

Analysis on the Application of Data Mining in Science and Technology Information Management

Li Du¹, Yue Zhang¹, and Qian Sun²

¹Chengdu Neusoft University, Dujiangyan Sichuan, China, 611844

²Sichuan Electric Power Design & Consulting Co., LTD, Chengdu Sichuan, China, 610041

Keywords: science and technology information management; data mining; rules; contact; decision

Abstract: Science and technology information management has accumulated a great deal of data in its development process while improving the level of economic and social development and promoting the construction of a harmonious society. With data mining techniques, many useful data can be presented to provide reliable decision-making basis for management and decision. However, traditional information processing methods have been unable to cope with large data groups. Therefore, modern information technology must be used to conduct comprehensive research and discover the rules of association between data mining and science and technology information management, and carry on the correlation analysis between the science and technology project, personnel and result, in order to discover and speculate the potential law and relation between science and technology information, provide support in the science and technology information management decision-making, thus help the standardization of science and technology information management.

1. Introduction

Nowadays, with the popularization and application of emerging technologies such as the Internet, Technology management informationization has achieved tremendous development and accumulated a large amount of data, it is in the sense that it has entered the era of big data. But at the same time, how to correctly handle these data and extract useful information has gradually become a difficult problem, the traditional extraction method has been difficult to meet the current precision requirements and big data decision requirements. Therefore, we must borrow the latest technology and tools of the information society and use data mining tools to explore the laws and links of huge data groups controlled by the technology management informatization, thus providing support for decision making in technology management and to help standardize management of science and technology.

2. Data Mining and Association Rules

"Data mining" is also called "data exploration" or "data acquisition."^[1] It is a key step in database knowledge discovery. Usually its specific meaning refers to searching and extracting a certain type of information through a specific algorithm in a large amount of data, and then selectively extracting and refining from these disorganized information, and finally associating with the research object to explores the law in the development history of the research object, and makes predictions and plans for the future development trend of the research object. At present, commonly used data mining technologies include: web page mining, decision trees, association rules, deviation analysis, regression analysis, and so on^[2-7]. In data mining technology, association rules are a common and very important model. Its form is $X \rightarrow Y$ implied, where X and Y are the leader and successor respectively. In simple terms, its working mode is to find the strong rules that both the credibility and the support satisfy the given value requirements in a given database, and then analyze the database again to find the hidden association rules between these data.

In data mining association rule mode, Apriori algorithm plays a very important role as a frequent

itemset algorithm. Apriori algorithm mainly adopts the recursive idea^[8-10], carries out layer-by-layer search iteration, and finds out the links between data sets among a large number of data. The workflow is roughly as follows: retrieval in all related matters according to the preset minimum support(minsup) and minimum confidence(minconf). In order to find frequent 1 item set, first recorded as L1, then find out the frequent 2 item set according to L1, and write it as L2, then find 3 frequent items according to L2, mark it as L3, and so on, until all frequent item sets or frequent item sets are found out. The next task is to find the required association rules we need in these project sets or item sets. In the follow-up content of this paper, this Apriori algorithm will be used to analyze the relationship between the examples of scientific and technological information, so as to find out the laws and connections between the examples of scientific and technological information, and provide support and help for relevant decisions.

3. Application of Data Mining in Science and Technology Information Management

The experimental data selected in this article comes from a city's scientific and technological information management system. Since the establishment of the system, it has collected 512 scientific and technical personnel in various fields such as agriculture, medical treatment, education, health, and chemical engineering. In addition, 899 scientific and technological achievements and 2014 science and technology projects were also collected, involving various types of workers, peasants, merchants, international trade and cooperation. What needs to be stated here is that data mining includes three stages: data preparation, data mining, and results analysis. After selecting suitable examples, this article explores the laws and links between basic information such as the age, title, workplace of scientific and technical personnel, published science and technology research results, and participating mobile research projects. Specific work steps and processes are as follows:

3.1 Data preparation.

In the data preparation phase, data selection, preprocessing and transformation are mainly performed. First, data selection refers to the selection of the relevant data to be studied from the given database to form the target data. According to the research content selected in this article, the database should be the recorded scientific and technical personnel table, title list, unit scientific and technological achievements and awards, etc., select the appropriate data from this database to build a data warehouse that connects science and technology personnel (A), technology projects (B), and scientific and technological achievements (C), such as A-B, A-C, A-B-C, etc. The specific content is as shown below:

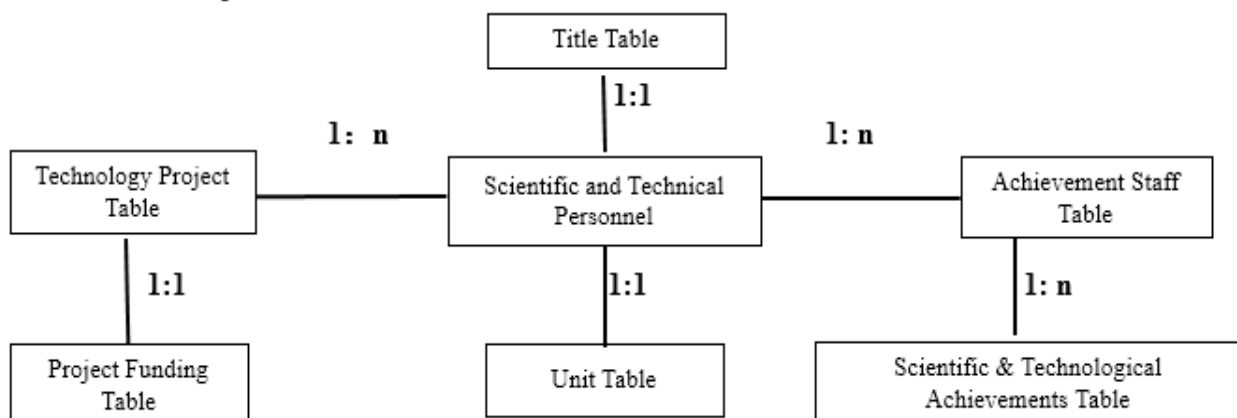


Fig.1

The rules of the data mining algorithm, organize the extracted relevant data, make possible values and encode the data (see Table I). Then establish the scientific and technical personnel-scientific and technological achievements-technical project data structure as described above (see Table II).

Table I: Symbolic representation of possible data values

	Significance	Symbol	Significance	Symbol	Significance	Symbol	Significance
I 01	20-30 years old	I 27	Municipal achievements	I 35	>200,000 RMB
I 02	30-40 years old	I 21	P Unit	I 28	Provincial Progress Award	I 36	Key Project
I 03	40-50 years old	I 22	Senior Engineers	I 29	Municipal Progress Award	I 37	Major Project
I 04	50-60 years old	I 23	Deputy Senior Engineers	I 30	Provincial Project	
I 05	60 years old +	I 23	Mid-level Engineer	I 31	Municipal Project		
I 06	A Unit	I 25	Primary Title	I 32	Within 50,000 RMB		
I 07	BUnit	I 26	Provincial level achievement			

Table II: Scientific and Technical Personnel - Scientific and Technological Achievements - Structure of Science and Technology Project

No.	Age	Title	Unit	Achievement	Reward	Project type	Funds
001	52	Deputy Senior Engineers	C	Provincial Level	Provincial Level	Tackle Key Problems	5
002	37	Deputy Senior Engineers	A	Provincial Level	Provincial Level	Major	15
003	56	Senior Engineers	P	municipal Level	municipal Level	Tackle Key Problems	12

3.2 Data mining.

According to the requirements, set Minsup to 6% and Minconf to 79%. Then use Apfiofi algorithm to extract the selected data for mining. The following results are obtained:

Table III Scientific and Technological personnel - Scientific and Technological Achievements - Association Rules for Science and Technology Projects

Item set 1	Item set 1 Support	Item set 2	Item set 2 Support	Association Rules	Credibility
I36	140	I26I36-I27I36	116	$I36 \geq I26-I27$	82.86%
I37	81	I37I30	65	$I37 \geq I30$	80.25%
I30	59	I26I30-I27I30	48	$I30 \geq I26-I27$	81.36%
I33	60	I26I33-I27I33	52	$I33 \geq I26-I27$	75. 36%
3036	42	I26I30I36-I27I30I37	30	$I30I36 \geq I26-I27$	71.42%

Table IV Scientific and Technological personnel - Association Rules for Technology Projects

Item set 1	Item set 1 Support	Item set 2	Item set 2 Support	Association Rules	Credibility(%)
I11	76	I11I23	70	$I11 \geq I23$	92.11
I14	43	I14I23	36	$I14 \geq I23$	83.72
I17	58	I17I23	46	$I17 \geq I23$	79.31
I36	140	I36I23	115	$I36 \geq I23$	82.14
I32	84	I32I23	71	$I32 \geq I23$	84.52
I33	69	I33I23	59	$I33 \geq I23$	85.51
I03I37	37	I03I37I23	30	$I03I37 \geq I23$	81.08
I22I36	59	I22I36I30	54	$I22I36 \geq I30$	91.53
I30I333	25	I30I33I37	20	$I30I33 \geq I37$	80
I31I34	40	I31I34I36	34	$I31I34 \geq I36$	85
I03I22I36	28	I03I22I36I29	26	$I03I22I36 \geq I29$	92.86
I30I33I37	31	I30I33I37I23	27	$I30I33I37 \geq I23$	87.10

Obviously, after preliminary analysis of the data in the table, we can see some conclusions: 1. Among the scientific and technical personnel who participated in scientific and technological projects and produced scientific and technological achievements, senior engineers aged 40-50 accounted for 89.23% (obtained by the scientific and technical personnel-association rule of scientific and technological achievements, $I02 \geq I23$); 2. 71.42% of the scientific and technical personnel involved in the provincial scientific and technological research projects have provincial or municipal scientific and technological achievements (obtained by the table scientific and technical personnel - scientific and technological achievements - association rules for science and technology projects); 3. Among the 40-50-year-olds who participated in major science and technology projects, the deputy senior engineers accounted for 81.08% (obtained by the table scientific and technical personnel- association rules for technology projects). In addition to these conclusions, it can be queried in the above three tables according to requirements.

4. Analysis of Data Mining Results and Related Opinions and Suggestions

After careful analysis of the preliminary mining association rules, the following results can be obtained:

(1) The 40-50 years old senior/ deputy senior engineer is the mainstay of science and technology projects, occupying almost all the positions of science and technology project leader. From this, it can be initially concluded that researchers in the 40-50 age group have strong research consciousness and scientific research capabilities. In view of this situation, relevant scientific research departments can be recommended to increase the degree of emphasis on this part of scientific research personnel, and give enough scientific research funds and certain policy support.

(2) People who have achieved major scientific and technological research results are all over the age of 40. Given the age of this group of scientific researchers, we can suggest that relevant scientific research departments should give full play to their scientific research leading functions and bring more scientific researchers to participate in more scientific research projects.

(3) The scientific and technical personnel who generally have scientific research achievements above the provincial level have more than one municipal-level or higher scientific research achievement, and these scientific research achievements are rarely published on their own, and most of them are done in cooperation with other scientific researchers, enterprises, or schools. As a result, the establishment of scientific research results publishing and exchange platforms can be increased, scientific researchers can be encouraged to have in-depth exchanges with each other, and actively participate in cooperation projects between universities and social enterprises.

5. Conclusion

In today's era of big data, rational use of data mining technology, comprehensive research and analysis of technology management information, access to the inherent laws and links hidden in modern technology management information, and the rational use of these laws and linkages can play a crucial role in the management of science and technology projects, the evaluation of scientific and technological personnel, and the evaluation of scientific and technological achievements. In turn it can then provide support in scientific and technological management decision-making, thus to help standardize management of science and technology and help the further development of scientific and technological information management, thereby forming a virtuous circle.

References

- [1] Rong Ye. Application of Web Data Mining in Information Management[J]. Technology Square, 2017(03):146-149.
- [2] Jianxun Zeng, Jidong Cao, Jing Su. Construction of National S&T Management Information System and Its Impact on S&T Intelligence Work[J]. Journal of Information, 2016,35(09):900-910.
- [3] Yidong Wang. Research on Application of Data Mining Technology in Bank Supervision[J]. Information Security and Technology, 2014,5(03):85-86.
- [4] Zhiwang Jiang, Hongxia Zhang, Huiqin Zhou. Research on Bank Customer Information Management Based on CRM[J]. Heilongjiang Science and Technology Information, 2014(07):166.
- [5] Hui Zeng. Application of Big Data Mining in Engineering Project Management[J]. Technological progress and countermeasures, 2014,31(11):46-48.