Research on Data Visualization Based on Big Data

Shasha Xu, Kouquan Zheng, Wenjing Yang, Yanming Sun

School of information and Commmunication National university of Defense Technology Xi'an, 710061, China

Keywords: Big data; data; visualization

Abstract: In the context of large data, with the progress of information technology, a large number of data can only be obtained by visualization. The author studies the application of data visualization under the background of large data. This will help us to develop and innovate data visualization technology. And summarize the research characteristics, in order to provide reference information for the in-depth development of big data research and the development of future big data. The research results show that in order to complete the rapid processing and visualization of big data, researchers not only need to purchase and maintain a distributed cluster environment, but also need programming capabilities in a distributed environment and corresponding front-end data visualization technology.

1. Introduction

With the arrival of the era of big data, the amount of data used in scientific research is increasing. Data analysis tools in a single computer environment are not enough to support the needs of scientific researchers [1]. The "big" of "big data" lies not only in the "big capacity", but also in the discovery of new knowledge, creation of new value and bringing "big knowledge", "big profits" and "big development" through the exchange, integration and analysis of massive data. Its characteristics can be summarized into four "v": variety, volume, velocity and value [2]. Mc Kinsey, a well-known consulting firm, claims that "data has penetrated into every industry and business function area and become an important factor of production [3]. People's mining and application of big data heralds a new wave of productivity growth and consumer surplus"[4]. This also prompts us to put the data in the database for data management and storage, and also greatly improve the quality of data and information processing requirements [5]. At present, with the rapid development of smart grid, the extensive deployment of smart meters and the wide application of sensing technology, the power industry has produced a large number of data with diverse structures and complex sources. How to store and apply these data is a difficult problem for power companies [6].

The high integration of human, machine and material world has led to explosive growth of data scale and high complexity of data model. The world has entered the era of large data network. Artificial intelligence-related industries and technologies have also been highly concerned and actively invested by governments [7]. It means a variety of data sources, a huge amount of data, fast and low-cost processing. Big data brings opportunities and challenges, but the intuitive feelings of data are always very different [8]. In order to visualize the relationship between data and information and data, we need to make data concrete [9]. For example, when the data file size reaches the GB level, Excel can not open all the files, and it is difficult to analyze and visualize all the data. The US government believes that big data is “new oil in the future.” The size of a country's data and the ability to use it will become an important part of its overall national strength. The possession and control of data will become the new focus of competition between countries and enterprises. [10]. Visualization not only visualizes figurative data and information, but also transforms abstract data information into concrete and visual representations, which makes it easier for us to observe the changing laws of data.

  DOI: 10.25236/iwmecs.2019.056

2. Methodology

Current data analysis tools are mainly divided into two kinds: data analysis tools in stand-alone environment and data analysis tools in network environment. The existing data application means still remain at the level of data query and statistics, and lack of targeted, deep-seated analysis and information extraction means for data. The value of application data aggregation has not been fully exploited, and the ability of decision support information acquisition, which is more concerned by application personnel, is insufficient. Visualization technology plays an important role in important scientific discoveries. Data visualization appeared in the 1950s, referring to the use of computer graphics and image processing technology. Presents data in charts, maps, label clouds, animations, or any graphical way that makes the content easier to understand, making the content expressed by the data easier to understand. Cites pace III is an information visualization software for measuring and analyzing scientific literature data. It has the characteristics of multiple, time-sharing and dynamic. It uses the time-sharing dynamic visualization to show the macro structure and development of scientific knowledge. And intuitively display the entire content of information in a certain field, identify and display new trends and new developments in the scientific development of a certain field, and show research hots pots and frontier directions.

Taking sorting as an example, this paper compares the performance of the "sorting" algorithm DC-Top-k used in this project with the performance of the traditional serial algorithm. The effect is shown in Figure 1.

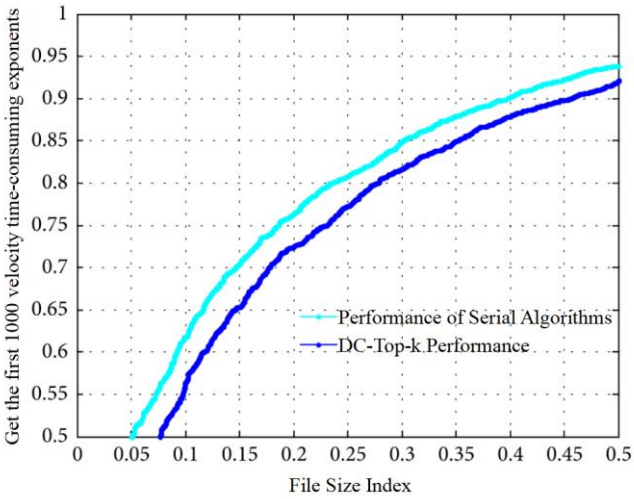

Fig.1. Performance Ratio of DC-Top-k and Serial Algorithms

The visualization of information technology, which we generally understand, refers to the transformation of a large number of concrete and abstract information and data into images and lines that we can intuitively understand, so as to achieve the purpose of easy understanding. In real life, the changes of a country's economy will cause economic shocks and changes in the lives of many countries and regions concerned. By collecting, analyzing and collating these data, journalists can find out the reasons. It reveals the economic laws and predicts the possible changes in the future, thus providing a certain basis and support for the formulation and development of relevant policies. This is what we need to translate a large number of concrete and abstract data into line trends and images that we can see at a glance and visualize, through which we can draw a clear conclusion. In this way, the Cites pace III software can be used to perform keyword co-occurrence analysis, institutional cooperation analysis, author collaborative analysis, literature co-citation analysis, author co-citation analysis, and journal co-citation analysis. And draw the corresponding scientific knowledge map to reveal the hots pots and frontiers of research in the field of big data. In view of the problem that the existing application data has more basic data and less decision support data, under the support of application information top-level planning and effective data integration, comprehensive use of data statistics, data analysis, data mining and other methods to obtain from

multiple data collection channels. In the large-scale application data, refinement can guarantee the knowledge of decision support information serving the application command. Improve the ability of application actions to respond quickly.

The development of big data in recent years is receiving widespread attention and ushered in a development boom. It can be clearly seen from Figure 2 that global big data development is in contrast to the development trend of artificial intelligence, and research is entering a stage of rapid development.

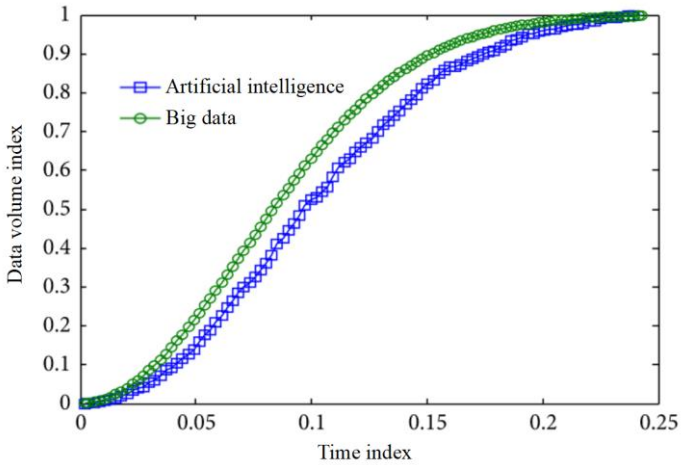

Fig.2. Contrast the Trend of Big Data and Artificial Intelligence in Recent Years

3. Result Analysis and Discussion

However, the above applications have not been able to support large data very well. The growth and regularity of scientific knowledge are closely related to the growth and regularity of scientific literature. The change of the quantity of scientific literature reflects the change of the quantity of scientific knowledge. Therefore, the quantity of scientific literature is one of the important measures to measure the quantity of scientific knowledge. Mining users' behavior habits and preferences, finding products and services more in line with users' interests and habits behind the messy data, and adjusting and optimizing products and services pertinently are the value of big data. Multi-level decision support information needs, can provide dynamic application data services according to application needs, and can ensure that the command process at any time, anywhere can enjoy the scope of application information services. Under this background, large data visualization technology, as a tool that can effectively simplify and refine data flow, has gradually developed to visualize large and complex data. Finally, according to the characteristics of the corresponding data and information, the corresponding modeling is carried out, and the data and information are analyzed and processed through modeling, so as to obtain the rules and conclusions expressed by the data and information. There are also a lot of multidimensional data in a lot of data and information, and visualizing these multidimensional data is a key point. There are many ways to visually design these data, choosing different methods for different data types.In addition to simply presenting data state, data visualization also has a very practical function, which is to compare several visualized data with relevance. It can mine the important association between data or present a reasonable data development trend. The Atlas of scientific knowledge involves the theories of mathematical statistics, computer science, sociology, information science and image science. Combining with the methods of Co-word Analysis and co-citation analysis of scientometrics, a series of visualized graphs are drawn through the mining and processing of scientific knowledge, and the development process and structural relationship of subject knowledge are visually displayed.

From the cited frequency of Table 1, we can see that the articles with high citation frequency in the field of big data research in China mainly appeared in 2011 and 2012.

Table 1 Keyword appearance frequency and centrality table

Key words	Frequency of occurrence	Centrality
Big data	817	0.25
Map Reduce	158	0.18
Cloud computing	126	0.16
Hadoop	109	0.14
Data mining	65	0.11

In the era of big data, everything is closely related to the numbers and data behind it. The prosperity of data news is closely related to the current era background. In today's big data era, we are faced with a large number of data and information to be processed every day, a small amount of data we can analyze the rules and conclusions. But for a large number of data and information, we can not analyze the law of data change and draw some conclusions in a short time. The fundamental of big data lies in data mining, which is a decision support process. The knowledge map is one of them. It is a graph that describes the relationship between the development process and structure of scientific knowledge. The application data construction will focus on service application and security command, and can integrate and accumulate relevant data resources obtained from various channels such as the business support department, and obtain valuable decision support information from large-scale application data resources. In other words, if big data is likened to an industry, then the key to profitability in this industry is to increase the “processing power” of the data and “add value” of the data through “processing”.

4. Conclusion

In recent years, data visualization technology has become more and more mature, and the types of charts have become more and more complicated. Domestic research on big data mainly focuses on the application of algorithms and technologies in the IT field, and the field of library and information is still in the preliminary exploration stage. As users' demand for data analysis grows, so does the need for data visualization. Regular users are constantly being asked or actively involved in the process of designing and creating visualization projects. Facing complex, dynamic and uncertain application environments, it can support portable terminals to quickly access and accurately obtain information, and at the same time provide personalized information services for different application personnel. Support the flexible organization, multi-perspective and all-round display of information content according to the focus of different application personnel, to meet the needs of application personnel at different command levels to grasp the application information conveniently and timely. Overall, cloud computing, Internet of Things, mobile Internet and other emerging computing forms are not only the place where big data is generated, but also the field where big data analysis methods are needed. Data visualization can improve production efficiency, save production time and promote economic progress.

References

[1] Chen M, Mao S, Liu Y. Big Data: A Survey [J]. Mobile Networks and Applications, 2014, 19(2):171-209.

[2] Varian H R. Big Data: New Tricks for Econometrics [J]. Journal of Economic Perspectives, 2014, 28(2):3-27.

[3] Lazer D, Kennedy R, King G, et al. The Parable of Google Flu: Traps in Big Data Analysis[J]. Science, 2014, 343(6176):1203-1205.

[4] Walker S J. Big Data: A Revolution That Will Transform How We Live, Work, and Think[J]. Mathematics & Computer Education, 2014, 47(17):181-183.

[5] Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics[J].

International Journal of Information Management, 2015, 35(2):137-144.

[6] The rise of “big data” on cloud computing: Review and open research issues[J]. Information Systems, 2015, 47:98-115.

[7] Chen C L P, Zhang C Y. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data[J]. Information Sciences, 2014, 275(11):314–347.

[8] Jagadish H V, Gehrke J, Labrinidis A, et al. Big data and its technical challenges [J]. Communications of the ACM, 2014, 57(7):86-94.

[9] Kambatla K, Kollias G, Kumar V, et al. Trends in big data analytics [J]. Journal of Parallel & Distributed Computing, 2014, 74(7):2561-2573.

[10] Specht, D. Book Review: The Data Revolution: Big Data, Open Data, Data Infrastructures and their Consequences [J]. Media, Culture & Society, 2015, 37(7):1110-1111.