

Study on Application of Research Methods of Information Science Based Bibliometrics

Hongfei Sun

College of Economics and Management
Northeast Dianli University, NEDU, Jilin, China
Sunny_bird@126.com

Yanyan Wang, Wei Hou

College of Economics and Management
Northeast Dianli University, NEDU
Jilin, China

Abstract—This paper selected relevant data as the basis data from 19 kinds of Core Journals and analyzed the sample papers with the use of Bibliometrics. The author summarized four characteristics of the application of Informatics research methods of China in the past five years and found some problems and deficiencies of the application of Informatics research methods, the author also put forward the corresponding countermeasures and suggestions.

Keywords- Informatics; Research Methods; Bibliometrics; Cluster Analysis; Co-word Analysis; Word Frequency Analysis

I. INTRODUCTION

One important sign of a mature discipline and field of study is that the discipline or field of study has formed a complete and scientific system of research methods. Since the 1980s, our intelligence circles has risen an academic discussion about the research methodology system of information science, the "three levels", "four levels", etc are relatively more accepted and mature^[1]. With the rise of Bibliometrics and other quantitative methods, this paper analyzed statistically relevant literatures from 2008 to 2012 and constructed co-occurrence matrix of related items and then conducted cluster analysis with the use of relevant data to show the application of Informatics methods in our country in the past five years.

II. ANALYSIS RESULT OF BIBLIOMETRICS

A. Journal distribution

Sample papers are distributed unevenly in the 19 kinds of journals, the basic situation presented is: more relevant papers are embodied in the journals named after "Information" except for "Information and Documentation", "Library and Information". Overall, the ratio that the papers applying Informatics research method published in the journal which named after "Information" is 86.9%, it illustrated that the journals named after "Information" is an important publish position for the application of Informatics research method. Meanwhile, it can be seen that some high-impact journals such as "Intelligence Journal", "Library and Information Service", "Information Theory and Practice" and "Information Science", etc. play an important role on the field of the application of Informatics research method.

B. Authors analysis

There are 7050 authors in the sample papers, in which 4853 authors only have published 1 paper of the application of Informatics research method respectively in the past five years, accounting for 68.8% of all authors. If the authors possessing 15 or more papers are considered of high yield, there are 36 authors meet the condition.

The above table indicates that these methods are rarely used among the authors who studies the application of Informatics research method, a minority belong to high-frequency authors in this field, which is not good for promotion, innovation and development of Informatics research method. If the authors possessing 3 or more papers are considered to be active users of Informatics research method, there are 1100 authors accounting for 15.6% of all, which means 84.4% of the authors are not active in applying Informatics research method. At the same, we can tell that there are a lot of space for the application, promotion and development of Informatics research method^[2].

C. Keyword analysis based on research field

After counting statistics the keywords which have been treated, we calculated word frequency threshold of the keywords according to "word frequency boundary formula"^[3] proposed by Donohue in 1973. In the paper, $I_1=7643$, substituting it into the formula, we got the word frequency threshold 123.13. However, the high-frequency words which meet 123 are "Competitive Intelligence", "Digital Library" and "Knowledge Management", so we can not discover the laws from high-frequency words. Therefore, we selected the keywords which frequency is higher than 19 to be high-frequency words, then we got high-frequency keywords table as is shown in table 1.

The high-frequency words were built into a 69*69 matrix, then we calculated the coefficient Ochiai of the matrix and convert the co-word matrix into a correlation matrix, at last, we converted the correlation matrix into a dissimilarity matrix with 1 minus the data in correlation matrix^[4]. Now, the data in dissimilarity matrix is the distance between keywords of line and row correspondingly. We put the dissimilarity matrix into SPSS19.0 to carry on hierarchical cluster analysis and select "Ward (Deviation squared and Law)" to get the dendrogram of 69 high-frequency keywords, as is shown in figure 1.

It can be seen that the 69 high-frequency words can be divided into 10 categories, One is the algorithm for text

classification; the second category is information retrieval and query expansion; the third category is China and the United States; the fourth category is performance evaluation, government website and E-government (the third and the fourth categories can be put together with slightly broader classification); the fifth category is college students and information literacy, information system, information behavior, information services, information needs and university library; the sixth category is E-commerce and recommendation system, keyword, similarity, search engine and website and labels; the seventh category is competitive intelligence, enterprise and information visualization, indicator system, internet public opinion, information resource, network, evaluation and information, network information resources, progress, quality of service, evaluation model, digital resources, performance evaluation, library, knowledge services, open access, database, analysis, patent, technology innovation, evaluation, model, knowledge organization, digital library and intellectual property rights; the eighth category is factors, virtual community, knowledge sharing, tacit knowledge, knowledge management, knowledge transfer and knowledge innovation; the ninth category is Chinese social sciences citation index, academic evaluation, quotation, journal evaluation, periodical,

academic influence; the tenth category is Informatics, Library science, hotspot and research frontier, core journals and Library information science. These are exactly the ten categories of field in which Informatics research methods are mostly used.

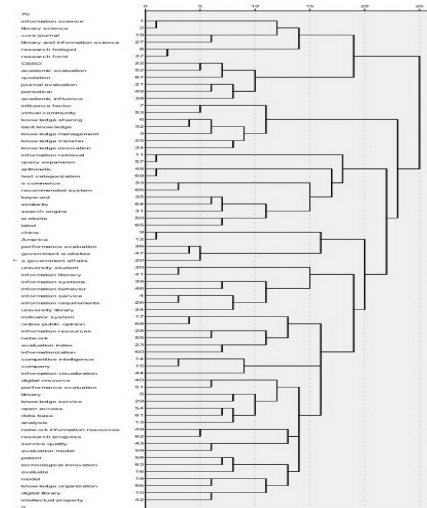


Figure 1. High-frequency words dendrogram.

TABLE I. STATISTICS OF HIGH-FREQUENCY KEYWORDS

Keywords	Frequency	Keywords	Frequency	Keywords	Frequency	Keywords	Frequency
Competitive Intelligence	143	Library and Information Science	58	Information System	32	Research Frontier	25
Digital Library	139	Knowledge Transfer	56	Core Journals	32	Knowledge Organization	24
Knowledge Management	135	Library Science	52	Periodical	31	Government Website	24
Informatics	112	Indicator System	50	Keyword	31	Academic Influence	23
Creative Commons	110	Information Visualization	48	Intellectual Property	30	Evaluation Model	23
Information Retrieval	107	Analysis	47	Network Information Resources	30	Quotation	22
Factors	92	Evaluation	46	Digital Resources	30	Recommendation System	22
library	90	Enterprise	44	Technology Innovation	29	Algorithm	22
Information Services	74	Internet Ppublic Opinion	43	USA	29	Virtual Community	21
Journal Evaluation	72	Search Engine	43	Performance Evaluation	29	Quality Of Service	20
E-commerce	72	Chinese Social Science Citation Index	43	Database	27	Academic Evaluation	20
Text Categorization	68	Information need	43	Informatization	27	Progress	19
Evaluation	65	Information Literacy	42	Knowledge Innovation	26	Similarity	19
E-government	65	China	41	Information Behavior	26	Performance Evaluation	19
Hotspot	64	Open Access	40	Web site	26	Query Expansion	19
model	62	Knowledge Services	37	label	26		
Information Resource	61	Patent	34	net	26		
University Library	60	Tacit Knowledge	33	University student	26		

D. Frequency analysis of the application of research method

In this paper, if the frequency of the research method is used more than 20, it is considered to be high-frequency method, then, we can get the conclusion that there is huge distance among the frequency of the application of research methods by accounting the method of high frequency, as is shown in Table 3. There are 1973 research methods applied in the sample paper, the mean frequency of each method applied is 5, however, the frequency of the application of the three method: experiments, empirical study and statistical analysis, are all more than 500 times which indicates that the application preferences of information science researchers are concentrated relatively.

Based on the statistical frequency of research methods, we can draw the conclusion that some research methods such as quantitative methods including empirical study and statistical analysis; qualitative methods including comparative analysis and comparative study; data analysis methods including survey and research; data analysis methods including cluster analysis and factor analysis; visualization, experiments and examples of verification proof represented by knowledge mapping are frequently used in Informatics research in the past five years.

The author specifically conducted a statistical analysis targeting at research tools in the sample paper and regarded the tools which were used more than 3 times as high-frequency research tools, the statistics table of high-frequency research tools drafted is shown in table 3.

It can be seen that the following tools are the research tools which are frequently used in the process of the application of Informatics research methods: the tool of factor analysis, cluster analysis and multidimensional scaling analysis—SPSS, the tool of visual knowledge map drawing—CiteSpace, Pajek and Netdraw; the tool of social network analysis—UCINET; the tool of empirical studies of structural equation—AMOS; the tool of simulation—Vensim, Netlogo and Matlab; commonly used tools—Excel, Bibexcel; the tool of network information search-- Web Spiders; the tool of web crawler and programming—JAVA.

TABLE II. STATISTICS OF RESEARCH TOOLS

Method	Number	Method	Number	Method	Number
PSS	116	Netdraw	11	Web Spiders	4
CiteSpace	97	JAVA	11	Web Crawler	4
UCINET	50	Bibexcel	9	WordNet	3
Pajek	21	Netlogo Platform	8	Wiki	3
AMOS	19	Vensim	7	VOSviewer	3
Protege	15	Lisrel	6	TDA	3
Excel	15	HistCite	6	EndNote	3
Matlab	14	Thesaurus	4	Altavista	3

The same research method as in the sample paper, the use of tools is also showing a lot of volatility: The utilization of

the well-known tools for example SPSS、CiteSpace etc. is ten times or even dozens of times than the utilization of general tools which shows the concentration of tool preferences of researchers. The high-frequency tool are mainly SPSS, CiteSpace and UCINET which indicated that data analysis methods, visualization and network analysis etc. are of high frequency to use in the past 5 years.

E. Correspondence analysis of the application of research methods and the field of research

“Correspondence analysis” also called “Correlation analysis” aimed to reveal the link between variables through cross summary table which is constituted by analyzing qualitative variables. Correspondence analysis is a kind of visual data analysis methods which can unfold several groups of data of no obvious link through intuitive positioning map. Factor analysis of common R-type and Q – type usually processed separately for objects and attributes so it is very difficult to research the intrinsic link between the properties of the sample by use of factor analysis while correspondence analysis can unify the two very well [5].

We counted separately the research method of each category based on ten types of research in Informatics and marked each research method of each research field by digital respectively, finally, we got a matrix of 3216*2. The two-dimensional correspondence analysis diagram was gained by conducting correspondence analysis through SPSS, as is shown in figure 2. In the figure, the red boxes represent for the field of research, the green circles represent the research methods used, the closer the distance is, the closer they are related.

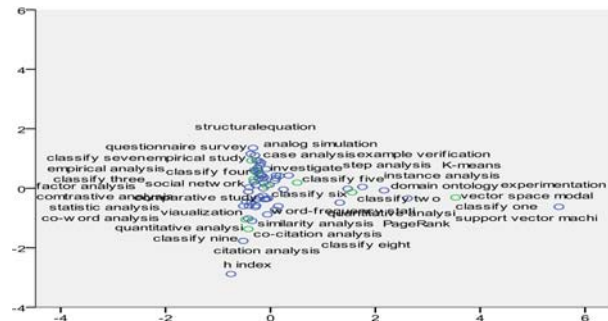


Figure 2. Two-dimensional correspondence analysis diagram.

It can be seen from the figure that the first field separated a lot from other fields, the research method is experiments and vector space model etc. which emphasized on data processing and results validation, it is more appropriate for research content in this type of research field; the main research methods in the second field are experiments, k-means, domain ontology and PageRank etc. these are also common methods information retrieval and information inquiry; the research methods of the third and the forth fields are empirical analysis and social networks and data mining etc.; the fifth field included the research methods of web survey, case study, empirical study, fuzzy comprehensive evaluation, data envelopment and research etc.; the research methods of the sixth field are patent analysis, analytic

hierarchy, examples of verification and link analysis etc.; the research methods of the seventh field are literature analysis, comparative analysis, factor analysis and content analysis etc.; the research methods of the eighth field are interview, impact factor, structural equation, survey, regression analysis, empirical analysis and simulation etc.; the research methods of the ninth field are citation analysis and co-citation analysis; the research methods of the tenth field are co-citation analysis, knowledge mapping and correlation analysis.

Among them, the application of research methods of many fields are clustering into two categories: one category included the statistical analysis methods such as frequency statistics, econometric analysis, quantitative analysis, principal component analysis, visualization, statistical analysis, co-word analysis, multidimensional scaling analysis, cluster analysis and content analysis; the other category included the methods of empirical research such as regression analysis, case study, empirical study, fuzzy comprehensive evaluation, data envelopment, research, web survey, literature research, SWOT analysis, empirical analysis.

III. CONCLUSION

A. Problems of research methods in Informatics

Based on the Bibliometrics statistical analysis of relevant data in sample papers, we found the following deficiencies and problems of applications of Chinese Informatics research methods.

1) methods of data collection are not standardized

The barbarism of methods of data collection in Informatics papers seriously affected the scientific of research result. Since there is great dependence between the scientific of Informatics papers and the standardization of the collection of data, we need to regulate the use of data collection methods in Informatics papers.

2) Lack of awareness of innovative methods and scientific logic

In sample papers, there are 1409 papers in which the authors have proposed new methods, accounting for 24% of all. Moreover, there are 815 papers which have not only mentioned new methods and new models but also proposed to verify the scientific, legitimacy and logic of the conclusion with experimental and empirical methods, accounting for 57.8% of the papers which have proposed new methods and new models and accounting for 13.9% of all the sample papers. Meanwhile, there are only 144 papers truly reflect the normalization and rigor of testimony from the papers which have mentioned supporting data, proof methods, proof process and results, accounting for 10.2% of the papers which have proposed new methods and new models and accounting for 2.5% of all the sample papers. This reflects the lack of innovation of research methods or the application of innovative approaches, and the lack of awareness of scientific logic of applying research methods.

3) Abuse and misuse of research methods concept

Since the application of measurement methods and new methods is becoming a trend in recent years, many researchers have used these methods without knowing the concept, the specific content, the deep-seated logic or the methodological paradigms of the methods in order to cater to this trend which led to the abuse and misuse of the concept and name of many methods.

4) Lack of collaborative applications of research methods

In sample papers, there are 1135 papers which simultaneously used three or more methods (This paper considered the papers which simultaneously used three or more methods as collaborative applications research methods) accounting for 19.3% of all the papers. However, after statistically analyzing the foreign sample papers of the same period, we found that there are more than half of the papers using various research methods in SSCI journals. So, the proportion of collaborative applications of research methods of Chinese Informatics is small and the ability is also poor.

B. Recommendations for improvement

For the above problems, the paper presents the following two proposals and measures wishing to improve the application of research methods of Informatics.

1) Strengthening the education of research methods and methodology

We can avoid many problems caused by the application of research methods through the implementation of education of research methodology. Therefore, the innovative research methods can be increased; the application of research methods are more consistent with scientific logic of consciousness; the process of data collecting will be more standard and scientific; the abuse and misuse of some methods concepts can also be avoided.

2) Encouraging researchers to do interdisciplinary research

Encouraging researchers of Informatics to do interdisciplinary research allows other disciplinary research approaches to be used more widely in Informatics which has the vital significance in advancing collaborative applications, diversity and pluralism of research methods and also in strengthening the innovation of research methods and the application of innovative methods of Informatics.

REFERENCES

- [1] Jun Qian, Xin Yang and Juan Yang, Cluster Analysis of Methods in Intelligence Research, Information Science, 2006, pp.1652-1653.
- [2] Huili Tan, Hotsopt Research Direction and Development of Information Science in China from 2003 to 2007, Modern Information, 2008, pp.22-25.
- [3] Donohue, J.C, Understanding Scientific Literature, A Bibliographic Approach, The MIT press, Cambridge, 1973.
- [4] Jiewang Chu and Chunxia Guo, General Principles of Co-words Analysis Method and Achieved by EXCEL, Information Science, 2011, pp.931-934.
- [5] Qiang Du and Liyan Jia, SPSS statistical analysis from entry to the master, The People's Posts and Telecommunications Press, 2009, pp.419-420.