International Journal of Smart Business and Technology
Vol.6, No.1 (2018), pp.1-14
http://dx.doi.org/10.21742/ijsbt.2018.6.1.01

Research on Supply Chain Application based on Big Data

Antonio Risi[1] and Pietro Schipani[2]

[1,2]University of Turin, Italy
[1]antonio.j.risi@gmail.com

Abstract

With the advent of the information age, the application of big data in supply chain management between different industries has become more and more extensive. The methods and ways to obtain big data and various data types determine the type of big data and whether it can obtain technology for enterprises. The competitive advantage in the market promotes the development of various industries. This article is based on the basic concepts of the big data supply chain and the research of related theoretical methods by experts and scholars. Summarizes the current application status of big data in supply chain management, and provides certain support for future in-depth exploration. Quantitatively analyze the keywords of 601 English documents on the topic of "big data supply chain" obtained by crawlers. And based on scientific research methods, the quantitative data transformed by the text content is sorted, summarized, and analyzed. Then according to the summary content, the current research hotspots of the big data supply chain are compared and analyzed, and the future development direction of big data application in the supply chain is predicted.

Keywords: *Big data, Supply chain, Cluster analysis, Social network analysis*

1. Introduction

The concept of big data has brought about a brand-new change in supply chain management. The way of competition in the new era is completely different from the previous way. If today's competition between enterprises is a competition between business models and a supply chain, the emergence of big data makes this competition a digital rivalry.

Nowadays, the abundance of data sources, quantity and types has promoted the wide application of big data technology in many fields. Big data technology has improved the real-time and comprehensiveness of corporate business behavior and the rapid development of market analysis. Supply chain management is to improve the overall efficiency of the enterprise and market competitiveness. By adopting an efficient management model, we can effectively manage the entire supply chain from upstream suppliers to downstream customers. In recent years, due to the development of advanced technologies such as the Internet, cloud technology, and the Internet of Things, the application of big data has become more extensive, and its effective value in the market has continued to be demonstrated.

The advent of the data age has prompted the reshuffle of different industries, and the business model and structure of enterprises have also undergone fundamental changes. The current big data technology is affecting the traditional supply chain towards the integration of big data. Through big data technology, it conducts in-depth analysis and control of a series of logistics links in the supply chain, such as procurement, design, and inventory, and predicts demand

Article history:
Received (November 12, 2017), Review Result (December 21, 2017), Accepted (February 3, 2018)

Print: 2288-8969, eISSN: 2207-516X, IJSBT

based on big data analysis, optimizes supply and distribution networks, and promotes the transformation of big data capabilities into corporate core competitiveness.

In the process of enterprise supply chain management and operation, big data technology can provide rich data information covering upstream suppliers, downstream customers, product quality, and market dynamics to ensure the rationality of relevant strategic decisions and provide directions for enterprise development. In addition, the application of big data to supply chain management can also assist management in making decisions while providing accurate services to customers [5]. Big data analysis has now become an important research method for the supply chain management. Based on meeting the basic requirements of supply chain management operations, supply chain management is combined with big data analysis. Not only can we obtain effective data in time, but also optimize the supply chain management operation mode while accurately positioning, and improve corporate efficiency.

Under normal circumstances, when new business models first appear in the market environment, supply chain management content may become complicated and numerous, leading to imbalances in supply and demand in the supply chain, making it impossible for companies to predict demand and supply normally. And the inability to integrate resources has become a bottleneck in the development of the enterprise supply chain. In the process of information upward, due to some delays, normal procurement and inventory planning cannot be made. The increase in procurement, production, and inventory costs harms the self-interest of enterprises, which also puts forward higher requirements for the integration of information resources. To solve these problems, make proper use of big data analysis, business intelligence, supply chain management, and other theories and technologies, and use more insightful data to conduct in-depth analysis and adjustments to each organizational link of the supply chain, grasp the characteristics of key business, and discover Problems and improve processes, coordinate supply chain relationships, improve supply chain operating efficiency, and realize changes from simple management to efficient and intelligent management. Research on the application of big data in supply chain management requires not only the use of effective resource data, but also the visual analysis of existing resources based on the existing theoretical methods, and the application of the application to the development process of the enterprise based on the actual needs of the enterprise. The applied research on the big data supply chain is not only a summary of the theoretical review but also a clear direction for future research through analysis and prospects.

Since the 1970s, scholars have begun to use co-word analysis to discover the system structure of related theories. Scholars [1] used the co-word analysis method to study the changes in the aquaculture industry over the years, compared similar maps and tolerance maps in different years, and used different index words for comparative analysis, and obtained the development trend of the aquaculture industry and the changing trend of the theme. Experts such as Law [2] combined theory with reality, studied the actual role of literature in the development of science and technology and studied the scientific driving force by summing up the principles of co-word analysis, which has a profound application in the future use of predetermined template structure. Impact. Corley et al. [3] combined bibliometrics and co-word analysis methods, and used software to visualize logical relationships into graphs, providing a successful model for subsequent research; scholars such as Neff and Corley [4] used co-word clustering methods Applied to the field of artificial intelligence, it has further improved the research on the use of computer software to achieve visual display of logical relationships and development trends. Scholars such as Khasseh and Soheili [5] evaluated and analyzed the advantages and disadvantages of co-word analysis based on existing theories, and successfully verified some classic hypotheses in co-word analysis research.

The American sociologist Granovetter [6] put forward the "weak connection power hypothesis" in 1975, combining social network analysis with the four dimensions of emotional strength, interaction frequency, intimacy, and reciprocal exchange between groups to study the human relationship network. Gould R V [7] not only verified the interrelationship between human activities and social networks with the help of mathematical models but also proposed that the interaction between people is closely related to the location and complexity of the social network structure. Based on social network analysis, Jocob Moreno [8] put forward the concept of "community diagram", which vividly exposes the complex relationship and structure of human society, and this research method has been widely used by experts and scholars in many research fields. Otte and Rousseau [9] systematically introduced the concept, development, and application of social network analysis in the field of analysis. This research has become an important reference material for scholars to learn early social network analysis methods for intelligence analysis. Scholars such as Lerman [10] used the social network method to explore the emotional power and reciprocal exchange between human families, and the logical family relationship was transformed into a graph that can be directly displayed. Ghosh [11] and others put forward the view that the social network analysis method is random due to the interaction between nodes in the social network. Therefore, the interactive nature between nodes must be considered in the social network analysis algorithm, otherwise, this randomness will not be conducive to the research of epidemics and information diffusion. Experts such as Andrade RL [12] used the social network analysis method to explore the factors that affect the relationship between animal types and constructed a method to compare the attributes of the research object with edge weighting and unweighted processing. In addition, foreign scholars have integrated social network analysis into economics, psychology, and other research fields, which have derived social capital, resource theory, and "embedding" theories [13].

2. Theoretical basis and data sources

2.1. Theoretical basis and related concepts

Co-word analysis is a method used to reflect the strength of association between keywords. First, the research topic words are divided, and the common word matrix is constructed by clustering analysis based on the statistics of the frequency of a group of words in the same document. Draw a visual map, systematically display the intuitive quantitative information of the research object, and finally analyze the hot spots of the research object. The calculation formula for co-word analysis is as follows:

$$C_N(k_i) = \sum_{J=1}^{N} P_J(k_i), P_J(k_i) = \begin{cases} 0 & k_i \notin P_j \\ 1 & k_i \notin P_j \end{cases} \quad (1)$$

$$C_N(k_q{:}\,k_r) = \sum_{J=1}^{N} P_J\left(k_q{:}\,k_r\right), P_J(k_q{:}\,k_r) = \begin{cases} 0 & \left(k_q{:}\,k_r\right) \notin P_j \\ 1 & \left(k_q{:}\,k_r\right) \notin P_j \end{cases} \quad (2)$$

$C_N(k_i)$ represents the cumulative frequency of keyword k_i in the research sample, and N represents the total number of samples. If the keyword k_i belongs to the document P_J, the value of $P_J(k_i)$ is set to 1, otherwise, it is set to 0. $C_N(k_q{:}\,k_r)$ represents the cumulative frequency of two-dimensional co-words in the research sample of $\left(k_q{:}\,k_r\right)$. If the keywords k_q and k_r belong to the document P_J, the value of the two-dimensional common word is set to 1, otherwise, it is set to 0. Multidimensional co-words can be deduced by analogy.

Co-word clustering analysis is the main type of co-word analysis, which clusters keywords that are relatively close together. Through accumulation, an independent cluster with low inter-group similarity and high intra-group similarity is formed, which plays an important role in discovering close connections between keywords and mining the hidden information of research objects.

Multi-dimensional scale analysis is a method of dimensionality reduction that uses the similarity of research objects with multi-dimensional features to classify statistics. Through multi-dimensional scale analysis to classify samples and visualize applications, it is possible to obtain clear relationships between variables and potential influencing factors of the network.

The multi-dimensional scale analysis method can use space and distance to determine the position between the points, and then divide the research object into a multi-dimensional computing space. In 1944, Guttman [14] proposed the scale analysis method for the first time, and then through the improvement and development of Marshalek, Lingoes [15], and others, successfully transferred the intimacy of the variable to the Euclidean space of the internal correlation matrix of the variable [16].

Generally, vector coordinates are used to calculate the distance in two-dimensional Euclidean space. The vector coordinates of the keyword Q_1 are set to $Q_1 = (X_1, X_2)$, and the vector coordinates of the keyword Q_2 are $Q_2 = (Y_1, Y_2)$, the values of Q_1 and Q_2 The calculation formula of Euclidean distance is:

$$L = sqrt((X_1 - Y_1)^2 + (X_2 - Y_2)^2) \tag{3}$$

In the same way, the calculation formula of 3D Euclidean distance is

$$L = sqrt((x - x_2)^2 + (y_1 - y_2)^2 (z_1 - z_2)^2) \tag{4}$$

The Euclidean distance calculation formula for n-dimensional space is:

$$L = sint\big(\sum_{n=1}(x_{11} - x_{12})^2\big), i = 1,2, \cdots n \tag{5}$$

Social network analysis is a quantitative analysis method for sociologists to study social structure based on mathematical methods and graph theory. It is not a specific subject, most of the time it is applied as a research method, and sometimes it becomes a research perspective. Social network analysis is closer to the reality of society and considers the interrelationship and role of individuals. With the help of software such as Ucinet for social network analysis, it is possible to more intuitively discover the individual's position in the social network and the overall structure to obtain the mutual relationship and structural information of the social network. At the same time, it is convenient to discover the action characteristics and mutual social relations of different participants in the network.

We call several nodes representing events or individuals in the network as participants and call the links connecting these nodes as relationships. The relationship symbolizes the social relationship established by the participants through the activity. In social network analysis, there are two commonly used methods to visualize this social relationship: matrix algebra and community diagrams. In the community graph, all participants form a point set: $N = (n_1, n_2, \ldots n_g)$, which can be divided into directed graphs and undirected graphs according to the relationship direction.

2.2. Data sources

The data source of English documents is limited to the Web of Science (WOS) database, the time is set from 2010 to 2019, and the subject is limited to "Big Data Supply chain". After retrieval, 601 English documents that meet the requirements are obtained.

In the 21st century, data is one of the most valuable products of the Internet. If the Internet is a huge spider web, then the data generated by the computer will become food after the spider web, and the crawler script is like a small spider, Catch the prey you want to get. To explain the image, the user obtains the data by sending a request through the browser, then the browser downloads the code, and the CSS parser renders the code into a page. The data crawler simulates the browser to submit a request, obtains the webpage code, and stores the required data in the database. The general flow is shown in [Figure 1].

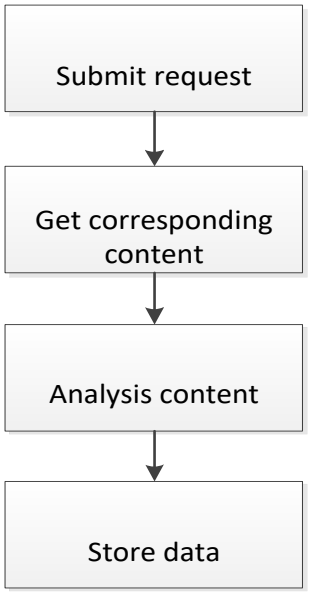

Figure 1. Crawler process

The specific target process is as follows. First, run Python to write an HTTP Request, and send the request to CNKI and Web of science respectively. After the request is sent successfully, the response status code of "200" is received. At this point, the returned Response content is processed, the HTML-related content is found in the obtained content with the help of the Python tool library, the required information is disassembled and analyzed, and the relevant data is saved through the PyMySQL library.

3. Word frequency analysis

3.1. Literature sample statistics

This article selects journals as the database in the crawling process. The reason is that journals are documents that are generally valued in the academic field and can represent the level of disciplines. Therefore, after summarizing the data, this article counts the ten journals with the most publications on related topics, and visually displays the statistical results with the help of a fan chart, as shown in [Figure 2].

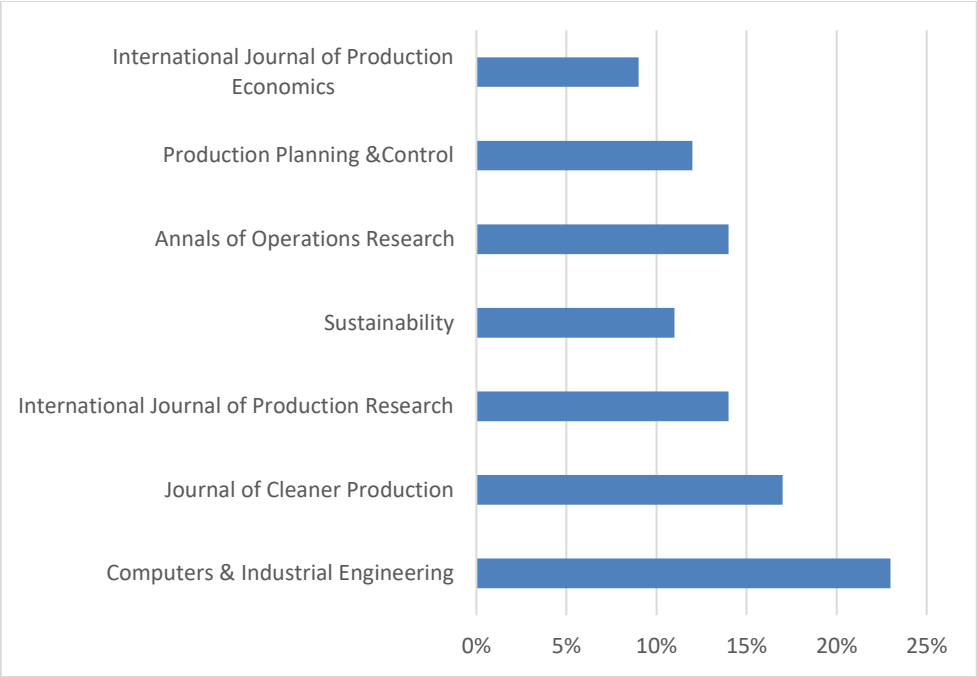

Figure 2. The main source journals of the web of science sample documents

It can be seen that the relevant papers of the big data supply chain are roughly concentrated in the fields of logistics, economics, and production, and the journals published by the relevant papers have a certain influence in their respective disciplines, which shows that the big data supply chain has gradually become very important Research topics.

3.2. Data preprocessing

Considering that there may be semantic similarities between keywords, to facilitate subsequent analysis of English keywords. This article processes and merges the English keywords of the obtained literature, and removes some keywords that are not related to the subject. With the help of Mysql to filter and delete, a total of 2862 English keywords are finally counted.

In the process of merging keywords and statistics in the previous section, it was found that the frequency counts of words with the same meaning of Italian and English keywords are quite different, to eliminate the influence of different bases. This paper adopts the zero-mean standardization method to standardize word frequency statistics. The specific steps of the Z-Score standardization method are as follows: 1) Calculate the average of all variables μ; 2) Calculate the standard deviation σ of all variables; 3) Subtract μ from the sample valueχ, and then divide by σ to get the standardized Z value, the formula is as follows

$$z = \frac{x-u}{o} \tag{6}$$

The word frequency after the above-mentioned Z-Score standardization process conforms to the standard normal distribution, and the size of the Z value can objectively reflect the popularity of the keywords and is positively correlated. After calculation, we selected English keywords with a statistical frequency greater than or equal to 5 to create a summary table of high-frequency vocabularies for big data supply chain research, as shown in [Table 1].

Table 1. High-frequency and Z value list of English keywords

Keywords	Word frequency	Z value	Keywords	Word frequency	Z value	Keywords	Word frequency	Z value
Big data	291	47.87	Simulation	14	1.89	Privacy	9	0.95
Supply chain	133	19.01	Information-technology	14	1.89	Food safety	8	0.74
Supply chain management	122	17.41	Operations management	13	1.56	ERP	8	0.74
Industry 4.0	73	10.32	Data mining	13	1.56	Circular economy	8	0.74
Big data analytics	70	9.89	Data quality	13	1.56	Application	8	0.74
Logistics	46	6.42	Cyber-physical systems	13	1.56	Value chain	7	0.58
E-commerce	41	5.69	Business intelligence	13	1.56	Sustainable	7	0.58
Supply chain finance	40	5.55	Blockchain	13	1.56	Supply chain risk management	7	0.58
Optimization	32	4.39	Machine learning	12	1.45	Sentiment analysis	7	0.58
Cloud computing	26	3.52	Business analytics	12	1.45	Research agenda	7	0.58
Internet of things	23	3.24	Io T	11	1.32	Sustainability	7	0.58
Digital manufacturing	20	2.97	Competitive advantage	11	1.32	Environmental Sustainability	6	0.45
Supply chains	17	2.51	RFID	9	1.01	AHP	6	0.45
Transportation	16	2.32	Mathematics	9	1.01	Risk management	6	0.45

3.3. Total word frequency analysis

The above is a sample statistic of the literature obtained by the crawler, and also incorporates keywords and standardizes the processing. This section summarizes and analyzes the common research hotspots and unique research hotspots of related courses by comparing the high-frequency English words used in big data supply chain research.

From the summary table of English high-frequency keywords, it can be seen that there are 8 keywords in the research of big data supply chain. Research hotspots not only cover major subject keywords such as "big data", "big data analysis", "supply chain management", and

"application". It also involves emerging fields such as "supply chain finance", "e-commerce", and "risk management". It can be judged that professional scholars at home and abroad are constantly expanding their research directions while in-depth research on the main subject.

English high-frequency words mainly cover the following research fields: 1) Big data. Big data and big data analysis have now become the most promising research fields in the new era, and they are also one of the key directions for scholars in future research topics. 2) Supply chain. Scholars are also concerned about supply chain management and supply chain finance. 3) Industry 4.0, the advent of the industry 4.0 era has realized the rapid development of the supply chain. Relying on big data technology not only promotes the intelligent application of the supply chain but also brings many opportunities for the development of the group transaction mechanism. 4) E-commerce. Only in line with the general background of e-commerce development in the 21st century, big data technology can truly enhance the business effect of enterprises and provide a complete decision-making plan for the supply chain management.

3.4. Annual keyword frequency analysis

To further explore the research hotspots of big data supply chains at home and abroad, this article counts the annual frequency of high-frequency English keywords from 2010 to 2019. The specific results are shown in [Table 2].

Table 2. Annual high-frequency keyword in WOS (part)

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Big Data	8	15	20	26	30	40	35	38	41	35
Big data analytics	2	1	3	5	10	9	16	8	6	14
Supply chain	7	8	14	10	11	11	16	18	20	21
Supply chain management	5	5	11	15	11	14	10	17	19	23
Supply chain finance	2	1	2	3	3	5	1	6	8	4
Risk management	0	0	1	0	0	2	4	2	2	0
E-commerce	0	0	1	2	3	5	7	8	9	10
Application	0	0	1	2	1	0	1	1	0	1

From the above table, it can be concluded that “Big Data” and “Supply chain finance” in English high-frequency keywords have fluctuated in the annual keyword statistics in the past ten years, but this is a topic. The core of research, on the whole, is on the rise. “Supply chain”, “Supply chain management” and “E-commerce” are the hottest research topics and represent the development direction of big data supply chain research. The annual word frequency statistics of “Risk management” and “Application” tend to be stable. It means that the relevant research direction is inseparable from the subject, and it is constantly being explored.

4. Keyword hotspot and trend analysis

4.1. Construction and standardization of common word matrix

The co-word matrix is a matrix to calculate classification information based on common keywords or numbers. Then through the arrangement of the common word matrix, we can see the correlation between the data and the logical calculation method.

The specific operation sequence for constructing the common word matrix is as follows: 1) Complete data statistics and construct a distribution table of high-frequency words and keywords. 2) Use the tool to filter out the single keywords that have no common words and then delete them. 3) Combine the processed keywords in pairs. 4) After the combination is completed, calculate the number of combinations according to the formula, and perform simple cross-copying of the two sets of data. 5) Finally, use the pivot table for layout and successfully construct a common word matrix.

$$\begin{bmatrix} n \\ k \end{bmatrix} = \frac{P_{n,k}}{k!} = \frac{n!}{k!(n-k)!} \tag{7}$$

The initial co-word matrix is shown in [Table 3].

Table 3. Co-word matrix of high-frequency keywords in WOS (part)

	Big data	Supply chain	Supply chain management	Industry 4.0	Big data analytics	Logistics
Big data	280	70	65	48	33	20
Supply chain	87	145	32	29	15	13
Supply chain management	54	24	132	15	6	6
Industry 4.0	48	27	16	79	4	4
Big data Analytics	27	14	6	5	78	2
Logistics	17	12	8	3	1	49

This section will standardize the above two tables and use Ochiia coefficients to construct a dissimilarity matrix. First, enter the corresponding high-frequency keyword frequency on the diagonal of the form, and enter 0 in the rest of the form. Then use the Ochiia coefficient to complete the standardized operation of the co-word matrix. Then subtract 1 from each item in the standardized matrix, and finally successfully construct a dissimilarity matrix.

4.2. Cluster analysis of keywords research topics

In this paper, the cluster analysis of high-frequency keywords is performed. The specific operation is to systematically cluster the English high-frequency word dissimilarity matrix obtained in the previous section by running the SPSS software to obtain the corresponding clustering pedigree diagram. The clustering pedigree graph is essentially a dendrogram using average connections, with all cluster categories listed on the far left, and the numbers on the horizontal axis reflect the relative distance between each category. The horizontal lines continue to merge as they go to the right, and the categories are highly concentrated, finally forming a figure that looks like a growing tree horizontally.

This paper introduces the cluster adhesion force for the next analysis. Adhesion can be used to measure the contribution degree of each keyword in the group to the group. The calculation of the adhesion force generally uses the average value of the co-occurrence frequency of the keyword and other keywords in the same document. The larger the average value, the more important the position of this keyword in the cluster. Usually, we use the maximum adhesion force. The keyword is called the headword. When there are clusters of n keywords; keyword A_i $(i \leq n)$ for other keywords B_j, the adhesive force $N(A_j)$ of A_j is:

$$N(A_1) = \frac{1}{n-1} \times \sum_{j=1}^{n \neq i} F(A_i - B_j) \quad (8)$$

After calculation, the adhesive force of English high-frequency keywords is shown in brackets in the table below.

4.3. Multidimensional scale analysis

Cluster analysis is performed on English high-frequency keywords and the concept of cluster center is introduced. To further study and analyze the interaction between keywords within clusters, this paper uses multi-dimensional scale analysis to study the relationship between keywords. And with the help of SPSS software to build a multi-dimensional analysis map, the key research areas of the big data supply chain are obtained.

Multi-dimensional scale analysis is an exploratory data analysis technique that can not only express actual distance but also display data for subjective similarity judgments. By comparing the results of multi-dimensional scale analysis with cluster analysis, it is found that the two methods have similarities in the classification of clusters, and the multi-dimensional scale analysis map also shows the mutual influence between clusters. The multi-dimensional analysis of high-frequency words divides the two-dimensional space into four quadrants. The keywords in the first quadrant belong to the important research content in the subject research, and the structure is close, indicating that there is a certain theoretical research foundation. The keywords in the second quadrant are far apart and cannot be self-contained, indicating that they are very unstable and may be decomposed and evolved into other parts in the subsequent development. The keywords in the third quadrant have a relatively small degree of dispersion and close internal links. In the future, they may develop into active topics in the research of the big data supply chain. The internal structure of the keywords in the fourth quadrant is relatively loose, and the centripetal is low. They are at the edge of the research network, indicating that such research is not yet mature.

From this we have summarized the three major application areas of big data in supply chain management:

(1) The financial industry. In 2008, Euromoney magazine referred to supply chain finance as "the hottest topic in bank transactional business". Since then, the new behavior of supply chain finance has not only become a new direction for the financial sector to pursue high profits, it has also attracted the attention of many scholars. At the same time, as the focus of the application of big data in the supply chain, scholars at home and abroad use the relevant research of supply chain finance as the starting point to conduct theoretical explorations and provide continuous and active help for the stable development of the supply chain finance industry.

(2) In the high-tech industry, under the background of the Internet, the development of the high-tech industry aims to create higher industrial value by relying on the efficient data collection, processing, and analysis capabilities of big data. With the emergence of a new round of consumption, mobile terminals, e-commerce platforms, and social software have undoubtedly become new existences that stimulate the economic development of enterprises. And these coexisting traffic and data through specific processing analysis and utilization will create unexpected business opportunities for the high-tech industry. At the same time, it also provides more precise consumer needs for e-commerce platforms and plans a more suitable development direction for the high-tech industry.

(3) Agriculture and medical industry, the application of big data in agriculture and medical industry not only has certain directional characteristics but also has the significance of theoretical innovation. The specific practice of the "Internet agriculture" theory is not only an innovation in the perspective and direction of supply chain services for big data in the Internet era but also an enrichment of supply chain management theories. In addition, the combination of big data and artificial intelligence technology for risk assessment can not only improve the ability of doctors to analyze tumors and make accurate diagnoses but also provide the same accuracy and detailed information as real tissue samples. It can also effectively reduce nursing costs, improve community health, and reduce the cost of providing medical services.

4.4. Social network analysis

The point degree centrality is the simplest of the three centrality calculations, and it represents the concentration of the overall network. This article first obtains the absolute point degree centrality, and then uses the formula to calculate the relative point degree centrality (standardized) of English high-frequency keywords, where $d(i)$ represents the absolute point centrality, which represents point i to other points j the sum of distances:

$$d(i) = \sum_j x_{ij} \tag{9}$$

$$C = \frac{d(i)}{n-1} \tag{10}$$

From [Table 4], it can be drawn: (1) "Big Data", "Supply chain" and other high-frequency words with the nature of parent disciplines not only have a higher degree of centrality, but also Large relative centrality. Compared with the previous year-by-year statistical analysis of English keywords, we can expect that the core research field of this subject will not change much in the future research and development process, and it will be a general trend for the deepening of theories and the integration of theories with reality. (2) "Application" and "Supply chain finance" are both at the forefront of the Google Scholar and Web of science keyword point degree centrality tables. This means that the application of big data in supply chain finance may become a hot topic in future big data supply chain research.

Table 5. Degree centrality of high-frequency in WOS (part)

Keywords	Centrality	Relative centrality	Keywords	Centrality	Relative centrality
Big data	176.0	12.1	E-commerce	28.4	2.5
Supply chain finance	113.0	7.8	Risk Management	22.3	1.7
Supply chain	98.0	7.2	Big data analytics	20.3	1.7
Logistics	54.0	3.6	Internet of things	15.0	1.4
Industry 4.0	37.0	2.9	Optimization	14.8	1.1
Supply chain management	31.0	2.4	AHP	14.6	1.0

Among the high-frequency keywords in English, the keywords “Big Data”, “Supply chain”, and “Supply chain management” with the nature of the parent discipline rank among the top in terms of centrality. Hot-frequency words such as “Industry 4.0” and “E-commerce” also have high intermediate centrality. This is because the goal of Industry 4.0 is mass customization. The core is intelligent manufacturing, and the realization relies on big data and cloud computing. In addition, in the process of e-commerce service model innovation and upgrading, the advantages of big data analysis will be used in the corresponding construction, which will be more conducive to the long-term and stable development of e-commerce.

5. Analysis and conclusions of research results

The article obtained the relevant literature on the big data supply chain direction in the Web of science academic literature database through the data crawler. After sample statistics and data preprocessing, we first use the word frequency analysis method to compare and analyze English high-frequency words and obtain relevant research hotspots and the annual development trend of each keyword. Then through the construction of a common word matrix, combined with cluster analysis, multi-dimensional scale analysis, social network analysis, and other research methods, the analysis of the interconnection between high-frequency keywords and the future research direction of the big data supply chain is the follow-up work Lay the foundation for development.

First, the integration with the fourth industrial revolution. The so-called fourth industrial revolution is an era in which high and new technologies are used to guide industrial changes, that is, the “intelligence era”. “Industry 4.0” is a growing keyword in the annual statistical tables of high-frequency keywords abroad, and its research enthusiasm is increasing, and it is a key research topic for experts and scholars. And many published documents are standing in the background of the Fourth Industrial Revolution, exploring how to face the challenge of big data to streamline operations and maximize efficiency and profit.

The second, the integration with emerging technologies. Experts and scholars combine the big data supply chain with the iconic emerging technologies of the "Internet" era such as the Internet of Things, cloud platforms, and machine learning, aiming to explore and innovate based on traditional models and methods to extend the business of the big data supply chain field.

Third, the countermeasure research of supply chain finance. Scholars continue to pay attention to the corporate financing constraints of supply chain finance, aiming to optimize the division of labor between the upstream and downstream of the supply chain through research on the overall structure of the supply chain, and continuously improve the overall value of the supply chain through innovation.

Fourth, simulation model research. Modeling and simulation play an important role in big data applications. Modeling and simulation can help developers perform analysis under different system configurations. One of the challenges of big data development is to balance its cost and performance by optimizing the configuration of software and hardware. In addition, supply chain analysis also provides a new method for big data simulation, helping companies to focus on maintaining the interconnection between supply chain operations and system integration.

The future research trend of the foreign big data supply chain: (1) It is roughly the same as the domestic big data supply chain research trend, “Big Data’, “Supply chain management” and other key subjects with the nature of the main subject Words have a relatively mature theoretical system and a high degree of centrality. They are mostly located in the first and second quadrants of the multi-dimensional analysis spectrum. Therefore, as the field of big data supply chain research continues to deepen, the importance of such keywords Research is bound to decline; (2) Keywords such as "Industry 4.0" and "E-commerce" have a high degree of intermediate centrality, reflecting the experts' and scholars' understanding of the big data supply chain and combining an exploration of the current situation and mainstream trends, it is foreseeable that many hot issues will continue to flood in future research; (3) Keywords such as “Demand Forecasting” and “Information security” are in the multidimensional analysis graph. In the second quadrant, the proximity to the center is also low. Although the current theoretical research is immature, it is getting more attention.

References

[1] J. Law and A. Rip, "Policy and the mapping of scientific change: A co-word analysis of research into environment acidification," Scientometrics, vol.14, pp.251-264, **(1986)**

[2] M. Callon, J. Law, and A. Rip, "Mapping the dynamics of science and technology: Sociology of science in the real world," London: The Macmillan Press Ltd, pp.103123, **(1986)**

[3] A. G. L. Herrera, M. J. Cobo, E. Herrera-Viedma, and F. Herrera "Visualizing the hybridizations between the fuzzy logic field and the other soft-computing techniques," Proceedings-8th International Conference on Hybrid Intelligent Systems, pp.252-257, **(2008)**

[4] M. W. Neff and E. A. Corley, "35 years and 160000 articles: A bibliometric exploration of the evolution of ecology," Scientometrics, vol.80, no.3, pp.657-682, **(2009)**

[5] A. A. Khasseh, F. Soheili, and H. S. Moghaddam, "Intellectual structure of knowledge in I metrics: A co-word analysis," Information Processing and Management, vol.53. no.3, pp.705-720, **(2017)**

[6] M. Granovetter, "The strength of weak ties," American Journal of Sociology, pp.1360-1380, **(1973)**

[7] R. V. Gould, "Collective action and network structure," American Sociological Review, pp.182-196, **(1993)**

[8] I. Nonaka and H. Takeuchi, "The knowledge-creating company," New York: Oxford University Press, pp.273, **(2008)**

[9] E. Otte and R. Rousseau, "Social network analysis: A powerful strategy also for the information sciences," Journal of Information Science, vol.28, no.6, pp.441-453, **(2002)**

[10] Bott, "Ambiguity and the process of knowledge transfer in strategic alliances," Strategic Management Journal, no.20, pp.595-623, **(2010)**

[11] R. Ghosh, K. Lerman, "Rethinking centrality: The role of dynamical processes in social network analysis," Discrete and Continuous Dynamical Systems - Series B (DCDS-B), vol.19, no.5, pp.1355-1372, **(2014)**

[12] R. L. de Andrade and L. C. Rêgo, "The use of nodes attribute in social network analysis with an application to an international trade network," Physical A: Statistical Mechanics and its Applications, **(2017)** S0378437117308579.

[13] L. Yong, Z. Wendi, and X. Yejing, "Discussion on the application of embeddedness theory in archives management and services," Archives Management, no.1, pp.4-10, **(2015)**

[14] E. M. Tucker-Drob and T. A. Salthouse, "Confirmatory factor analysis and multidimensional scaling for construct validation of cognitive abilities," International Journal of Behavioral Development, vol.33, no.3, pp.277-285, **(2009)**

[15] B. Marshalek, D. F. Lohman, and R. E. Snow, "The complexity continuum in the radix and hierarchical models of intelligence," Intelligence, vol.7, no.2, pp.107-127, **(1983)**

[16] W. Kaixuan, "Factor analysis and mu[[plh

This page is empty by intention.