A Comparative Study of Statistical Analysis Techniques

Zhiqiong Zou

Department of Materials Science, Jingchu Institute of Technology, Hubei, 448000, China

Keywords: Big Data; Statistical Analysis; Technical Comparison

Abstract: Due to the continuous progress of science and technology, big data has also been rapidly developed, and people's access to resources and information is becoming more and more diverse. This situation has brought unprecedented challenges to the statistical work of grassroots data in the past. It has not only ushered in new challenges, but also accordingly ushered in new opportunities. Therefore, the author studies the comparison of big data statistical analysis techniques. In the context of big data, the development of statistics should take a proactive approach, including designing more complex methods to collect data and design more unique methods to analyze data. Research shows that statistical analysis needs to combine big data ideas, which is not only the need of government service work, but also marks a new height in the transformation of government departments' work methods.

1. Introduction

The Internet technology is becoming more and more mature, and the huge and diverse data and the huge value it has produced have been paid attention to in all fields of natural science and economic market. How to effectively use data output information and explore value has become a common concern [1]. Because of the arrival of big data, China's previous grassroots data statistics work mode has undergone tremendous changes [2]. The mining and application of massive data indicates a new wave of productivity growth and the arrival of consumer surplus. In the so-called big data thinking, when conducting statistical analysis, effective integration is not an accidental opportunity. In addition, while the times develop, its own working mode is also gradually changing [3]. And the application of big data statistics is also an inevitable need of the development of the times. The development of big data platform has played an important supporting role. Throughout the wide range of modern network information, a large number of users access social information through the Internet every day [4]. In this network of big data aggregation, the working mode of "big data +" has become the inevitable result of the development of the times. At present, the statistical work of data and information in our country has gradually begun to use the information technology of computer Internet to statistics [5]. The use of computer Internet information technology for statistical data and information greatly speeds up the time of data and information statistics, improves work efficiency, and also enhances the accuracy of data and information statistics.

It is not accidental to merge the thinking mode of big data with statistical analysis, but the requirement of changing the working mode by the development of the times [6]. With the rapid development of network confidence, the advent of the era of global information, cloud computing and other rapid development, resulting in an unprecedented increase in the amount of data information around the world, data information penetrates into people's daily life and work [7]. Therefore, we can see that we are gradually entering the era of big data development. Big data expands the research field of statistics. It is of great significance to exploit the social and commercial value of big data resources by making use of the changes in the way of thinking brought about by big data, to improve and perfect statistical work and to enhance the value of statistical work [8]. In the traditional statistical analysis work of our government, the quality of statistical analysis work is low due to the lack of accuracy and completeness of data management. To this end, the combination of big data thinking mode and statistical analysis work came into being [9]. The unstructured and real-time nature of big data is different from the traditional statistical data source. Various analysis techniques suitable for various big data applications have emerged [10]. The “big data” era has arrived, and in business, economics,

  DOI: 10.25236/isaicn.2019.048

and other areas, decision making will increasingly be based on data and analytics, not based on experience and intuition.

2. Theoretical Overview of Big Data

Scientific computing systems have been developed for some complex and large-scale problems in important areas such as economic development, technological innovation, and national defense. This definition accurately reveals that the size of the data involved in big data has become so large that it is impossible for humans to extract, manage, and process all kinds of complicated and complicated information in a reasonable time through the existing mainstream software tools. It will be researched and organized, and will eventually become more effective and effective information to help enterprises make business decisions. However, under the form of big data, its existence is also the opposite. It can not only effectively break the time limit of its existence, but also make the data collection work more extensive and comprehensive. . Because the data transmission of the Internet is not limited by time, it can make the statistical analysis of data more accurate. Big data refers to any data that exceeds the processing power of a computer. Not only the workload is large and the efficiency is low, but also the accuracy of the whole data is insufficient due to the limitation of space. After years of continuous exploration and innovation by researchers, many new computing programs have been developed, which have faster and stronger computing power. Therefore, it is necessary to update the statistical mode of its data. In addition, under the background of the Internet, the statistical mode of its big data and statistical analysis are effectively combined. It can not only collect and organize the data accurately, but also display the data in the statistical analysis according to its graphics and tables.

In recent years, the research on statistical analysis technology based on big data has attracted more and more attention of scholars in China. Figure 1 shows the increasing and decreasing trend of research on such projects in China.

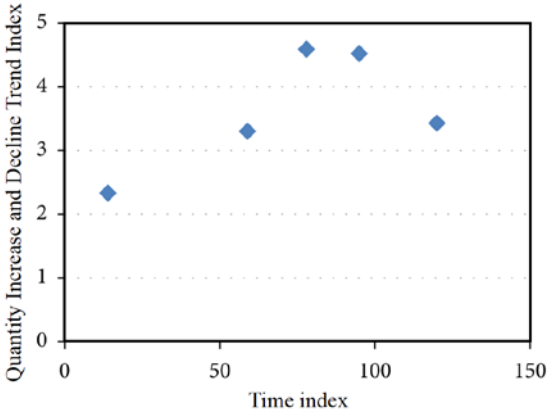

Fig.1. The Trend of Increasing and Decreasing Number of Statistical Analysis Technologies under Large Data in China in Recent Years

In the so-called traditional statistical analysis work, when government departments collect information, their collection channels are usually single, mainly for a specific place to make the corresponding collation of data. On the contrary, statistical analysis in the form of large data can not only break the limitation of space, but also make data collection more extensive and comprehensive. Regression analysis modeling is simple and fast, and it can be used to check the accuracy of residuals easily. However, the selection of specific models requires a priori knowledge of the original data. Therefore, improper selection is prone to over-fitting, and the model is sensitive to outliers when it is simple. Therefore, the development of IT technology promotes the development of statistical analysis under big data in China, and is also the basis of leading technology in the world. We can see from the development of big data volume from PB to EB, and then from EB to ZB, the growth of data volume of big data has produced a qualitative leap. Usually, it is mainly reflected in texts, etc., but with the rapid development of society. The textual approach adopted by the government

departments in China today has no way to meet the actual requirements of rapid social development. Moreover, the data transmission of the Internet is not limited by time, which makes the statistical analysis of data more efficient and accurate.

3. Application of Statistical Analysis Technology under Big Data

Because the decision analysis is uncertain in the process of visualization, when the analysis results are tested and predicted, the robustness of the analysis is emphasized. For traditional statistical work, due to the geographical restrictions, its data is also subject to more serious space constraints. At present, the application of big data thinking is the comprehensiveness of the statistical analysis work of the Chinese government, and the statistics of various data. The ability to use the network to integrate data from various organizations in our society fully guarantees the universality of data on the macro level. Data statistics focus on statistical analysis and modeling. Generally, models are built through statistics and probability to mine data content. Due to the huge amount of data in the process of big data sorting, the complexity of data types, and the efficiency of data processing. Therefore, the problems reflected are only reflected in the local, so there is a certain amount of data for this basis, and the accuracy of data can not be guaranteed accordingly. It can shorten the statistical period greatly, reduce the error of statistical analysis, and then let the government make decisions based on effective and accurate data, achieve "prophecy" and predict the future economic development form and trend through big data.

Traditional national population distribution statistics is a huge project. The government divides each province into groups one by one, and invests a lot of manpower and financial resources to conduct a "visiting" census. But after statistical analysis and utilization of big data, according to the network platform of big data, we can make corresponding statistics for each netizen' s disease, so that we can generate a graph by ourselves. It is also able to effectively break the limitations of traditional time and make real-time detection of medical conditions in each region. Traditional statistical methods, such as grouping, summarizing, compiling statistical tables and drawing statistical charts, are no longer competent for the statistical work of large data information. They will gradually be replaced by the two links of auditing and storing data. The statistical analysis under big data has received extensive attention in the era of social information development. Applying big data can improve the efficiency of data information statistics and promote the good development of big data. The complexity of big data, such as chaotic, noisy, mixed, internal and unstable, is objective, but we can also find some hidden relationship patterns and knowledge from the noisy data. Small samples are more valuable. It is also possible to make a correct pre-judgment for each disease for its corresponding disease situation and number of visits, so that prevention of different diseases can be achieved.

4. Conclusions

As a young and active research area facing application, big data statistical analysis technology is being driven by solving practical problems. The statistical analysis of big data has gradually replaced the original statistical methods of grassroots data, which can be completed faster and better, and statistical analysis of a large amount of data information. Clear the connection and feedback relationship between the two basic links of statistical design data collection and mining data. In short, for the rapid development of society, the work of statistical analysis also needs to be able to keep pace with the development of the times. Establish a multi-disciplinary work style, based on the concept of "disciplinary group", integrate natural science, social science, and data science, and the leading role of substantive science in domain applications should receive sufficient attention. As far as possible, break the time and space restrictions, so that our statistical analysis work is based on the thinking of big data, the establishment of big data platform, so that government departments through the efficiency and quality of work can be guaranteed. Therefore, we should strengthen the research of statistical analysis of large data, so that it can better meet the needs of rapid development of the social era. To make statistics education more close to the practice of data analysis, by employing industry

experts to teach, vivid and practical case teaching to guide students to understand the application and needs of statistics in practice, and strengthen the practical teaching of statistics. Make the personnel training mode more actively adapt to the needs of social development for high-level statistical personnel.

Acknowledgement

In this paper, the research was sponsored by the Jingchu institute of technology school-level scientific research project (youth project) QN 201608, Feasibility study on waste classification in Jingmen city.

References

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

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

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

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

[5] Hu H, Wen Y, Chua T S, et al. Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. IEEE Access, 2014, 2:652-687.

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

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

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

[9] Kitchin, Rob. The real-time city? Big data and smart urbanism. GeoJournal, 2014, 79(1):1-14.

[10] Richtárik, Peter, Taká? Martin. Parallel coordinate descent methods for big data optimization. Mathematical Programming, 2016, 156(1-2):433-484.