Based on CiteSpace Knowledge Graph Analysis of the Status Quo of Research on Expenditure Forecast

Haoning Tang1,2, Xin Liang1*

1 Department of Management Engineering and Equipment Economy, Naval University of Engineering, Wuhan 430033, China
2 Military training base of Xinjiang Armed Police Corps, Urumqi, Xinjiang 830099, China
*Corresponding author’s e-mail: jessie@0xng.com

Abstract. This paper uses CiteSpace software to conduct text mining and visual analysis of 744 academic papers related to funding forecasts in the CNKI database in the past 20 years, sort out the literature time, journals, authors, and institutions of funding forecasts, and use keywords co-occurrence and subject terms Clustering and other methods reveal the current research status and evolutionary laws of funding forecasting, hoping to provide reference and reference for the selection of other types of funding forecasting methods.

1. Introduction
This article uses the CiteSpace information visualization tool, takes the journals included in the CNKI database as the object, and takes the period from 1980 to 2021 as the time range, and conducts an in-depth analysis and analysis of the funding forecast related literature from the aspects of keyword co-occurrence, cluster analysis and hot word emergence Comment, grasp the status quo and laws of development, summarize research deficiencies, predict development trends, and provide reference and reference for the selection of other types of funding forecasting methods.

2. Research methods and data sources
2.1. Analysis tools
This paper uses the information visualization software CiteSpace developed by Professor Chen Chaomei from the School of Computer and Information Science of Drexel University to conduct text mining and visualization analysis on the research topic of "expenditure prediction" [1]. CiteSpace is the abbreviation of citation space. CiteSpace is a multivariate, time-sharing and dynamic citation visualization analysis software that focuses on analyzing the potential knowledge contained in scientific literature. It can extract useful information from a large amount of literature and convert it into visual scientific knowledge. The map can visually present the laws hidden in a large amount of data, and effectively display the network, structure, interaction, evolution or derivation of many complex relationships between knowledge units[2][3].

2.2. Data sources
Search the CNKI database according to the subject containing "expenditure forecast" or "expense forecast". As of February 24, 2021, a total of 887 papers have been retrieved. Non-academic papers
such as conferences and press releases or articles irrelevant to the topic are excluded. Secondly, the selected data is exported in Refworks format, named as a CiteSpace recognizable name, and filtered, deduplicated and converted by the data converter, and finally a total of 744 valid data are obtained.

3. Text mining and visual analysis

3.1. Analysis of the basic situation of the research

3.1.1. Analysis of Annual Paper Volume

The number of documents is an important indicator to measure the development of a certain subject or topic, and it is also one of the important scales of academic research attention. According to Figure 1, it can be seen that the research of funding forecast is directly affected by the needs of actual tasks and the degree of development of machine learning and various algorithms.

![Fig 1. Annual publication volume statistics](image)

3.1.2. Distribution of articles in journals

Academic journals are an important link and an important carrier of academic communication, as well as the test of academic research results. By analyzing the source journals of the literature, it is helpful to determine the main source of the research field, and to a certain extent reflect the research in this field. Degree and so on [4].

| Serial number | Journal name | Publication frequency |
|---------------|--------------|----------------------|
| 1             | China Health Economics | 12 |
| 2             | China Health Statistic | 11 |
| 3             | China's health service management | 9 |
| 4             | System Engineering and Electronic | 8 |
| 5             | Journal of the Academy of Equipment Command & Technology | 8 |
| 6             | China Health Resources | 9 |
| 7             | China's public health management | 10 |
| 8             | Firepower and Command Control | 11 |
| 9             | Forecast | 15 |
| 10            | Soft Science of Health statistics and decision-making | 16 |
| 11            | Journal of Sichuan Armamentary | 18 |
3.1.3. Co-occurrence analysis of research authors

This article uses the micro-scholar cooperation network provided by CiteSpace to find out the cooperation between authors in the field. In the co-occurrence network of authors drawn by CiteSpace, the size of the node represents the amount of literature published by scholars. As shown in Table 2, the author's co-occurrence network contains a total of 683 nodes (N) and 582 edges (E), and the density (Density) is 0.0025. It can be seen that although a total of 683 scholars participated in the completion of 582 academic papers, the connections between scholars are not close, and they are still mainly individuals or small groups.

Table 2. Core author from 1980 to 2021

| Serial Number | Author Name         | Article Frequency |
|---------------|---------------------|-------------------|
| 1             | Zhang Hengxi        | 13                |
| 2             | Jiang Tiejun        | 5                 |
| 3             | Sun Linkai          | 5                 |
| 4             | Wu Lifeng           | 5                 |
| 5             | Jing Qi             | 5                 |
| 6             | Guo Feng            | 5                 |
| 7             | Lu Zuxun            | 4                 |
| 8             | Zhao Lei            | 4                 |
| 9             | Meng Ke             | 4                 |
| 10            | Chai Peipei         | 4                 |
| 11            | Lei Haichao         | 4                 |
| 12            | Ma Guifeng          | 4                 |
| 13            | Wei Ruxiang         | 4                 |
| 14            | Zhao Yuxin          | 4                 |
| 15            | Li Jiyuan           | 4                 |
| 16            | Zhou Jiazhong       | 3                 |
| 17            | Chen Xia            | 3                 |

According to "Price’s Law", the expression formula for the number of articles published by the core author is

\[
M_p \approx 0.749 \sqrt{N_p \text{max}}
\]

Among them, \( N_p \text{max} \) is the number of articles published by the most productive scholars in this field [5]. According to the results of the CiteSpace software, Zhang Hengxi has published 13 papers, ranking first, so the minimum number of published papers for core authors is 3. From this, there were 50 core authors from 1980 to 2021.

3.1.4. Co-occurrence analysis of research institutions

Table 3. Main issuing agency

| Serial number | Institution name                                           | Article frequency |
|---------------|------------------------------------------------------------|-------------------|
| 1             | Air Force Engineering University                           | 29                |
| 2             | Naval University of Engineering                            | 26                |
| 9             | Central University of Finance and Economics               |                   |
| 10            | Beijing University of Aeronautics and Astronautic          |                   |
It can be seen from the figure that the distribution of funding forecast research institutions is affected by the nature of disciplines and units. It can be seen that in the field of engineering disciplines, the main problem solved by funding forecasts may be of a military and national defense nature. However, according to the knowledge graph, it can be seen that organizations with strong connections are often in the same department, and there is still less cross-departmental cooperation.

### 3.2. Research topics and hotspot analysis

#### 3.2.1. Keyword co-occurrence analysis

According to Fig 2, "prediction" is the keyword with the most occurrences, and it is also the keyword with the strongest centrality. Excluding the keywords included in the search formula, the reference is of little significance. On the whole, medical expenses and maintenance expenses are the main application areas of the research on expenditure forecasting; gray forecasting model, support vector machine, BP neural network, ARIMA model, time Sequence analysis, regression analysis, etc. are the methods that scholars focus on in the field of funding forecasting.
3.2.2. Keyword cluster analysis

Based on the results of the cluster analysis, the research topics of funding forecasting are finally summarized into the following two aspects: First, research on the application of funding forecasts. Key words include medical expenses, total health expenses, research and development expenses, funding requirements, medical expenses, treatment expenses, etc. The second is the study of specific methods of funding forecast. Key words include GM (1,1) model, regression analysis, combination forecasting, grey system, quantitative analysis, etc.

3.2.3. Keyword emergence analysis

Use the mutation test function of citespace to get the keyword emergence map in this field, so as to understand the research frontier of this field. Based on the cluster analysis, this paper uses citespace to sort the keywords according to the length of the keyword emergence duration, the emergence start time, and the emergence intensity.

Based on the analysis of the emergence time length, emergence intensity, and emergence start time, the most prominent methods are support vector machines and ARIMA model methods. When making expenditure forecasts, a combination of forecasting methods can be used to combine the two to obtain the best prediction result.

3.2.4. Analysis of Research and Development Process

Keyword clustering sequence diagram is used to describe the historical evolution process of a certain cluster and the relationship between each cluster. Combining clusters with similar meanings together, cluster #0 (expense prediction), cluster #1 (prediction), and cluster #3 (prediction model) have continued to be research hotspots since 1986. The attention so far has not decreased. Cluster #2 (total health expenses), cluster #4 (hospitalization expenses), and cluster #5 (medical insurance) have been paid attention to by experts and scholars from 1987 to 2021, and they are still applying funding forecasts. The research on cluster #6 (model), cluster #7 (gray prediction), and cluster #9 (gray system) started in 1990, and the research continued until 2018.
Fig 4. Keyword timing diagram

4. summary
Research on funding forecasts has been an academic hotspot since 1986. Its application fields and covered subjects were initially mainly in the medical field and engineering equipment fields, but now they are more and more widely used in economics, education, sports, transportation, computers, etc. each field. With regard to the choice of the method of expenditure forecasting, according to the comprehensive analysis of the visualized results of citespace, the most promising and more mature methods are the support vector machine and ARIMA model.

Although funding forecasting research has always paid more attention and the scope and fields of application are more extensive, there is a lack of innovation in the selection of specific methods. Most of the widely used forecasting methods have been applied for more than 20 years and have universal applicability, but they are different. Funds have different characteristics. Therefore, when making fund forecasts, it is necessary to improve on the basis of existing forecasting methods in order to obtain more scientific and accurate forecast results.

References
[1] Chen Yue et al. (2015) Methodological function of CiteSpace knowledge graph. J. Research in Science of Science, 33(2): 242-253.
[2] Chen Yue, Chen Chaomei, Hu Zhigang, etc. (2014) Citation Space. Analysis Principle and Application: CiteSpace Practical Guide. Beijing: Science Press, Beijing.
[3] Li Jie, Chen Chaomei. (2017) Citespace: Science and Technology Text Mining and Visualization (Second Edition). Capital University of Economics and Business Press, Beijing.
[4] Sun Yusheng. (2014) The research progress of domestic knowledge graphs——Based on the analysis of CiteSpace II. J. Modern Information, 34(1): 84-88.
[5] Price. (1982) Little science, big science. World Science Press, Beijing: