Analysis Of Agricultural Logistics Literatures Based On Cnki

Yi LIN,Qilan ZHAO
School of Economics and Management, Beijing Jiaotong University, P.R.China,
17120582@bjtu.edu.cn

Abstract Since its introduction in 1980s, agricultural logistics has been developed for more than 30 years in China and has basically achieved localization. From the perspective of literature analysis,this paper applies the methods of literature statistics, word frequency analysis, co-term clustering analysis and social network analysis to 1981.(3The beginning of the study of agricultural logistics) CSSCI: a core journal in the field of agricultural logistics(Chinese Social Science Citation Index)，EI(Engineering Index),SCI (Scientific Citation Index)Literature statistics, the establishment of a total of 1331 papers database as the data source of this study, and its bibliometric laws and characteristics of analysis. The knowledge map of co-occurrence matrix was established by word frequency analysis, and the top 20 high-frequency keywords were classified by cluster analysis, and the hot fields of agricultural logistics research were summarized.
Key words Agricultural logistics; China Network of knowledge; Word frequency analysis; Cluster analysis; Development trends

1 Introduction

Bibliometrics is based on the literature system and document metrology features, using mathematical, statistical and other quantitative research methods to study the distribution of literature and information structure[1], the relationship between quantitative changes and quantitative management, and then explore some of the structure, characteristics and laws of a discipline of science and technology. Compared with the agricultural logistics organizations in developed countries, the research work in the field of agricultural logistics in China started late and did not form a mature system. The level of agricultural industrialization is still at its infancy, and the system research literature on agricultural logistics organization is also relatively scarce[2]. In order to make further judgment on the future research direction in the field of agricultural logistics and clarify the problems of agricultural logistics in practical application in recent years, various types of journal articles in CNKI were consulted and sorted out.

2 Collected and collated academic papers

These papers mainly come from the cnki.net. According to the search term “agricultural logistics”, the core journals, CSSCI, EI, and SCI journal articles were classified and retrieved in the CNKI database. Combining the number of valid samples collected for the four major categories of papers, a total of 1,331 papers related to agricultural logistics were obtained. These papers are classified according to the content and the paper information were summarized.After a large number of statistics to form a database, as the data source for this study.

In the research process of this article, quantitative research has always been carried out around this database. Various types of periodical articles in the field of agricultural logistics have been consulted since 1981, and the various types of journal inclusion was explored. According to relation chart as Figure 1, excluding EI papers from core papers and CSSCI papers,and the core / CSSCI overlap is only counted once and attributed to CSSCI journals.

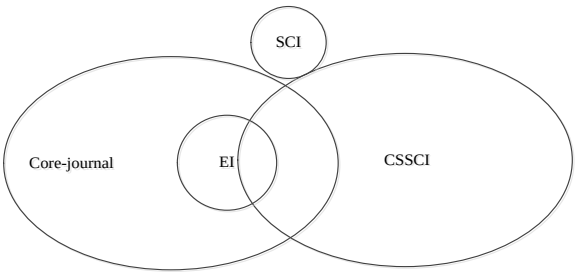

Figure 1 The Inclusive Relationship among Various journal articles

Copyright © 2019, the Authors. Published by Atlantis Press.
This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

3 Keywords analysis and classification of papers

Taking 1331 selected documents as samples, they are exported as EndNote in CNKI and saved as text files. Using SATI, a statistical analysis tool for bibliographic information of bibliographies developed by Liu Qiyuan and Ye Ying[3], the key words of these 1331 articles were extracted, and the number of effective keywords finally obtained was 4080,and these keywords are sorted according to the frequency of occurrences. This section keywords analysis process shown in Figure 2.

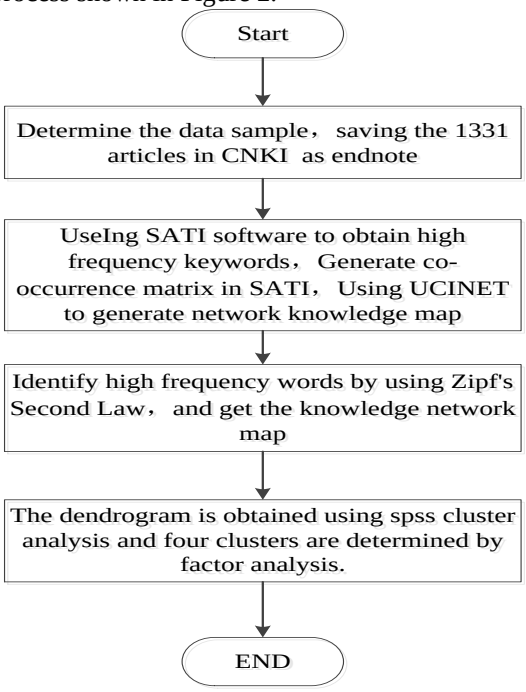

Figure 2 Keyword analysis process

According to the process summarized in Fig.3, using SATI to extract the key words of 1331 papers, and count the frequency.SATI arranges the keywords from highest to lowest according to the frequency of appearance, and generates the co-occurrence matrix (100 * 100) of the top 100 appearances of the keywords. The frequency of descending order of the corresponding number of entries in the table can be used as a unit of knowledge elements to construct knowledge unit co-occurrence relationship matrix. In order to eliminate the influence of frequency disparity on the statistical results in the multivalued co-occurrence matrix, the software adopts the Equivalence coefficient (formula (1)) to transform the multivalued matrix into a similarity matrix of [0,1]. T_i and T_j indicate the number of times the entry appears.The numbers in the similarity matrix represent the similarities between the matrix elements. The greater the value, the greater the degree of association. In addition, due to too many 0 values in the similarity matrix, the error in statistics may be easily caused. On the basis of this, the software automatically generates a dissimilarity matrix, that is, the sum of similar matrix values and -1 is the value of the dissimilar matrix element[4].

$$E_{ij} = \frac{F_{ij}^2}{F_i * F_j} \tag{1}$$

In the formula, $E_{ij}$is the value of the similarity matrix element;
F_{ij}Is the co-occurrence times between T_i and T_j;
$F_i$is the total frequency of occurrence of T_i;
$F_j$is the total frequency of occurrence of T_j.

Import matrix data into social network analysis tool UCINET to generate ##h and ##d files, Then import the ##h format file generated by UCINET in NetDraw, perform visual data analysis, and generate the co-

occurrence matrix knowledge map as the Figure 3.

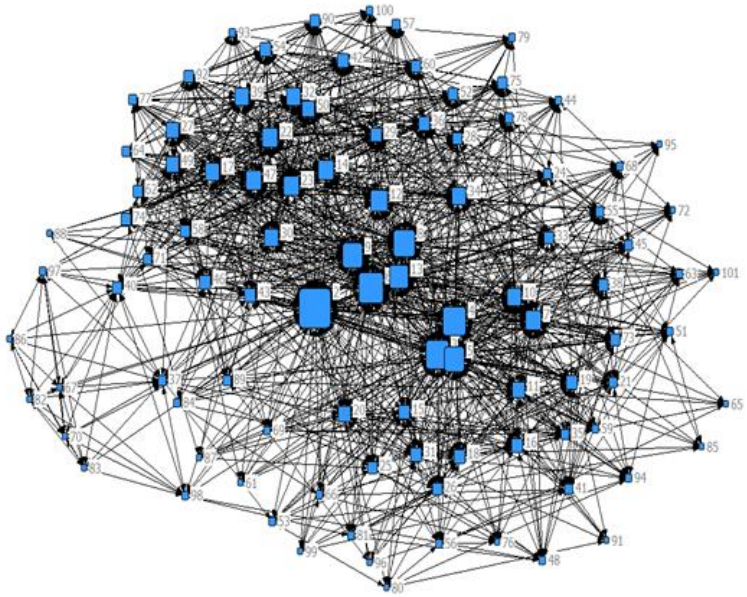

Figure 3 The knowledge map by searching all literature keywords Co-occurrence matrix

The node size in the figure reflects the word frequency and centrality of the keyword they are represented by the node. Connections between nodes reflect co-occurrence relationships between keywords. Serial numbers represent different keyword codes. Point center degree is the focus of social network analysis, and it is also the main indicator used to evaluate the status of keywords in social networks. It is used to measure the degree of direct connection between a keyword and many keywords in a social network. The higher the node's degree of centrality, there is more closely the relationship among the nodes, and it has a more central position in the network.

Due to the large number of keywords involved in the literature, there is a need to further narrow the scope of the keyword study. So we choose the Zipf's second law to get the threshold n of high frequency words. Zipf's second law is a very important law in bibliometrics[5]. The predecessors have proposed the critical value calculation formula of high-frequency words and low-frequency words according to Zipf's second law,which is,

$$n = \frac{1}{2}\left(-1+\sqrt{1+8I_1}\right) \tag{2}$$

I_1 is the number of keywords that appear once, the number of the keywords appear only once that are obtained by SATI's field extraction is 903, calculating according to formula (2), the threshold n=42. According to the above figure, there are 20 keywords that meet the criteria. The frequency of occurrence and the sorting of the point degree center are shown in Table 1.

Table 1 High-Frequency Keywords

Number	Key words	Frequency	Point degree center
1	Agricultural products	442	2.071
2	Agricultural production logistics	301	1.869
3	Logistics	246	1.711
4	Countermeasures	197	1.608
5	Agricultural logistics	155	1.598
6	E-commerce	133	1.494
7	Rural logistics	107	1.410
8	Cold Chain Logistics	89	1.397
9	Question	65	1.382
10	Modern agriculture	64	1.320
11	Supply chain	63	1.280

Table 1, cont.

12	Development	60	1.253
13	Agricultural Industrialization	58	1.225
14	Existing condition	52	1.169
15	Modern logistics	49	0.922
16	SWOT analysis	47	0.900
17	The Third Party Logistics	47	0.796
18	Agriculture	47	0.784
19	Development Strategy	45	0.747
20	Internet of Things	42	0.713

The frequency of keyword occurrence is positively correlated with the value of centrality，and the "agricultural product" is the central point of 20 high frequency keywords. "Cold chain logistics," "third-party logistics," "development," "countermeasures," and "e-commerce," appear at higher frequency.it constructs that the study on agricultural product sales, problem status and coping strategies is relatively common.

4 Keywords Analysis

In the research of agricultural logistics, the number of key groups that should be divided cannot be clearly determined. For this reason, factor analysis can be used to determine the optimal number of classifications. Therefore, the correlation matrix of the above 20 keywords was introduced into SPSS23 for factor analysis. The KMO sampling fitness score was 0.835, which was close to 1, and the Bartlett test sphericity test significance was 0.002<0.05. As can be seen from Table 2, the common factor variance extraction results, the extraction value is close to 1, the variable common degree is higher, so it is suitable for factor analysis.

Table 2 Common Factor Variance

Key words	Initial	Extract	Key words	Initial	Extract
Agricultural products	1.000	0.636	Supply Chain	1.000	0.996
Agricultural logistics	1.000	0.731	The wholesale market	1.000	0.847
Agricultural production logistics	1.000	0.995	The Third Party Logistics	1.000	0.979
Logistics	1.000	0.657	Question	1.000	0.660
Rural logistics	1.000	0.675	Modern Agriculture	1.000	0.997
Countermeasures	1.000	0.663	Logistics facilities	1.000	0.497
Cold Chain Logistics	1.000	0.985	The Internet of things	1.000	0.997
E-commerce	1.000	0.997	Agricultural modernization	1.000	0.756
Agricultural Industrialization	1.000	0.975	Development	1.000	0.993
Agricultural	1.000	0.996	Agricultural market	1.000	0.605

According to the principle that “a factor with a eigenvalue larger than 1 is retained and the eigenvalue is less than 1” is discarded[6], the eigenvalues of the first 4 factors are greater than 1 and the cumulative variance contribution rate is 68.184%. Therefore, according to the result of factor analysis, the top 20 keywords in this article are divided into 4 categories.

Table 3 Interpretation Of Total Variance

Initial eigenvalue			Extract the sum of squares of the load			the sum of squares of rotating load		
Sum	***Variance %***	***Cumulative %***	***Sum***	***Variance %***	***Cumulative %***	***Sum***	***Variance %***	***Cumulative %***
1.443	7.215	22.191	1.443	7.215	22.191	1.428	7.138	10.194
1.394	6.972	45.132	1.394	6.972	45.132	1.376	6.882	39.132
1.184	5.920	54.178	1.184	5.920	54.178	1.145	5.725	48.178
1.112	5.558	68.184	1.112	5.558	68.184	1.061	5.307	68.184
.983	5.426	74.683						
.940	5.313	79.998						
.720	5.286	83.183						

Table 3, cont.

.674	5.283	85.279							
.546	5.258	87.982							
.485	5.246	88.100							
.363	5.240	92.801							
.257	5.218	96.399							
.157	5.199	96.884							
.098	5.045	97.239							
.049	5.006	97.335							
.048	4.915	98.423							
.044	4.701	98.734							
.040	3.599	99.207							
.009	3.372	99.771							
.001	.229	100.000							

In this paper, SPSS 23 is used to cluster the correlation matrix composed of 20 high frequency keywords processed by equivalence coefficients in SATI software. Using the clustering method of "connection between groups", the interval is "Euclidean distance", select "from 0-1" for standardization range and classified by variables.Obtain the following chromatographic cluster analysis tree, as shown in Figure 4.

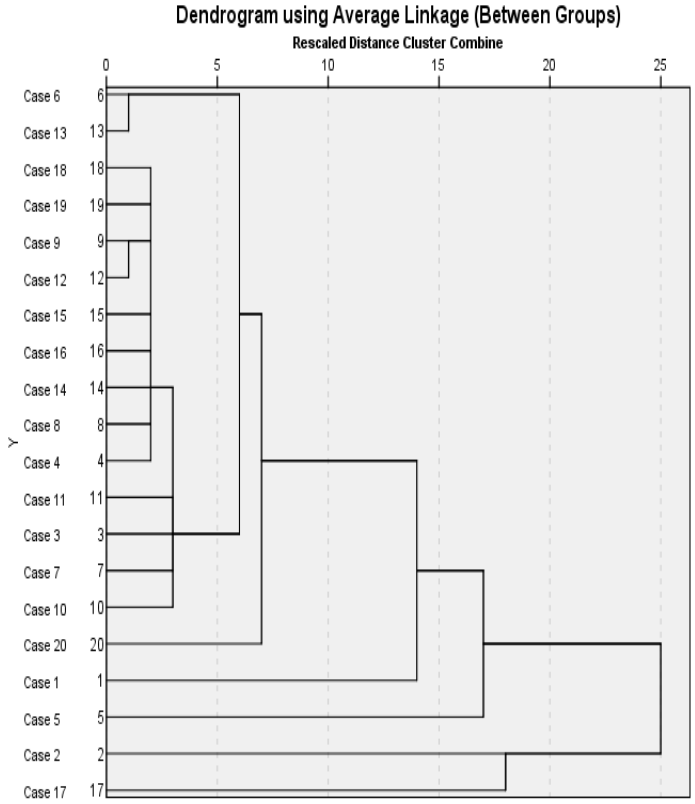

Figure 4 Co-word clustering tree

Summarize the following research hotspots in the field of agricultural logistics:

（1）Modern Agriculture and Supply Chain Logistics.

The keywords of this type of research are "logistics," "agricultural products," and "subjects of logistics," which are the areas of research that combine theory with practice.

（2）Problems and Countermeasures in Rural Logistics.

logistics mainly starts with the aspects of logistics and transportation time, storage styles, facilities and

equipment, talents absorption conditions and cost levels.

（3）Development and Application of High-tech in

Modern Agriculture. Key words of such research include "Internet of Things," "e-commerce," and "cold-chain logistics," and it are closely related to the sales of agricultural products. Along with the rise and development of agricultural logistics hardware facilities, wholesale markets and cold-chain trucks are rapidly developing.

（4）Promote the industrialization of agricultural.

products logistics industry. In this type of research, the explanation of the current status of China's agricultural industrialization and the measures for accelerating the industrialization of agriculture are more extensive.

5 Conclusion

(1)After consulting more academic papers related to “bibliometrics” and conducting a comprehensive comparison, this paper selected co-word clustering and social network analysis methods to conduct literature research in the field of agricultural logistics, and the results prove that the selected method Can better solve this problem.

(2)The clustering analysis of key words in the field of agricultural logistics has yielded the top 20 high frequency words, summing up the four major research hotspots of China's agricultural logistics.

(3)Increase the intensity in scientific research on agricultural logistics, increase the number and level of documents.

References

[1]Boccolini, Patricia de Moraes Mello; Boccolini, Cristiano Siqueira; Chrisman, Juliana de Rezende.NonHodgkin lymphoma among Brazilian agricultural workers: A death certificate case-control study.Archives of environmental & occupational health .2017-May-04 (Epub 2016 Apr 20),72 ,166-172

[2]Jackson, Sarah S; St George, Diane Marie; Loffredo, Christopher A.Nonoccupational exposure to agricultural work and risk of urinary bladder cancer among Egyptian women.Archives of environmental & occupational health. 2017-May-04 (Epub 2016 Mar 30),72,166-172

[3]Zhang, Defu; Cai, Sifan; Ye, Furong.A hybrid algorithm for a vehicle routing problem with realistic constraints.INFORMATION SCIENCES.JUL 2017,394 :167-182.

[4]Gu, Jiangxin; Yuan, Mengxuan; Liu, Jixuan.Trade-off between soil organic carbon sequestration and nitrous oxide emissions from winter wheat-summer maize rotations: Implications of a 25-year fertilization experiment in Northwestern China.The Science of the total environment,2017-Apr-05 (Epub 2017 Apr 05),595, 371-379

[5]Liu, Y.; Tang, K.; Long, J.The association between hepatitis B self-awareness and depression: Exploring the modifying effects of socio-economic factors.JOURNAL OF VIRAL HEPATITIS,APR 2017,24,330-336

[6] Porat, Ron; Lichter, Amnon; Terry, Leon A. Postharvest losses of fruit and vegetables during retail and in consumers' homes: Quantifications, causes, and means of prevention. POSTHARVEST BIOLOGY AND TECHNOLOGY. 135-149 MAY 2018.