line graphs, etc.), character images combined with user tags. The size of tags in the tag cloud is
proportional to the significance level of user characteristics, which can be used for visual presentation of single and group users. Statistical graphics are more used for group users. Portraits combined with tags can reveal the characteristics of users in multiple dimensions, and are mostly used for single users. There are many tools for generating tag clouds (word clouds), such as Word Art, Wordsift, TagCrowd, etc. You can also use Python Word Cloud Package to write code. There are a wide variety of tools for generating statistical graphics, which can be selected according to the characteristics of the data and the application field. If you want to generate a user distribution map, you can choose Modest Maps, Leaflet, etc; if you want to use more comprehensive user data, you can choose I Charts, Fusion Charts Suit etc. There are fewer tools to generate character images combined with user tags, and it is necessary to draw them by using visualization software according to the effect of user portrait. In order to describe the characteristics of scientific researchers more accurately, this paper uses the method of character images combined with user tags to draw user portraits of scientific researchers.

3. Instance Verification

In order to verify the effectiveness of the method, this paper selects researchers from D University as an example for verification, collects multi-source heterogeneous data, forms a related structured database through data fusion and integration, and then builds user portraits of researchers through portrait visualization technology.

3.1. Data Collection and Data Fusion

Data collection is the basis of user portrait. The main sources of scientific research achievement are CNKI, Wanfang, Weipu, Web of Science and EI databases. Scientific research project data comes from the National Natural Science Foundation of China, the National Social Science Foundation project database, and the National Science and Technology Report Service system provides special reports, progress reports, final reports and organizational management reports of scientific research activities. In order to achieve real-time acquisition of multi-source data, the Python Scrapy framework is used to implement distributed data crawling. The collected data includes the following categories:

Information of scientific research personnel: name, department, gender, position, education, age, professional title, education background.

Scientific research project: project name, department, person in charge, project type, project nature, project source, start time, completion time, approved amount.

Standards: standard name, department, maker, approval number, standard category, application time, approval time, personal ranking, and unit ranking.

Patent: Patent name, department, owner, application number, approval number, patent category, application time, approval time, individual ranking, unit ranking.

Paper: paper name, department, author, journal name, journal category, retrieval status, publication time, funding status.

Since data collection is obtained from multiple sources, there are problems such as conflicts, inconsistencies, and ambiguities in the data. It is necessary to integrate the information of scientific researchers on multiple data sources in order to have a more comprehensive understanding of the overall situation, and provide a basis for accurate portraits and multi-level portraits of users. Multi-source data fusion is mainly to eliminate data conflicts, inconsistencies, and ambiguities, improve data quality, and lay the foundation for more comprehensive user portraits.
3.2. **Construct User Portrait of Scientific Researchers**

User portraits of scientific researchers are the process of transforming fragmented scientific research data into high-precision label. Therefore, it is necessary to construct a reasonable, multi-level and comprehensive labeling system, in order to reasonably evaluate scientific research performance of scientific researchers. According to a set of defined labeling systems, on the basis of data analysis and mining, the knowledge and labeling of scientific researchers’ data are realized. Accurate labeling and clustering analysis of the feature tags of the user portraits can help provide users with highly matched scientific research services and assist scientific researchers in conducting high-efficiency and high-level scientific research activities.

The user portrait of scientific researcher contains accurate descriptions of scientific researcher's information, standards, scientific research awards, scientific research projects, paper achievements, patents and other data. The machine learning model such as Bayesian network is used to model the processed data, continuously train the model, and update user tags to make the modeling results more and more accurate. Finally, the user tags are displayed in the form of data visualization in order to more accurately describe the user characteristics. The constructed user portrait of the scientific researcher is shown in Figure. 2.

**Figure 2. User Portrait Visualization of Scientific Researchers**

4. **Conclusion**

User portraits make full forthputting of multi-dimensional user data to describe and model users, which can comprehensively and meticulously abstract the full picture of tagged users, which has important application value. More in-depth data mining and analysis based on user profile can provide auxiliary decision support for scientific research management to improve the scientificity, correctness and predictability of decision-making. In terms of scientific research management, user portraits of scientific researchers describe the characteristics of scholars’ interests, research directions, scientific research capabilities, scientific research cooperation relationships, etc, which can provide decision-making references for finding suitable scientific research partners or experts in specific fields.

This article starts with the specific scientific research activities of university scientific research users, comprehensively summarizes the various characteristics involved in the user portraits of scientific researchers, and analyzes the data sources and data characteristics of various
characteristics. The user portrait of the scientific researcher was constructed from the dimensions of scientific researcher information, standards, scientific research awards, scientific research projects, paper achievements, patents, etc, and finally displayed in a visual manner, and the feasibility of the method was verified. User portraits not only fully describe the information of scientific researchers, but also make the display of information more intuitive. It plays an important role in the evaluation of scientific researchers, the recommendation of review experts and team formation. It improves scientific research management and service strategies for university and realizes effective and accurate intelligence. Provide certain reference for scientific research decision-making.

Acknowledgments
The research results of this thesis were funded by "North China University of Technology Yuyou Talent Support Program Project (20XN213/011)".

References