

Study of Chinese Traditional Culture under the Background of "Big Data"

Yang Wang^{1, a}, Shuyan Yi^{2, b, *}

¹Jingyue Nanhuan Primary School, Changchun 130033, China

²College of Humanities & Sciences of Northeast Normal University, Changchun, Jilin, China, 130117

^a626705792@qq.com, ^byishuyan2013@126.com

*Corresponding author

Keywords: Big data, Chinese traditional Culture, Network Science.

Abstract: The era of big data has come. Big data not only changes our lives, but also changes our way of thinking. With the rapid development of cloud computing, image retrieval and other technologies, literary research is bound to enter a new era. [1] By expounding the present situation of literary research related to big data at home and abroad, this paper finds out all kinds of new changes in the field of classical literature, and puts forward some suggestions on how to carry out the study of Chinese traditional culture with the help of big data. With the help of this kind of promotion, this paper not only comprehensively discusses the specific influence of "big data" in the literary world, but also provides a reference for the new feasibility study of Chinese traditional culture.

1. Introduction

With the popularization of computer and the development of network technology, the new stage of information age, big data era, is coming quietly. It has not only changed the development mode of all aspects of social life, but also broken all kinds of traditional research models in the field of humanities and social science. In today's humanities research team, in addition to the traditional people with academic background such as history, geography, literature, art and so on, there are also new people with a background in science and engineering such as statistical science. In the aspect of research and application, it breaks the shackles that traditional literary research relies solely on text analysis, and makes digital computing, association network, data statistics, text recognition, subject classification, mathematical modeling and other technologies also play a role in the field of humanistic research. This can be seen in the seminar on "Digital Humanities: academic Frontier and Exploration in big data era" held at Nanjing University in July 2017. [2] In foreign countries, the practice of using "big data" technology to carry out literary research has been relatively mature. It has become a consensus in academic circles to establish a complementary relationship between big data analysis technology and scholars' individual reading. Scholars have explored a set of effective research models, and the corresponding results are rich and fruitful. In China, although some scholars in various fields of literature have also taken corresponding measures, the influence of "big data" in all directions is quite different. In contrast, the research direction of Chinese traditional culture has been carried out at a great pace, whether it is the development and utilization of various databases, or the use of "big data". Some fields have even taken the lead in entering the "era of analysis." [3] However, it should also be noted that the development of the whole field of traditional culture research is uneven. Big data technology is more widely used in the field of verse, especially in the direction of Tang and Song poetry and Song Ci, and its research is already in the stage of "data" analysis, while other fields are only in the stage of "digitization".

The paper will discuss how to apply big data technology to the study of traditional culture from the following three aspects.

2. Definition and Development of Big Data

Big data is written in English as "Big Data", which literally means "a lot of data". McKinsey interprets big data as: the amount of data involved is so large that it can not be intercepted, managed, processed and sorted out into information that can be interpreted by human beings in a reasonable time. In the era of big data [4], Viktor Mayer-Sherberg of the United Kingdom further pointed out the various effects brought about by big data's emergence: Big data is in an unprecedented way, through the analysis of massive data, to obtain products and services of great value, or profound insight'. Big data marks that the 'information society' is finally worthy of its name. All the digital information we collect can be utilized in a new way. We can try something new and start a new form of value. But this requires a new way of thinking and will challenge our social institutions and even our sense of identity. However, most people now think that data is a technical problem and should focus on hardware or software, and we think more consideration should be given to what happens when the data is spoken.”

The emergence of big data has caused great changes in all aspects of society. In the 2015 platform for Action for big data's Development, Premier Li Keqiang attached great importance to the role of big data in future economic and social development, he said: "persisting in innovation-driven development, speeding up big data's deployment, and deepening big data's application have become the internal needs and inevitable choices for stable growth, promoting reform, readjusting the structure, benefiting the people's livelihood, and promoting the modernization of the government's governance ability." [5] In addition, he also stressed many times that he will promote the development of big data and the Internet, so as to transform and enhance the development potential of traditional industries and promote the transformation and upgrading of China's economy. It can be seen that under the background of big data's view, big data is closely linked to the grand strategy of promoting China's economic transformation and upgrading.

3. Big data and the study of traditional Culture

Big data technology appears to be more closely and directly related to business and economic development, especially in the field of scientific research, especially in science and engineering. However, it has unbelievably established a subtle docking relationship with literary research at present, and relevant scholars have found a more mature research system. This not only provides practical and effective convenience for literary research, but also brings new research results.

In 2000, the foreign scholars began to use big data to carry out literary research. Franco Moretti, a professor at Stanford University in the United States, is a leader in this field. He led the team to set up the Stanford Fiction Laboratory and has achieved a wealth of research results. By using quantitative analysis to study literature, it is proved to the world that it is practical to study literary ontology with the help of data analysis, and it can provide new research ideas for literary research. In his book *The Great Analysis*, Qiao Kes specially introduces how to apply big data's analysis to literary interpretation, in order to make data statistics a necessary means for literary researchers to accept. [6]

China's relevant scholars have also actively explored this new field. In 2005, Li duo and Wang Yi pointed out very forward-looking: "the future relationship between people and computers is not a matter of who replaces who, but of communication and inspiration, dialogue and integration. Of course, this is not absolutely equal. Human subjectivity is the first, but we also have to learn from the computer, including entering its way of thinking: to integrate all kinds of knowledge, but also to make up classes. When cultivating some graduate students in the field of classical literature research, we should open courses such as electronic information, statistics, etc., we should draw lessons from sociological data statistical methods, and so on, in the knowledge structure, especially in the design of research methods and research paths. Make up for the obvious lack of our discipline in the past. "Although this understanding points out the correct direction for the current literary research, the process is extremely slow in concrete practice. According to the retrieval results of

China knowledge Network, the papers with "big data" in the title show the trend of blowout in recent years, and the specific data are shown in figure 1.

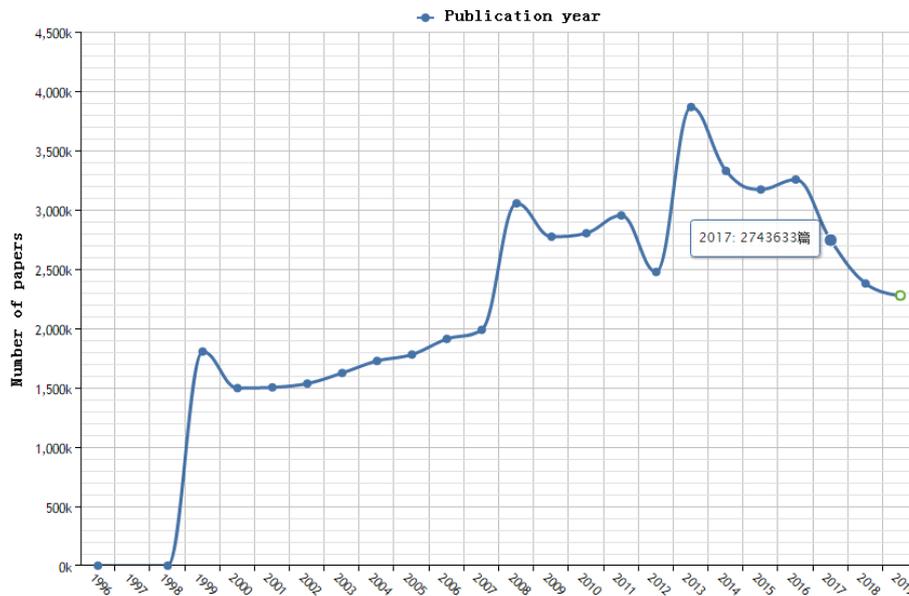


Fig. 1 The Development trend of the number of big data's papers in the past years

4. "Digitization" and "data" in the study of Chinese traditional Culture

It can be found through the search of related papers in recent years: the academic research and discussion with "big data" is relatively concentrated in the research fields of Chinese classical literature, literature, network literature and linguistics, especially Chinese classical literature as the most. [7][8].

Compared with other literary fields, the construction of thematic database in classical literature is more complete and rich. Such as the Si Ku Quan Shu, the Compilation of historical materials of stone carvings in the past dynasties, the Book of Treasure of Chinese Studies, the Basic Ancient Book Bank of China, the Twenty-five history, The poetry of the whole Tang Dynasty, the Full Song Poems, the various "Notes on the Tang and Song Dynasties" and the like. Such resources provide great convenience for researchers engaged in traditional culture in literature collection, all they have to do is type in the name of the document they need in the search bar, and all versions of the book and literature will be displayed in the form of electronic information in front of the searcher. This provides great convenience for the research work both in increasing the richness of literature and in improving the efficiency of literature search.

The existing large-scale classical literature retrieval platform is often developed and designed by commercial companies, and can only provide information services "full-text retrieval". The establishment of this kind of data platform is the first step in the digitization process of classical literature, although it can help researchers to deal with the empirical research of pure literature, such as proofreading, collection, exegesis and other pure literature, notes, appreciation, chronology and other aspects of literature and art and history in classical literature research. However, there is nothing he can do about the research at a more advanced level, such as the rational thinking of the universe, society, life and so on. To some extent, big data is the frontier technology of data analysis, which can quickly obtain valuable information from various types of literature data, which is big data technology and big data's core value.

In view of this deficiency, the relevant scholars have actively explored and made some progress in technology. The whole Song Poetry Analysis system can be regarded as a model of data analysis database. In short, this kind of database has more intelligent characteristics, not only can the computer provide the information that has been analyzed and screened, but also can do some very important things directly, such as "word frequency analysis", "lattice analysis", "use the analysis",

“repeat poem extraction” , “word system analysis” and so on. The development of such a system has opened an exhaustive study mode for better combining computer technology with artificial intelligence for panoramic scanning.

In recent years, although the data process in the field of classical literature has been greatly improved, the database in many research fields is not perfect, and there is a lot of room for improvement both in time and form. The point of interest of the scholar is only limited to the poetry text of the Tang Dynasty, and more is the combing work of a poet, an image and other thematic literature history. Obviously, this single-point database is far from enough for the whole field of classical literature. There is an urgent need for scholars to carry forward this kind of database success example and put it into the study of poetry texts and other styles in other times, so as to achieve the beautiful blooming of big data's flowers in classical literature.

5. Conclusion

In short, we are currently in an era of rapid development and rapid popularization of "big data" and "cloud computing". As literary researchers, we should change our thinking quickly, take the lead in using "big data" to carry out the relevant research of Chinese traditional culture, turn machine analysis into a tool and means of personal interpretation, and realize the blending and cooperation between literary "big data" analysis and individual scholars "small reading", so as to promote the progress of Chinese traditional culture.

Acknowledgements

This paper was funded by the project of Jilin Province Department of Education (NO. JJKH20181309SK). Shuyan Yi is the corresponding author of this paper.

References

- [1] Q.Z. Zhang, L. Han, et al. 2017. Study for Modern and Contemporary Chinese Literature in “The Big Data”Era. *Journal of Chongqing normal University*, Vol. 5 (5-9).
- [2] Q. Liu, Z.P. Lu, et al. 2018. The Frontier and Exploration of Humanities Research in big data era: a Summary of the Symposium on "Digital Humanities" in Nanjing University. *Library forum*, Vol. 38, No. 3 (37-45).
- [3] D. Li, Y. Wang, et al. 2005. The Dialogue concerning Ancient Cultural Information Processing and Classic Chinese Literary Studies. *Literary heritage*, Vol. 1 (126-137).
- [4] X.H. Chen. 2016. Literary Research Methodology in Big Data Era: A survey of Franco Moretti's Quantitative Analysis. *Theoretical Studies in Literature and Art*, Vol. 36, No. 2 (70-77).
- [5] <https://baike.so.com/doc/10955176-11483156.html>.
- [6] http://kns.cnki.net/kns/brief/default_result.aspx.
- [7] X.Y. Zheng. 2014. Speeding up the Transformation from "digitization" to "data"-the Theory of 'big data' and "Cloud Computing" and the study of Classical Literature. *Literary Heritage*, Vol. 6 (141-148).
- [8] J.C. Liu. 2015. The study of Classical Literature in big data's era-centered on data Analysis, data Mining and Image Retrieval. *Literary Heritage*, Vol. 3 (182-190).