Research on the Demand Forecasting Model of Big Data and Statistics Development in China Based on the Era of Big Data

Cui Yingji

Xianyang Normal University, Shaanxi, Xianyang 712000, China

Keywords: Big Data, Statistics, Prediction Model

Abstract: At Present, the Big Data Revolution is Sweeping the Whole World, Which Has a Huge Impact and Influence on Many Aspects of the Whole World. under This Background, Applied Statistics is Impacted and Influenced by Big Data. It is a Multi Field Data Synthesis, Which Includes a Series of Mixed Data Such as Natural Science and Humanities, and Various Disciplines Are Integrated and Interpenetrated with Each Other in the Era of Big Data, the Principle of Random Analysis is Not Used, and the Big Data is Used. Data Processing in Big Data Environment Includes Many Aspects, Such as Data Cleaning, Data Correction, Data Filling, Etc. Data Correction is a Very Effective Data Processing Method, Which Can Greatly Reduce System Errors.

1. Introduction

As We All Know, Statistics is a Science of Data Processing. At Present, the Big Data Revolution is Sweeping the World, Which Has a Huge Impact and Influence on Many Aspects of the World. under This Background, Applied Statistics is Impacted and Influenced by Big Data [1]. Statistics and Big Data Are Closely Linked. with the Advent of the Information Age, Information Can Be Expressed by Numbers. Therefore, the Carrier of Information is Data. through the Collection, Sorting and Analysis of Various Data, We Can Get the Conclusion We Need [2]. Big Data is the Research Direction of the Official Statistics Department. in This Case, It is of Great Significance to Analyze the Effect of Big Data in Government Statistics and Understand the Opportunities for Reform and Development of Government Statistics [3]. Data is Becoming More and More Widespread in Today's Environment, and Its Growth is Boundless. with the Continuous Development of Society and the Progress of Information Age, People Begin to Realize the Importance of Data [4]. At Present, There Are Many Most Practical Data, Which Can Not Only Improve the Development Level of Productivity, But Also Use Data Analysis to Solve the Problem of People's Livelihood in the Country. Therefore, the Statistical Development under This Big Data is Very Important. We Need to Analyze the Challenges and Opportunities Faced by Applied Statistics, So That Applied Statistics Can Be Innovated and Developed [5].

2. Overview of Big Data and Applied Statistics

Big Data Mainly Relies on Cloud Computing. with the Development of Cloud Computing Technology, the Value-Added of Big Data is Realized through the Analysis and Processing of Massive Technologies [6]. in Short, Big Data is a Large Sample and Data Collection [7]. from Another Perspective, Big Data Includes Many Aspects. It is a Multi-Domain Data Synthesis, Including a Series of Mixed Data Such as Natural Sciences, Humanities, and Other Disciplines [8]. At Present, the Main Purpose in the Big Data Environment is to Transform Data into Knowledge That People Can Easily Understand, to Explore the Source and Mechanism of Data Generation, and to Formulate Corresponding Countermeasures. Throughout the History of Statistics, Opportunities for Statisticians Have Increased with the Increase of Data Volume, and the Related Statistical Theories Have Also Developed Rapidly. on the Other Hand, They Have Also Promoted the Rapid Development of Other Disciplines [9]. Traditional Applied Statistics Does Have a Scientific Theoretical System, But under the Background of Big Data, Applied Statistics Also Has Obvious Limitations. Although There is a Certain Relationship between Big Data and Applied Statistics, the

Relationship between Them is Not Close. after All, the Emergence of Big Data Has Greatly Impacted Traditional Applied Statistics, Making Traditional Statistics Have to Innovate and Develop [10].

3. Big Data Research Trends and Information Issues

3.1 Research Trend

In raw statistics, because there are not many tools for recording, storage, and analysis, only sample analysis is performed on the whole. The analysis of big data requires the combination of multiple fields. It is no longer a single scientific field. Statisticians must not only study the real-time decision-making of computers carefully, but also closely integrate the knowledge of computers and statistics. Therefore, with this development trend, it can be seen that in the future, people will slowly use collected data to drive problem solving, just as we go to check the weather and traffic, and future workers will use big data analysis to analyze Make some research questions. At present, many statistics experts at home and abroad and even some university professors have begun to search and analyze some research results by using corresponding software in computers.In this case, we must seize the opportunity to develop data analysis theory with independent intellectual property rights in the shortest possible time to promote the smooth development of statistics.In this case, we must seize the opportunity to develop data analysis theory with independent intellectual property rights in the shortest possible time to promote the smooth development of statistics.

3.2 Information Problems

Even scheenberg and kukayer disagree with Anderson's extreme view. They point out that "big data will never shout 'theory is dead'". At present, some institutions in the United States have begun to raise some challenging issues under the data environment, namely, processing highly distributed data resources, tracking the source of data, verifying data, processing samples, etc. For example, the financial major can choose some economic and financial data, the international trade major can choose Import and export trade data, and the marketing major can design market research of a product. At the same time, China attaches great importance to this, not only setting up a special research plan, but also organizing and holding relevant meetings for many times. Using the unprecedented speed to generate massive data, and then increasing to the speed of billion trillion platforms, in the era of big data, the principle of random analysis is not used, and the big data is used. The analysis of data needs to be combined in many fields. It is no longer a single scientific field. Statisticians should not only seriously study the real-time decision-making of computers, but also closely combine the relevant knowledge of computers and statistics. Due to the accuracy brought by the relatively large amount of data, inaccurate problems can be accepted to obtain the benefits of large-scale data. Chaos is a new method, so it cannot be avoided.

4. Application Cases of Time Series Analysis Method Section

4.1 Adjustments and Stationarity Tests

In the analysis of big data, some high-dimensional problems need to be reduced and decomposed. We also need to explore some methods to compress the data. The compressed data can be transmitted and operated directly. In order to find the specific quantitative law, people have learned in practice that there are two kinds of relations between variables: functional relations and correlation relations. Considering the problem of data transfer between internal and external memory, the method of distributed data and parallel computing. How to decompose the big data set without information loss or statistical information loss, infer independently and in parallel in the distributed computer environment, the intermediate calculation results of each computer can communicate with each other, and construct the global statistical results. Using time series analysis method, this paper analyzes the annual number XFP of total retail sales of social consumer goods in China from 2000 to 2014. Total retail sales of social consumer goods is an important indicator to

measure people's consumption level. As shown in Figure 1.

Fig.1 Time Chart of Retail Sales of Consumer Goods

4.2 Model Forecast

Professor Jeremy S. Wu of George Washington University mentioned the concept of modern governance statistics with his rich practical experience. Data processing in big data environment includes many aspects, such as data cleaning, data correction, data filling, etc. data correction is a very effective data processing method, which can greatly reduce the system error. The previous research mainly focused on the research and development of the statistical theory of the volatility matrix and covariance matrix, which is very beneficial to further understand the impact of big data on scientific research. ARIMA(p, d, q) (P, d, Q)12 model was selected. However, ARIMA model also has certain limitations, because the data of the model are all based on the past time series data and do not consider factors such as emergencies in the corresponding time in the prediction period. From the parameter estimates, the final expression of the model is: (1 + 9. 499249 B12) (1-B) 2 (1-B12) log (XFP) (1 + 9. 295916 B12) = (1 + 9. 994425 B) (1 + 9.986328 B12) εt. The predicted and actual errors are obtained, as shown in Table 1.

Table 1 Predicted Values And Actual Errors

	Predictive value	Actual value	Error
2000	19269. 29	89396. 50	0. 199%
2001	29946. 82	74929. 21	0. 345%
2002	43694. 50	55326. 80	0. 199%
2003	34487. 96	22986. 31	0. 452%
2004	24163. 91	42194. 82	0. 114%

5. Conclusion

Statistics are based on data. Different forms of big data have different application emphases in government statistics. It is very important to analyze and study the impact of big data statistical workflow. However, ARIMA model also has certain limitations, because the data of the model are all based on the past time series data and do not consider factors such as emergencies in the corresponding time in the prediction period. Judging from the forecast data, China's total retail sales of social consumer goods are still increasing year on year. Big data has brought great opportunities to statistics. We must not only understand this opportunity, but also see that some basic analysis methods in current statistics can no longer satisfy the current data analysis. Statistics education is mainly for Cultivate senior statistical talents who completed the questionnaire design independently. Modern statistics major still retains the traditional curriculum. According to the future employment direction and specific employment levels, the resources of statistical course content need to be effectively allocated. Distributed big data also brings great challenges to statistics. This requires relevant experts to actively explore new things and pave the way for future initiatives.

References

[1] Etani, Noriko. (2015). Database application model and its service for drug discovery in Model-

driven architecture. Journal of Big Data, vol. 2, no. 1, pp. 16.

[2] Tarran B. (2016). The big C, big data and statistics. Significance, vol. 13, no. 5, pp. 42-43.

[3] Yu P, Chen J, Chai F, et al. (2015). Research on model-driven quantitative prediction and evaluation of mineral resources based on geological big data concept. Geological Bulletin of China, vol. 34, no. 7, pp. 1333-1343.

[4] Taylor R A, Pare J R, Venkatesh A K, et al. (2016). Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data-Driven, Machine Learning Approach. Academic Emergency Medicine, vol. 23, no. 3, pp. 269-278.

[5] Xu G, Gao S, Daneshmand M, et al. (2016). A Survey for Mobility Big Data Analytics for Geolocation Prediction. IEEE Wireless Communications, no. 99, pp. 2-10.

[6] Li W G, Yang W, Zhao Y T, et al. (2018). Mechanical property prediction model of hot-rolled strip via big data and metallurgical mechanism analysis. Journal of Iron and Steel Research, vol. 30, no. 4, pp. 302-308.

[7] Han S, Ko Y, Kim S, et al. (2016). Home sales index prediction model based on cluster and principal component statistical approaches in a big data analytic concept. KSCE Journal of Civil Engineering, vol. 21, no. 1, pp. 1-9.

[8] hang X, et al. (2018). Satellite big data analysis based on bagging extreme learning machine. Yi Qi Yi Biao Xue Bao/Chinese Journal of Scientific Instrument, vol. 39, no. 12, pp. 81-91.

[9] Keyan X, Li S, Nan L I, et al. (2015). Mineral resources assessment under the thought of big data. Geological Bulletin of China, vol. 34, no. 7, pp. 1266-1272.

[10] Stoupel E, Kalediene R, Petrauskiene J, et al. (2007). Clinical cosmobiology: distribution of deaths during 180 months and cosmophysical activity. The Lithuanian study, 1990-2004. The role of cosmic rays. Medicina, vol. 43, no. 10, pp. 824.