ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

# Study on Information Recommendation of Scientific and Technological Achievements Based on Client Behavior Modeling and Big Data Mining

[1]TARINI PRASAD PATNAIK, *Gandhi Institute of Excellent Technocrats, Bhubaneswar, India*
[2]SHUBHANKAR NATH, *Ghanashyam Hemalata Institute of Technology and Management, Puri, Odisha, India*

***Abstract*— This paper brief presents the source, the idea and qualities of enormous information of logical and innovative accomplishments. The strategies and procedures of enormous information investigation are looked into. The most common way of offering customized assistance in view of client conduct demonstrating and huge information mining is investigated. The data proposal administration of logical and innovative accomplishments in light of huge information investigation is examined. Joined with the qualities of customized administration in large information climate, the development technique of client conduct model is proposed. The model structure technique and the customized administration plot are given toward the end.**

***Keywords-big data; scientific and technological achievements;information recommendation; ontology***

## I. INTRODUCTION

### *A. The Source of Big Data of Scientific and TechnologicalAchievements*

With the rapid development of science and technology, a large number of scientific research outputs data are accumulated in the process of scientific research, such as scientific papers, patents and software copyrights, research reports, etc. The information search, analysis and services become more and more important, and on this basis, some application oriented special databases are formed, such as library of scientific and technological achievement, library of scientific and technological talent, library of scientific research project. The scale of these data is becoming larger and larger, and the structure is becoming more and more complex, and the requirements for deep analysis and mining of data are becoming higher and higher. These rich data resources show great reference value for mining and decisionmaking through association and integration.

With the expansion of network information access channels, information acquisition becomes more and more easy. For huge amounts of information of scientific and technological achievements, the traditional search has been unable to meet the needs of users. Researchers are exploring the use of data mining, analysis and visualization tools, provide dynamic tracking, customized push, theme research, strategic decision-making research in-depth professional information service for users.

### *B. The Concept of Big Data of Scientific and TechnologicalAchievements*

From the perspective of the data scale, it is generally considered that the data size more than petabytes is called big data. From the perspective of techniques and methods, traditional database techniques and methods can't deal with massive or unstructured data sets, which are called big data. From the perspective of application value, big data is the sum of the massive data analysis of multi-source heterogeneous cross the associated domain generated by the decision-making process, business model, scientific paradigm, life style and concept of disruptive changes based on morphology. Therefore, big data has large size and structure diversity, timeliness and other characteristics. New computing architectures and intelligent algorithms are needed to deal with big data. The application of big data emphasizes the correlation rather than a causal relationship, which aims to discover new knowledge and insight andscientific decision making.

The big data of scientific and technological achievements is a multi-source heterogeneous large-scale data composed of scientific research activity elements and related factors. The elements of scientific research activities including scientific research input, e.g., R & D investment, scientific research project; the scientific research subject, e.g., research institutions and researchers; the scientific research platform conditions, e.g., scientific instruments and equipment; the scientific research process, e.g., scientific experiment data; scientific research communication, e.g., academic conference; scientific research output, e.g., papers, patents, and reports and transformation of scientific and technological achievements; as well as the scientific research management, e.g., project application, achievements management. The relationship between the data reflected by these elements and the relevant data of the scientific research activity constitute the content of the big data of scientific and technological achievements.

### *C. The Characteristics of Big Data of Scientific and Technological Achievements*

Big data from all areas, though representing different things and hidden different value information, can be described by the following characteristics: namely Volume (large capacity), Velocity (fast processing), Variety (type and nature)

ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

and Veracity (vary quality).

- Volume refers to a large scale of data, although there is no absolute standard of capacity, but generally it is more than ten terabytes. The scale of data collection, storage and distribution is beyond the management ability of traditional management technology.
- Velocity refers to the high speed of data processing, and the generation of big data is a fast dynamic process. All kinds of data flow and information flow are generated, transmitted and processed at high speed.
- Variety refers to the diversity of data types. Except for literal data, it also includes multimedia data, such as images, graphics, video and sound. That is, the objects that are handled include structured data, semi-structured data and unstructured data.
- Veracity means that high value information is hidden in the data, which needs to be extracted by machine learning and data mining methods. The data quality of captured data can vary greatly, affecting the accurate analysis. The value density of data is very low, so it also increases the difficulty of value mining.

## II. Methods and Techniques for Big Data Analysis

To achieve accurate information service of scientific and technological achievements, the relationship between bigdata and information service should be solved first, which requires the support of methods and technology. Based on information science and technology achievements of big data analysis elements, in addition to the traditional statistical indicators, also includes association relationship, temporal variation, spatial distribution and difference index. There are a series of analytical methods around these factors, such as scientific measurement, correlation analysis, trend analysis, signal analysis and cluster analysis. In addition to general big data techniques, such as distributed parallel computing, big data analysis and visualization, multi-dimensional data association analysis, it also includes multi-source heterogeneous data integration technology, knowledge system technology based on knowledge extraction and ontology construction, user demand detection technology, new computing technology based on scenario calculation.

### *A. Multi-source Heterogeneous Data Integration*

Policy texts, hot topics, technology trends, and cutting- edge technologies are all factors that have influence on users'needs. Each factor has multiple data support and different data are distributed in different sources or channels. In the era of big data, it is the basis and prerequisite for many work to obtain multi-source data quickly and accurately. Multi- source data reflect the industry from different perspective. These data fusion aggregation are analyzed by correlation. By cross-checking and complementing each other, they can reveal things more comprehensively. And these data can provide powerful data support and decision reference for strength comparison and evaluation, competitive environment scanning and situation analysis, strategic opportunity selection and expansion. Many problems need tobe solved, such as how to relate mapping between different data sources, unify and analyze data among heterogeneous data, data conflict, data loss, data duplication and so on. Howto construct a multi-source and heterogeneous data resource usually involves a series of technical problems.

### *B. Knowledge Extraction and Ontology ConstructionTechnology*

After constructing the big data information of scientific and technological achievements, fast acquire knowledge from unstructured data, expressed it as an ontology that can be understood and deduced by computer and combined with intelligent algorithms, such as deep learning, has become the key of intelligent recommendation service. Knowledge extraction is the process of acquiring all kinds of knowledge from all kinds of data and information resources. It is a process of extracting knowledge from a variety of media resources, such as text, image, video, and audio, and discovering important patterns from data set. Generated by the knowledge extraction from multi-sources of data obtained by using some form of knowledge representation, knowledge element and semantic relations are complete and correct and unambiguous. As the input of subsequent knowledge fusion, knowledge extraction and organization are completely dependent on the understanding of the relationship between knowledge mining and organization, and to a certain extent, reflects the implicit contact. Therefore, the support of a large number of ontology and thecorrelation analysis between cross-domain knowledge are essential. Ontology can provide effective support for the intelligent recommendation, machine translation, knowledge graph and so on.

### *C. New Computing Technology Based on ScenarioCalculation*

Under the environment of big data, analysis of data mining is more and more attention to the user scenarios. Only by analyzing user scenarios and focused on the changeof scene, real-time update the service contents and methods, the demand of users can be better understand and meet. Thus the degree of acceptance and satisfaction of the product and service could be greatly improved by the user. Traditional data mining is mainly calculate the commonality and association between data, such as

ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

association rule, cluster analysis, social network analysis, vector space model and so on. When information is huge, finding out some commonalities and associations is a very popular way of thinking. Usually, difference analysis is also needed, it will become a new challenge to transfer the common correlation calculation among documents to the difference comparison calculation.

## III. Information Recommendation Service of Scientificand Technological Achievements Based on Big Data Analysis

The information recommendation service scheme of scientific and technological achievements is composed of three parts, which includes user behavior ontology construction, user interest and requirement ontology mining, and personalized recommendation service based on ontology,as shown in Figure 1.

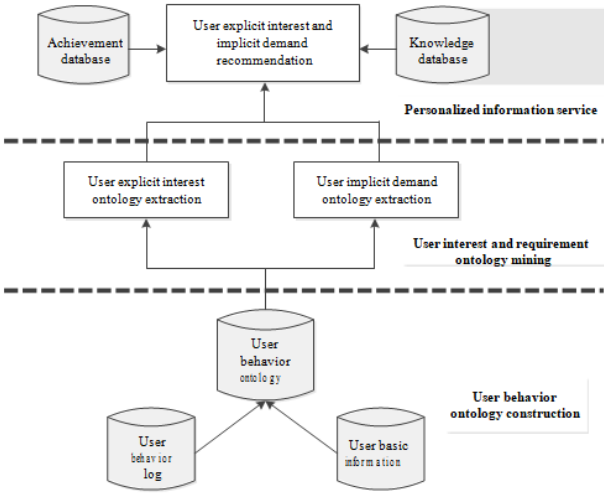


Figure 1. The information recommendation service scheme of scientific and technological achievements.

*User behavior ontology database construction process*: Generate the corresponding database field according to the user behavior ontology model schema. Combining with the ontology concept of knowledge domain, the data of corresponding fields is analyzed from the original user service logs and the database of user basic information. Building user behavior ontology database, the database needs to be updated according to change log in terms of the ontology model. That is, the concept is extracted by analyzing and mining the original log information of the user. After the specification of ontology technology, it is added to the ontology repository.

*User explicit interest and implicit demand ontology extraction process*: Based on the statistical analysis of each user behavior log database, the user's explicit interest ontology term is extracted to accurately reflect the user's preferences. The implicit demand ontology is extracted by data mining of all user logs of the library. After the systematic learning and mining of all users' historical behavior in the user behavior log library, the terms closely related to the user's explicit interest ontology term are discovered. It can meet the needs of user extensions and recessive.

*Personalized information recommendation process:* When the user login service platform, according to the user interest ontology and the query related database, e.g. research reports, papers and patents and software copyright, the results will be pushed to the user, which achieves personalized recommendation of user's explicit interest. If the user further retrieves knowledge on the service system, the retrieval statement entered by the user is analyzed based on context, the feature words are extracted, and the ontologyterms of implicit demand are matched, the potential result will be pushed to the user, which achieves the personalized recommendation of users' implicit demand.

### A. User Interest Ontology Extraction Process

The user's interest degree is used to describe the degree of user interested in the scientific and technological achievements. Through statistical quantitative rating of user behavior, such as browsing, retrieving, collecting and commenting on scientific and technological achievements, reflect user's interest in scientific and technological achievements, the calculation expression is: interest degree =browsing + retrieval + collection + comment. The calculationrules are as follows: for the "browse" behavior, first determine the time threshold of browsing, suppose the threshold is 30 seconds. When the user of a continuousbrowsing time is greater than or equal to the threshold value, the interest degree of the user is increased by 1 point. If the user continues to browse after discontinuous, then the interest degree is increased by 1 point, otherwise 0 point. In the same way, if the user has a "retrieval" and "collection" behavior, the interest degree is increased by 1 point, otherwise 0 point. When the user has "comment" behavior, ifthe comment is positive, then the interest degree is increasedby 1 point, otherwise 0 point. The cumulative score represents the interest of users for the achievement information of science and technology. When user interest degree value of the information of scientific and technological achievements to achieve predefined threshold, extract the corresponding terms of ontology, and stored in the user interest ontology. The user interest ontology is extracted from the user's behavior log database only for a single user. The extraction process based on the MapReduce framework under the Hadoop big

ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

data platform is as follows:

- Preprocessing user log data in the MapReduce framework. The user ID is mapped to the Key value of < Key, Value > in the MapReduce framework, and the number of users browsing, retrieving, collecting and reviewing is mapped to Value.
- Reduce statistical phase. The Value of the user's interest is calculated according to the "user interest" statistical rule mentioned above, and the various values of the same ID user are accumulated.
- Extracting and preserving the terms of user interest ontology. If the user's interest degree is greater than the threshold set by the system, the ontology terms are extracted and saved as the interest ontology. For example, if the user's "browsing, retrieval, collection, comment" behavior reach a certain number of times for the achievements information related to "artificial intelligence" of the ontology terms, when the user's interest degree reaches the threshold, the term "artificial intelligence" is extracted and saved as a user's interest ontology.

### *B. User Requirements Ontology Extraction Process*

Data mining algorithm is applied to extract user requirement ontology from user behavior log database. Related algorithms include association rule mining and collaborative filtering. The association rule algorithm is used to discover the correlation between the terms of the ontology from the log of all the users, and the frequent data sets are obtained by the statistical data items. Here, association rule is used to extract the process of user requirements ontology, which including 3 steps:

- Preprocessing user log data in the MapReduce framework. It is the same as the first phase of the user's interest ontology extraction process, only necessary to operate on the entire user ID.
- Calculation of support. Firstly, the frequency of single ontology term is accumulated by Reduce method. That is single term support degree P(A). Then, the two ontology terms are accumulated when they appeared simultaneously. That is P(AB), the support degree of binary group or two ontology terms.
- Calculation of confidence. The support degree of twotuple is divided by the support degree of single term, then the confidence degree is obtained. The greater the value of confidence, the higher the probability of the two ontology terms appear simultaneously. When the confidence degree is greater than the preset threshold, the algorithm considers that the ontology term A and the ontology term B are frequent correlation items and credible. A set of frequently associated items is composed of a large number of frequently associated items. When the user retrieval systems, the search words are composed of a feature word in a retrieval statement and the other ontology term that is frequently associated with it. As the implicit demand, the retrieved results are recommended to the user. Because of the frequent items sets are results of association rules mining based on the behavior of all user logs, the frequent association items can be shared by all users, not only for one user's personalized recommendation.

### *C. Personalized Recommendation Service based on UserInterest and Demand*

When the user logs in, the system automatically gets the ontology term "robot" that the user is interested in. And uses this term as a query word to query relevant knowledge of scientific and technological achievements from the knowledge base and recommend them to the user. Suppose this user input retrieval statement is "what are the core technologies in the field of artificial intelligence?" The system uses semantic word segmentation technology to map an ontology term "artificial intelligence". By querying frequent association item sets, "artificial intelligence" and "automatic speech recognition", the frequent items are found, so another ontology term "automatic speech recognition" is obtained. Then "artificial intelligence" and "automatic speech recognition" as key words, the results obtained from the associated results library are recommended to the user. The query results of "artificial intelligence" as ontology terms are directly derived from user problems, it is normal retrieval. The query of "automatic speech recognition" as the ontology term comes from the results of data mining, it is the personalized recommendation of the implicit demand.

## IV. Conclusion and Future Work

This paper studies the ontology modeling of user behavior and combined with big data mining technology, provides a personalized service scheme for users to recommend information of scientific and technological achievements. In this study, different processing logic is designed for "login" and "retrieval". When users log in, the system uses user's explicit interest ontology terms as retrieval words to achieve recommendation, it can get higherrecognition from users. When users search in the system, it indicates that users have deeper knowledge needs. Then the system recommends the retrieval results of the ontologyterms based on the data mining algorithm to the users, it canmeet the user's implicit needs. This

ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

design scheme that based on user behavior ontology model, combining explicit interest and implicit demand to implement personalized recommendation can help users to gain information of scientific and technological achievement accurately that they are interested in. In the process of model design, the application of associated data technology can facilitate the sharing and interoperation of massive data, such as achievements base and knowledge base. Based on Hadoop big data platform and MapReduce framework technology, the analysis and processing of user behavior log can be morereal-time and efficient.

The complexity of big data of scientific and technological achievements is mainly embodied in the width of the data. Therefore, further research will focus on strengthening scientific data management and effectiveness in the process of data collection, storage, processing, analysis and decision-making. Under the premise of reducing the load of big data analysis platform, system structure of big data analysis platform, the availability of the analysis algorithm and the value density of data resources will be improved constantly. It will provide a scientific, comprehensive, real- time and reliable big data analysis and decision support for personalized information recommendation service and intelligent management of scientific and technologicalachievements.

## REFERENCES

[1] Shim K, "MapReduce Algorithms for Big Data Analysis," International Workshop on Databases in Networked Information Systems, 2013, vol.5, no.12, pp.44-48.

[2] Dean J, Ghemawat S, "MapReduce: Simplified data processing on large clusters," Communications of the ACM, 2008, vol.51, no.1, pp.107-113.

[3] Lublinsky B, Smith K T, Yakubovich A, "Professional hadoop solutions," Birmingham: Wrox Press, 2013.

[4] M. Sarwat, J. Avery, and M. F. Mokbel, "RecDB in Action: Recommendation Made Easy in Relational Databses," In Proceedings of the International Conference on Very Large Data Bases, VLDB, 2013, pp.1242-1245.

[5] S. B. Roy, S. Thirumuruganathan, G. Das, S. Amer-Yahia, and C. Yu, "Exploiting Group Recommendation Functions for Flexible Preferences," In Proceedings of the IEEE International Conference onData Engineering, ICDE, 2014.

[6] Agrawal D, Budak C, Abbadi A, Georgiou T, Yan X, "Big Data in Online Social Networks: User Interaction Analysis to Model User Behavior in Social Networks,", In Proceedings of the 9th International Workshop on Databases in Networked Information Systems, 2014, vol8381, pp.1-16.

[7] Xing E P, Ho Q, Dai W, "Petuum: A new platform for distributed machine learning on big data," IEEE Transactions on Big Data, 2015, vol.1, no.2, pp.49-67.

[8] M. Sarwat, "Interactive and Scalable Exploration of Big Spatial Data
- A Data Management Perspective," 16th IEEE International Conference on Mobile Data Management, 2015, pp.263-270.

[9] M. Sarwat1, R. Moraffah, M. F. Mokbel, J. L. Avery, "Database System Support for Personalized Recommendation Applications," IEEE 33rd International Conference on Data Engineering, 2017, pp.1320-1331.

[10] Pulice C, "Discovering User Behavioral Features to Enhance Information Search on Big Data," Acm Transactions on Interactive Intelligent Systems, 2017.