Domestic Health Care Big Data Research Hotspots and Frontiers Based on CiteSpace Visualization Analysis

Haifeng Xu, Xinwen Chen, and Min Peng(✉)

Guangzhou College of Commerce, Guangzhou 511363, China
13986210591@139.com

Abstract. Health care big data reflects the development trend of medical science. It represents the developing direction of medical treatment. We use CiteSpace to mine the big data of health care in CNKI (2013–2022). The trend of publication, keywords emergence and clustering are analyzed in knowledge map visualization. We find health care, interconnection, medical association, medical health and medical center are the main research hotspots in this field. At the meantime, Big Data and Medical Health occupy frontier position .

Keywords: big data · medical health · CiteSpace · knowledge map visualization

1 Introduction

The innovation and application of modern science and technology such as big data, Internet and cloud computing have opened up a new era of informationization in human society, opened up a new situation of rapid development of information dissemination, brought more convenient experience to human society, and driven changes in various industries. Li Chunying and Zhang Weiwei analyzed the literature on big data and health management from PubMed through BI-COMB2.0, EXCEL and Gcluto1.0, and concluded that the application of data mining in the network of biological macromolecules such as proteins, electronic medical records and genome sequence disease prediction is a research hotspot in this field [1]. By analyzing WOS data (2010–2019), Yang Wenjing et al. obtained the research hotspots and frontiers of health care big data [2]. All of the above have made some contributions to the research of health care big data, but there are still limitations. Few scholars have analyzed the hot spots and frontier issues of domestic health care big data as a whole. In terms of methods, there are few studies on advanced measurement tools. The use of measurement tools, based on the perspective of knowledge map, can more objectively reflect the research hotspots and development frontiers in a certain field.

It can be seen that a comprehensive comparative analysis of the research hotspots and frontiers of health care big data is of great significance for promoting domestic theoretical research and practical innovation. This paper uses CiteSpace, a bibliometric software recognized by the academic community, as a research tool to visualize the

D. Kumar et al. (Eds.): IEIT 2023, AHSSEH 10, pp. 164–170, 2023.
https://doi.org/10.2991/978-94-6463-230-9_21

research results of health care big data from 2013 to 2022, and provides new ideas and directions for domestic research.

2 Research Methods and Data Source

By CiteSpace software, this paper uses the bibliometric method to carry out the visualization analysis of knowledge map, that is, using, High-frequency keywords emergence and clustering diagram, table and other research to obtain the research hotspots and research frontier of health and medical big data.

In this study, the data are derived from the CNKI network publishing database and the core collection database of foreign WOS. The specific search strategy of this study is: search with the subject word 'health care + big data', 1096 Chinese literature (2013–2022) were obtained and 1485 articles (2015–2022) were retrieved in WOS. In order to ensure the accuracy and scientificity of the data, A total of 2134 sample literatures on health care big data research at home and abroad were retrieved.

3 The Evolution Path and Hotspots of Health Care Big Data Research

3.1 Trend of Publication

The trend of annual papers can reflect the macro research trend in the field of health care big data at home and abroad. In this paper, 2134 articles from CNKI and WOS are sorted out, and the number of articles every year is drawn into Fig. 1.

From Fig. 1, it can be seen that the domestic health care big data in 2015 was only half of the number of foreign publications. During the period from 2015 to 2021, the number of foreign publications increased rapidly, which was related to the release of the white paper "Big Data for Development: Challenges and Development" by the

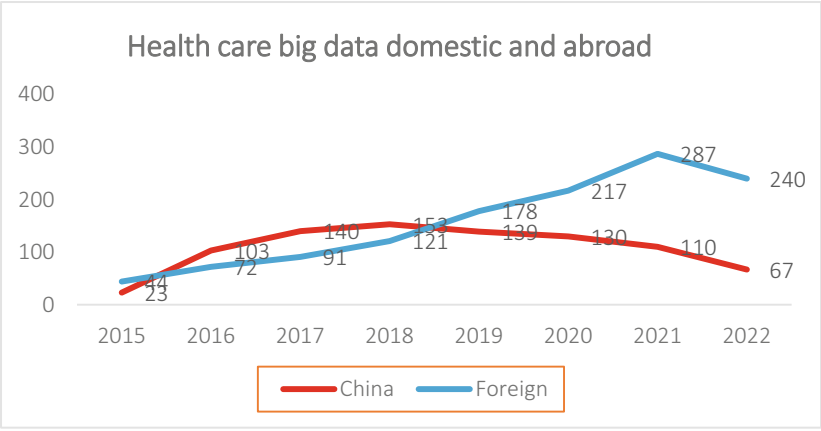

Fig. 1. Time series diagram of annual publication volume

United Nations in 2012.Former US President Obama announced the launch of the " Big Data Research and Development Plan." In 2016, China surpassed the amount of foreign literature and entered the outbreak period of China's health and medical big data publication. The reason was that in October 2016, the CPC Central Committee and the State Council issued the " Healthy China 2030" Planning Outline. Subsequently, China's publication volume peaked in 2018. Since 2019, foreign countries have also surpassed China. Domestic and foreign countries have begun to show a downward trend in 2021. China's reason is that the new coronavirus was first discovered at the end of 2019, and the focus of medical research has shifted to epidemic-related. With the opening of the epidemic in 2021 abroad, the attention of experts and scholars has also shifted to the epidemic.

3.2 Research on High-Frequency Keyword and Mutation

Keywords are the core of literature. High-frequency keywords are often used to identify hotspots in a research field. Chen Chaomei believes that the use of emergent subject terms is more suitable for detecting sudden changes and emerging trends in the development of disciplines than the use of high-frequency subject terms [3].

The emergence of keywords refers to the sharp rise in the frequency of keywords in a short period of time, which indicates that the research topics in this field have attracted the close attention of relevant scholars. The emergence of keywords in time can reflect the changes and dynamics of the topic. Table 2 shows the emergence intensity and mutation time of 25 keywords in CNKI health care big data literature from 2013 to 2022. Since 2014, the research topic of medical health has changed in stages. Researchers mainly focus on the policy, technology and management of health care.

In 2015, the mutation words appeared in the State Council, and the centrality was as high as 2.65, which indicated that there were many policy studies at the national level of health care big data.

Table 1. Keywords frequency top20

number	key word	frequency	year	number	key word	frequency	year
1	big data	209	2013	**11**	epidemic prevention	14	2020
2	medical health	79	2014	**12**	application	13	2016
3	health treatment	69	2015	**13**	medicine center	12	2018
4	smart healthcare	34	2014	**14**	industrial park	11	2016
5	artificial intelligence	23	2016	**15**	jinan city	11	2018
6	medical data	23	2016	**16**	internet plus	11	2016
7	cloud computing	21	2015	**17**	mobile health	11	2014
8	internet of things	19	2019	**18**	medical alliance	10	2017
9	data sharing	17	2015	**19**	healthy China	10	2017
10	health management	15	2015	**20**	blockchain	10	2019

Table 2. Top25 Keyword emergent graph

Keywords	Year	Strength	Begin	End	Keywords	Year	Strength	Begin	End
Health treatment	2014	3.92	2014	2016	Medical field	2015	1.16	2017	2018
State council	2015	2.65	2015	2016	Jinan city	2018	1.23	2018	2019
Public health	2015	1.25	2015	2016	Blockchain	2019	2.25	2019	2022
Data mining	2015	0.95	2015	2016	Ethical issue	2019	1.24	2019	2020
Information platform	2016	2.28	2016	2017	Health industry	2017	1.22	2019	2020
Interconnection and interworking	2016	1.99	2016	2017	Personnel cultivation	2017	0.96	2019	2020
health care	2016	1.49	2016	2018	China unicom	2019	0.82	2019	2020
Population health	2016	1.34	2016	2017	Epidemic prevention	2020	4.76	2020	2022
Medical alliance	2017	1.89	2017	2019	Data security	2020	2.81	2020	2022
Health information	2017	1.89	2017	2018	Artificial intelligence	2016	2.13	2020	2022
Graded diagnosis and treatment	2017	1.63	2017	2018	Medicine center	2018	1.78	2020	2022
Grassroots doctor	2017	1.48	2017	2018	Medical sciences	2018	1.25	2020	2022

In 2019, as can be seen from Fig. 3, blockchain was proposed. The application and development of blockchain technology plays an important role in promoting the development of the new generation of information technology industry.

In 2020, epidemic prevention and control appeared for the first time, and the centrality was as high as 4.76. Liu et al. believed that it was necessary to improve policies and regulations to promote the open sharing of prevention and control data and promote the standardization of services in this field [4] (Table 1).

3.3 Keyword Cluster Analysis

Based on the analysis of keyword co-occurrence and burst in the field of health care big data, this paper further understands whether the research hotspots in this field have commonality. This paper uses CiteSpace to cluster keywords using the ' LLR' algorithm (Log-Likelihood Ratio). As shown in Table 3, there are 8 clustering categories, which are # 0 health care, # 1 interconnection, # 2 medical association, # 3 medical health, # 4 medical center, # 5 intelligent medical care, # 6 data sharing, # 7 medical service. The module value Q is 0.55, which is greater than 0.3, indicating that the module structure is more reasonable. The average contour value S is 0.86, which is greater than 0.7, indicating that the clustering is convincing.

Cluster # 0 focuses on health care, and the core keywords include big data, health care, and national health. The keyword clustering timeline diagram shows that big data first appeared in 2013, and then big data was applied in medicine, from the technical level to the application level. In November 2015, the Fifth Plenary Session of the 18 th CPC Central Committee passed the '13th Five-Year Plan'. Cluster # 1 is centered on interconnection, and also includes applications, health management, mobile medical care, traditional Chinese medicine, and countermeasures. In 2020, Wang Shengfeng,

Table 3. Keyword clustering diagram

ClusterID	ClusterID	**Size**	**Silhou**	**mean**	**Label (LLR)**
0	0	52	0.887	2017	Health treatment; big data; medical health; health care big data; hospital management
1	1	28	0.802	2016	Interconnection; information platform; application; population health; medica.
2	2	27	0.902	2017	Medical association; big data, health information, grassroots doctors, health care
3	3	26	0.837	2016	Medical health; internet; healthcare: data sources Unstructured data
4	4	25	0.84	2018	Medical Centre; industrial park; jinan; jmc; medical science
5	5	24	0.882	2016	Smart medical; artificial intelligence; medical field; medical data: healthcare
6	6	20	0.687	2019	Data sharing; privacy protection; privacy; protection; privacy
7	7	16	0.954	2016	Medical services; personal health; personalized medicine; medical insurance; health data

Ning Yi proposed that interconnection is the biggest barrier to the use of health and medical big data in countries around the world [5]. Clustering # 2 medical association, its main core keywords are medical association, top three hospitals, specialized hospitals, primary hospitals, and off-site insurance. The main core keywords of cluster # 3 medical health include medical health, precision medicine, and talent training. Cluster # 4 medical center, the core keywords are medical center, industrial park, medical science, Jinan. Clustering # 5 Intelligent medical, the main keywords are artificial intelligence, medical field, medical data, health care. Clustering # 6 data sharing, the main keywords are data sharing, privacy protection, privacy, protection, privacy. The data generated by the medical industry involves the personal privacy of many patients. Clustering # 7 medical services, the main keywords are personal health, personalized medical care, medical insurance, health data.

As shown in the Fig. 2 and Fig. 3, this is a good example of using citespace to process big data, resulting in a timeline diagram and a cluster diagram. From the timeline chart, we can see that big data first appeared in 2013, and then it was applied in medicine, from the technical level to the application level.

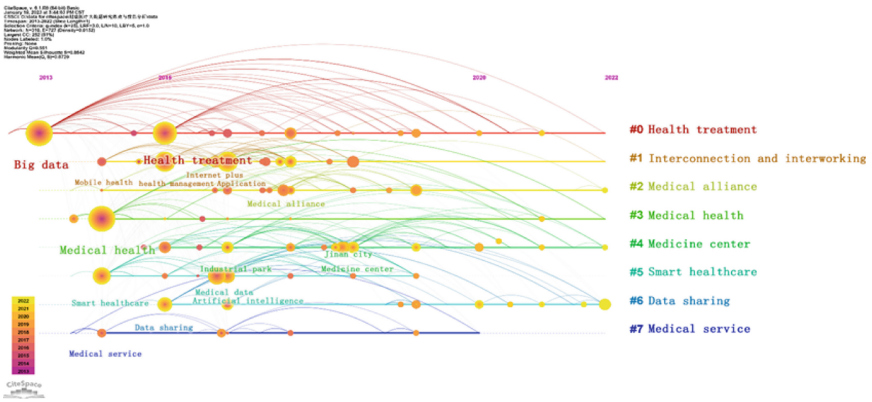

Fig. 2. Keyword timeline diagram

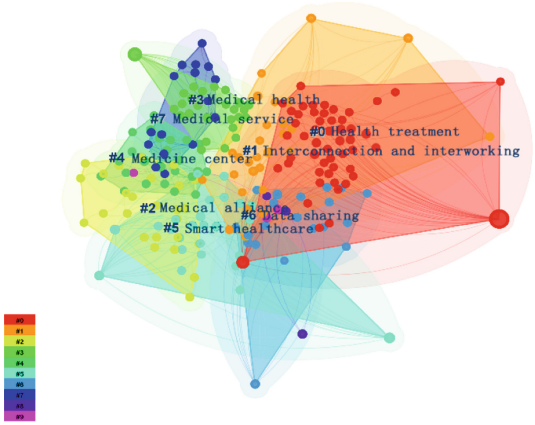

Fig. 3. Keyword clustering diagram

4 Conclusion and Future Prospects

It can be seen from the trend of annual publication that since 2013, health care big data has received extensive attention from scholars. At present, compared with foreign countries, the number of publications in this field is still in a relatively small state, but it is predicted that it will rise after 2023. By analyzing the keyword table, emergent graph and cluster analysis, it can be seen that the current research hotspots in China focus on health care, interconnection, medical association, medical health, medical center and other topics, while blockchain, epidemic prevention and control, data security, artificial intelligence, medical center and medical science occupy the forefront. Combined with China's national conditions and the progress and frontier of health care big data research, the future research in this field is prospected: (1) The privacy risk of health care big data is about to become a research hotspot. (2) The application of blockchain, artificial

intelligence and other technologies in health care big data will continue to heat up. (3) Health industry research started and will grow rapidly.

References

1. Li Chunying, Zhang Weiwei. Cluster analysis of research hotspots in global big data and health management [J].China Hospital Management, 2016,36 (10) : 63–65.
2. Yang Wenjing, Du Ranran, Zhang Ran, Gao Dongping, Chi Hui. Research hotspots and frontier analysis of health care big data based on Web of Science database [J].Chinese Journal of Health Information Management, 2020, 17 (06) : 809–814.
3. Lu Wanhui, Ma Jianxia, Zhao Yingguang. Research on burst word recognition and topic detection technology [J].Information theory and practice, 2012,35 (06) : 125–128.DOI : 10.16353 / j.cnki.1000–7490.2012.06.028.
4. Liu Yu, Cao Qingxin, Cao Wenwen, Han Xiong, Wang Jing. Ideas and countermeasures of health care big data to help the prevention and control of COVID-19 epidemic [J].China Digital Medicine, 2020,15 (05) : 70–72.
5. Wang Shengfeng, Ning Yi, Li Liming. Experiences and challenges of health care big data interconnection model [J]. Chinese Journal of Epidemiology, 2020,41 (03) : 303-309.