Research on Big Data Analysis and Visualization Tools

Yao Ru

Agriculture Information institute, Chinese Academy of Agricultural Sciences, China.
Email: 550825948@qq.com

Ning Jing

International Association of Applied Science and Engineering Technology.

Abstract—**In recent years, big data has drawn a heated discussion in our society. The wave of big data makes people pay more attention to data visualization. Actually, visualization technology can dig the potential value of data and show to users in a vivid way. This paper is designed to introduce and analyse several visualization tools in order to summarize the strategy of choosing tools. This paper selects some visual tools and methods of big data analysis, and summarizes the characteristics and functions of the tools, and gives the contrast analysis of the advantages and disadvantages of these tools. Different user has different requirement and emphasis. This paper puts forward some views on the selection strategy and development tendency of data visualization. This paper summarizes the background, objective, research significance and research status review of data visualization. The innovation points are that principle and method of big data analysis are concluded, and the basic situation of several typical visualization tools—R language, Processing, Google Chart API and D3 is introduced. This article sums up the advantages and disadvantages to provide users with some suggestions when choosing the visualization tools by contracting its performances and features. Meanwhile, it will promote the development of information society and minimize the weakness in the visualization research in future.**

Index Terms—**bigdata, visualization tools, visual analytics**

I. INTRODUCTION

The boom of big data has aroused the attention of all walks of life, and also led to a heated discussion on the big data from experts and scholars in all over the world. There are a series of related research and problems worthy of our deep thinking behind the big data boom. The massive amounts of data means greater value of data mining, and mining the deep value of these data can give the industry and enterprises tremendous business value, so as to enhance their own economic and social benefits [1]. The key is to see big data applications. The traditional computing technology and information system processing capacity has been unable to meet our needs because of the explosive growth of data and the rapid expanding scale of application system within the industry, which needs to seek effective big data analysis, processing techniques, methods and tools. Then how can we realize the vivid and visualize present of the results of

the big data analysis so that the public can easily accept and understand, and benefits decision-making? What we need is data visualization. Data visualization can reflect the value of the data, and the value can help companies to sort out the complex data, thus effective decision-making. Actually, data visualization has already turns out in all aspects of human life, for example, Amazon recommends commodity information for customers by using users' click data, and PayPal provide users with the end of the bill through associated data.

The current data visualization products are divided into two leading markets, commercial and non-commercial products. This survey was only aimed at commercial visualization products to investigate, such as Tableau, Microsoft, Qlik, IBM, Tibco software, SAS, SAP and Oracle because these kinds of products have too many types or amount [2]. And then it analyzed from three points of view: popularity, visibility and leader, as shown in table 1:

TABLE I.

THE DATA VISUALIZATION INVESTIGATION INTO ITS POPULARITY,

VISIBILITY AND LEADER

INDEX TOOL	Popularity
Tableau	13%
IBM	14%
Microsoft	13%
Qlik	12%
Tibco software	11%
SAS	13%
SAP	12%
Oracle	13%

From table 1 can be seen that the visibility of these software is not divided, the gap is minimal, the highest is IBM; In terms of popularity, SAP, IBM and SAS are more popular than the other software; And from the perspective of the leader, Tableau has an overwhelming advantage owning 40% of the proportion. And this provides the users with guidance when they are selecting

data visualization products.

Through the above analysis, it can be seen that the use of data visualization in China is still in the initial stage, and there is still much development space in the future. And foreign data visualization products have occupied leading position in marketing, whether it is in the visibility or its market share, which shows that the data visualization technology already have a certain foundation. Big data technology continues to develop deeply, the domestic focus on data visualization is also growing, and we have reasons to believe that domestic data visualization technology will be more mature and more and more data visualization products could occupy the market in the future.

In this paper, the methods, the technologies and the corresponding visualization tools of the big data analysis are summarized and classified, and it means that the proper tools should be used in proper scene. On this basis, we select some of the visual tools and methods of big data analysis, and summarize the characteristics and functions of the tools, and give the contrast analysis of the advantages and disadvantages of these tools in order to provide users with valuable references in selecting proper data visualization methods. At last, it puts forward some views on the development of data visualization.

II. BIG DATA ANALYSIS AND VISUALIZATION METHODS

A. Analysis Principles of Big Data Visualization

Big data analysis is a process of extracting the information and knowledge from large, incomplete, fuzzy and random data in which people are not known in advance, but potentially useful information and knowledge [3]. We need to seek the visualization method to discover knowledge and the visualization technology and tools, in order to understand the process and result of the knowledge and discover and understand the relationship between the data and the development trend of the data.

Among the numerous visualization tools, we should have the insight to select the correct tools to achieve the desired results. Of course, if these steps such as data collection, data analysis or data selection are not required, we could not get the expected results. So the steps of data processing are also very important. In the literature [4], Ben Fry divided data visualization into seven steps: acquisition, analysis, selection, mining, description, refining, interaction. The process flow is shown in "Fig.1":
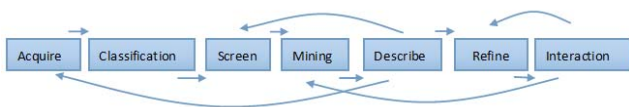

Figure 1. The Flow Chart of Visualization Process.

Through a series of steps of data visualization, we can roughly understand how to manage and visualize big data to offer users.

B. Analysis and Processing Methods of Big Data Visualization

Visualization is one of the ways to make the data meaningful. It is also a way to deal with the data. So it is important to understand the data visualization analysis. The visualization process is defined as three steps [5], as shown in "Fig. 2":
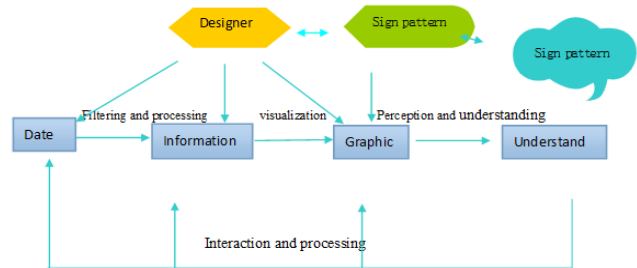

Figure 2. Visualization processing methods and steps.

The first step is to process data information: visual information is from the beginning of the data. Converting data into information consists of three steps: Firstly, collect and store relevant data. Secondly, process and convert data and delete the pre-processing errors and duplication. Thirdly, create a data structure and metadata based on meaning.

The second step is to transform the processed data into visual images: In this step, data visualization is achieved by selecting the view, rendering, and geometric transformations.

The last step is to visualize the data of the user: visual effects should be displayed to the user with the graphical symbols of the image. In this step, the user can easily understand the visual representation through human-computer interaction. During the interaction, the user can influence the visual effect by modifying, extracting and adjusting the parameters.

III. VISUALIZATION TOOLS OF BIG DATA

A. Necessary Conditions of Visualization Tools

With the development of information technology and the increased data of the industry, it is necessary to explore the new visualization tools that can collect quickly, filter, analyze, summarize and display users' necessary information, and explore the new-style data visualization tools that can update the new data in real time. Therefore, under the environment of big data, the following conditions are essential for the new-style tools [6].

- Instantaneity. In the era of the explosive growth of data, data visualization tools must be able to search, collect, sort, filter and analyze data rapidly, and update the data in time, thus it will benefit the enterprise to make decisions quickly and grasp the commercial opportunity.

- Operability. Data visualization tools should be able to develop fast, so that it can adapt to the condition that data information is complex and fickle, and it will be convenient for users.

- Colorful ways of showing. So it can offer numbers of options.
- A variety of data integration. Data comes from not only database, but also the data of some text and teamwork and so on, and it can be presented to the users in in vivid ways.

B. The Comparative Analysis of the Tools

It is to provide customers with better user experience that to analyze and display data to customers by visualization tools. After analyzing the function and situation of several popular visualization tools, this article carries through the contrastive analysis of the visual conditions, visualization, data processing power and compatibility, etc, looking forward to providing some help to users when choosing the suitable visualization tool.

Among the four visualization tools--R language, Processing, Google Chart API and D3, as a statistical analysis, drawing language and operating environment, the function of R Language is very comprehensive. It supports various computer operating systems and data mining algorithms, and supports both the data of different structures and different data sources, and meets the condition of the new data visualization roughly. The most important thing is that the visualization and the ability to handle big data of R Language are very strong, and R Language can take account of both data processing and display. Relative to the other software, R Language is the most comprehensive, but its running speed is very slow.

Processing approaches to visualization is the perspective of Art. It is the difference between itself and the other three software that Processing focus on the display of data visualization, give consideration to data visualization and creating animations, simple operation, good compatibility, and almost can be used on all platforms, but need to learn Processing Language when using this software [7].

Google Chart is an online visualization tools. It has the rich presentation, strong ability of data processing and visualization. But its compatibility is poor, only can be used in devices supporting JavaScript, and the results of visualization cannot be stored in other ways.

D3 is able to display Rich visual charts and support for a variety of data formats. It is the most suitable for displaying data on the Internet through interactive way, and with strong real-time performance. But its learning cost is higher, so we should get a deeper understanding of JavaScript before using this software. It is same as Google Chart that visualizing the data based on the JavaScript Library.

From this perspective, the R language is the most comprehensive data visualization tool among the four visual tools.

IV. THE STRATEGY AND TREND OF ANALYZING VISUALIZATION TOOLS OF BIG DATA

A. The Use of Visualization Tools of Big Data

Information visualization consists in analyzing the abstract data set, while data visualization combines the two fields and a close bond has been developed between them. When it is not feasible to be confined in observing and analyzing data, it is vital to discover the relationship of data and put it into practice showing to user in an intuitive way. Nowadays, the data visualization is applied to every aspect, and it will be more intuitive and easy to understand when enumerating its development history in the form of the time axis.

It is found that enterprise deploying data visualization is less and still in a good prospect in the survey on data visualization. It is estimated that more than half of enterprises will deploy data visualization application in the next two years. It is the deep value of analyzing the visual data that enterprise wants most on purpose. Visualization can help leaders make decision and facilitate business analyst. And it doesn't demand much on the aesthetic measure of visualization effect.

B. Strategy of Choosing Visualization Tools

With the data surge in the industry, the demand for visual products is becoming higher and higher. We should pay more attention to meet user requirements, and do not just emphasize the instantaneity, simplicity of operation, richness and universality.

The enterprises demand most for big data and visualization from comparative analysis of some popular visualization tools and data visualization investigations. And the enterprises achieve data visualization by using non-programming tools and programming tools [8]. One is to focus on show effect and the representative tool is processing. The other one is to focus on statistical and processing analysis data, and the representative tool is the R language, it also can handle the graph at the same time. There is a more perfect data visualization tool, which is represented by D3 data visualization product. It is very suitable for Internet interactive data display because of its superior data processing capacity and show effect. We can easily find the differences between these visualization tools, and hope that the analysis and comparison could provide non-professional data analysis users with assistance when choosing visual tools.

C. Development Tendency

This paper puts forward four pieces of development tendency of data visualization development based on introducing the data visualization tools and the requirements of visualization tools in big data era.

1. Data visualization not only considers data process but also the excellent visualization capacity. The value of data is putting into application. Data analysis is the key point of data processing, and the target of data processing is application, thus it is necessary for workers to show the results of data processing to the users in an intuitive and vivid way so that users could understand and accept it easily. And this is the reason why we need the excellent visualization capacity. However, most enterprises pay more attention to data and image processing, but few companies pay attention to visual effect nowadays. We believe that the effect of visual

aesthetics will be more close to user demand with technological development [9].

2. Optimize the compatibility of visualization tools. We came to a conclusion that our country's data visualization technology is still in the primary stage from domestic and foreign researches. It takes companies lager costs to analyze data. If companies could optimize the compatibility of visualization tools and improve the system scalability, there would be a variety of choices for companies on data visualization tools. Meanwhile, it can reduce the costs of data analysis and application.

3. Operating is more simple and convenient. Nowadays, domestic visualization technologies are at the stage of development. Professional data analysis is still relatively scarce, which will affect the development of big data. If everyone could analyze data, the market of data visualization will be more active in the future. Although the visualization product is rich full, most of the common users still do not know the usages, and learning how to use the tools costs heavily. Therefore it is more advantageous to enrich the performance of the tools and simplify the operation in the future.

4. Develop multi-platform of data visualization. Intelligent mobile clients have become more and more people's new favorite with the development of communication technology. The obstacles are that some visualization tools can only be used on the computer system. To some extent, it will be more convenient to develop the data visualization technology if most of the data visualization tools could be used in the mobile clients. And this is beneficial for meeting the requirements of real-time visualization tools, facilitating users at any time, satisfying the demands of displaying the internet interaction.

V. CONCLUSION

The data itself has no value, we can truly understand the revelation contained in its combination, analysis, mining and using. The benefits of data are valued by all walks with the arrival of the big data era. Government departments can make scientific decisions, the enterprises can preferably make development strategies and objectives, and individuals can comb information to realize personal value, actually, these all depend on big data analysis. It is the key to make the data analysis results shown to people in a vivid way and make it into application. Therefore, the development of big data visualization analysis tools is essential. There will be more and more people pay attention to this issue. And in the future, data visualization technology will greatly promote the development of information society.

REFERENCES

[1] Buxton B, Goldston D, Doctorow C, et al. Big data: science in the petabyte era. 2008.," *Nature*, vol. 455, pp. 1-136, 2008.

[2] J. Yang, X.Y. Wang, R.J. Bai and N. Zhu, "Status quo and development trend of scientific data analysis tools under big data environment," *Review and Comments*, vol. 38, pp. 134-144, 2015.

[3] C.X. Zheng and Y. Chen, "Visualization analysis under the big data environment," *Technology Research and Development*, vol. 20, pp. 32-34, 2013.

[4] B. Jotham, "*Computational information design,*" Massachusetts Institute of Technology, 2004.

[5] Yuri E, InfoVis Diagram, Retrieved April 5, 2016, form http://www.infovis.net/printMag.php?num=187&lang=2.

[6] R. Xie, "TOP50+5 big data visualization analysis tools," *China Internet Weekly*, vol. (17), pp. 58-59, 2014.

[7] X. Yang and D.W. Wu, "The application of R language in big data processing," *Science & Technology Information*, VOL. 23, pp. 19-20, 2013.

[8] Y. Zhang, M. Chen and X.F. Xu, "The present situation of the big data application and prospect," *Journal of Computer Research and Development*, vol. 50(S2), pp. 216-233, 2013.

[9] H. Zhang and C. Guo, "Data visualization technology trends and classification research," *Software Guide*, vol. 11, pp. 169-172, 2012.