

Mining Data Value of Group Prediction Wisdom in the Era of Big Data

BAO Xixi^{[a],*}

^[a]School of Business Administration, South China University of Technology, Guangzhou, China.

*Corresponding author.

Received 29 May 2018; accepted 27 August 2018

Published online 26 September 2018

Abstract

When making predictions about an uncertain event, the individual's ability is always limited and biased. Therefore, in order to predict the outcome of the event, it is necessary to collect prediction data from different individuals to make predictions. The aggregated results of individuals are always better than most individuals, and the performance of the group is called group wisdom. However, due to information technology and space constraints, the data collects is limited under normal circumstances, time-consuming and costly, and the group's ability to predict is limited. In big data era today, we can collect massive amounts of data that meet the forecasting requirements more quickly, aggregate these data, and through certain data processing methods, we can get the solution of the group, which is generated in this environment. The prediction of the data is often more accurate, and the predictive power of the group is greatly increased.

Key words: Big data; Prediction ability; Group wisdom; Value measurement

Bao, X. X. (2018). Mining Data Value of Group Prediction Wisdom in the Era of Big Data. *Management Science and Engineering*, 12(3), 57-61. Available from: URL: <http://www.cscanada.net/index.php/mse/article/view/10795> DOI: <http://dx.doi.org/10.3968/10795>

INTRODUCTION

Data is a logical induction of objective things and is used to represent raw material of objective

things(Anagnostopoulos, Zeadally, & Exposito, 2016). By obtaining data that different individuals give their own judgments about the same event, we use different analysis methods to process the data according to the type of the event, and a solution of the group can be obtained. The economist Herbert Simon believes that individuals are "limited rational animals"(Asir, Appavu, & Jebamalar, 2016), they are not able to give the most correct solution in face of most forecasting events, we often gather a certain group of people forecast the event to get the results of the data prediction and realize the value of the data. With the advent of the era of big data, the way people access data has become more diverse, and the speed of data generation has become faster and faster. IDC is a pioneer in the study of big data and its impact, defining big data in the 2011 report(Garcia et al., 2013) "Big data technology describes a new era of technology and systems, designed to be diversified from a large scale. The value of data extracted through high-speed capture, discovery, and analysis techniques in data." The arrival of the era of big data provides more data support for predicting events, and it also greatly reduces the cost of data acquisition(Gautam, Prajapati, Dabhi, & Chaudhary, 2015). In such an environment, data generation, diffusion, and proper processing of these data will increase the value of the data. For a group, the arrival of the era of big data has broken the limitation of space; the number of data acquisition has greatly increased, so the predictive ability of the group has also undergone major changes.

1. THE GROUP PREDICTION METHOD OF LIMITED DATA

We often make our own judgments about uncertain future events, such as predicting the future weather, trends in a stock's future time period, etc. Individuals use their existing knowledge to make future uncertain events predictions. However, in most cases, due to the

limited rationality of the individual, it is usually not possible to make an appropriate prediction. Therefore, the group prediction is used to generate the data, and the corresponding data analysis means is used to obtain the answer of the question and realize the value added of the data. This method has been widely recognized and applied.

Traditionally groups are defined as social collections with mutual awareness and potential interactions, and are relatively small and structured organizations (Hashem et al., 2015). The structure and cohesiveness of the group gives them an advantage in completing their tasks. Groups in the traditional sense are prone to predictive bias due to independence from independence or group polarization due to herding effects. In the book, Surowiecki (James, 2004) identified three requirements for groups to meet diversity, distraction, and independence in generating opinions. This allows groups to express different opinions when making predictions and to be able to profit from relative independence. The group referred to in this context refers to a collection of individuals who are able to express opinions independently.

In most cases, when we need to predict an uncertain event, in order to gain the wisdom of the group and help us make better decisions, we often convene a certain

number of individuals to make predictions in a relatively limited space. Out of the solution that the individual considers to be the most appropriate, and in the process of giving results to each individual, it is necessary to remind the individual not to exchange discussions to ensure independence between individuals. Under the premise of ensuring the diversity of the group, when each individual gives data after the prediction, the error items of the extreme cases will cancel out each other, so the average estimate is similar to a good guess (Erlenkamp, Becker, & Sciences, 2013). We can see that in group prediction, everyone uses their existing knowledge to give their own judgments on problems that need to be predicted. They may only be able to give a part of future results because of bounded rationality, but aggregate these results together. It can eliminate most of the errors and get a relatively correct prediction. For example, when letting people predict the event "Will it rain tomorrow?", some individuals may observe the clouds to make predictions, some individuals may observe animal behaviors, and some individuals may refer to farmers' yearbooks and so on. We can see that any method itself is inaccurate. But in general, their accuracy is expected to be high.

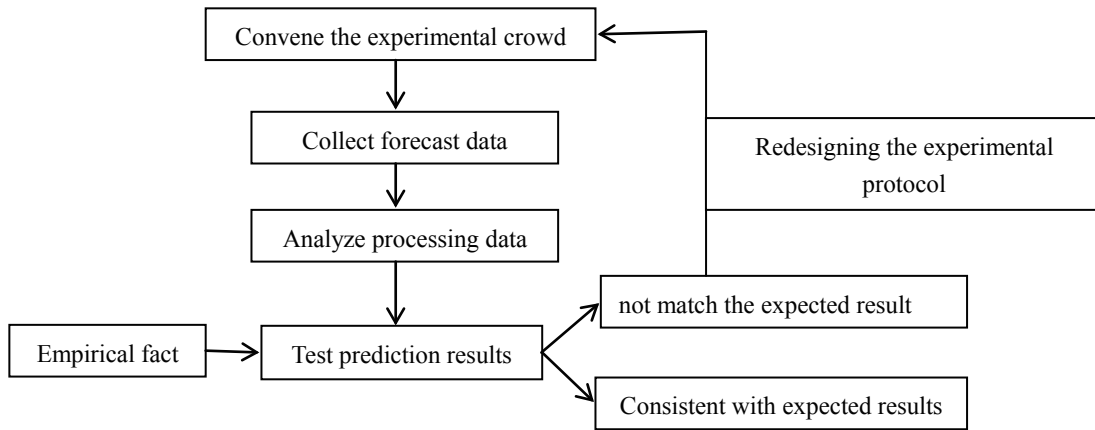


Figure 1
The Main Processes in Traditional Data Forecasting

Group predictions are expected to have better performance, and group diversity is critical to predicting outcomes. PAGE (Page, 2008) proposed the diversity prediction theorem to illustrate the importance of group diversity for the accuracy of group prediction results.

$$PD = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Predictive diversity (PD) represents the total squared difference between individual guesses and average guesses. It reflects the average level of the individual. n indicates the number of people involved in the forecast event.

$$IE = \frac{1}{n} \sum_{i=1}^n (x_i - x_{true})^2$$

Individual errors (IE) aggregate the squared error of all individuals. Therefore, it represents the average precision of individual guesses.

$$CE = (\bar{x} - x_{true})^2$$

The group error (CE) represents the square error of the population prediction and is used to measure the accuracy of the population prediction.

This theorem shows that population error (CE) = individual error (IE) - predictive diversity (PD), we can reduce group errors by reducing individual errors

(increasing individual expertise) or increasing group diversity.

Using the diversity prediction theorem, Wagner C(Page, 2008)used experimental and simulation methods to conclude that group abilities can be consistent with experts in the case of sufficient diversity of groups. Therefore, when making event predictions, we can use the data provided by individuals in the group to help me make more accurate predictions. In the past, due to information technology and space constraints, it was a time-consuming and labor-intensive task to gather enough individuals to provide prediction data for an event at the same time and space when using individual data to assist in the prediction. In such an environment, how to maintain individual independence is also a problem. When the number of individuals involved in the prediction reaches a sufficient number, the diversity increases relatively, but it becomes more and more difficult to maintain the independence between individuals. With the advent of the era of big data, the continuous development of information technology, the acquisition of data gradually breaks the limitation of space, and the ability of group prediction has changed dramatically.

2. THE DATA PREDICTION ABILITY IN THE ERA OF BIG DATA

Big data projections based on large and diverse data carriers and cloud computing, sensing solutions and distributed processing technologies “have led to new breakthroughs in productivity, technological strength and social structure” (Bolón-Canedo, Sánchez-Marroño, & Alonso-Betanzos, 2016). Big data is predictive of all aspects of life, including more accurate predictions of sports events, stock market movements, user behavior, and disease outbreaks. Big data provides greater accuracy for data prediction due to its large volume, variety, and fast flow rate.

Big data is influenced by the timeliness, statistical characteristics and sources of the data used compared to previous data predictions(Long, Zhao, Chen, & Software,

2014). The use of big data prediction enables real-time updating, monitoring, and processing of the required data, which greatly reduces the time difference between the predicted data and the predicted results. Due to the limitation of time and space, traditional data prediction mostly adopts the method of field experiment collection. Since the data collected by this method has hysteresis, the timeliness according to the prediction data cannot compare with the big data prediction.

In the era of big data forecasting, the statistical characteristics of data are “full data”, and almost all the data we need can be obtained through technologies such as the Internet. In the past data collection process, sampling methods were generally used to collect data, and the amount of data obtained was very limited(Demirkan & Delen, 2013). According to the data diversity prediction theorem, in order to reduce group errors as much as possible, it is necessary and effective to increase individual diversity. Because of the “full data” feature of the big data era, we are able to quickly collect large amounts of data for a certain event that needs to be predicted through social networks, which provides the possibility of data diversity.

The rapid development of information technology is more abundant in data sources. In the traditional data collection process, the individuals are generally single, and the diversity of data is also limited. The difference is that big data predictions take advantages of many data resources that have not been used before. For example, advances in data acquisition technology have allowed us to use Internet search behaviors such as Baidu and to obtain new data such as text messages from Weibo. In the past, these data were unstructured data, which could not be directly used for analysis and prediction, but advances in big data technology can convert these data into structured data. Although these new types of data are not as accurate as traditional data, they greatly increase the diversity of the population and make the predictions change dramatically.

Today, the big data forecasting process has also changed dramatically:

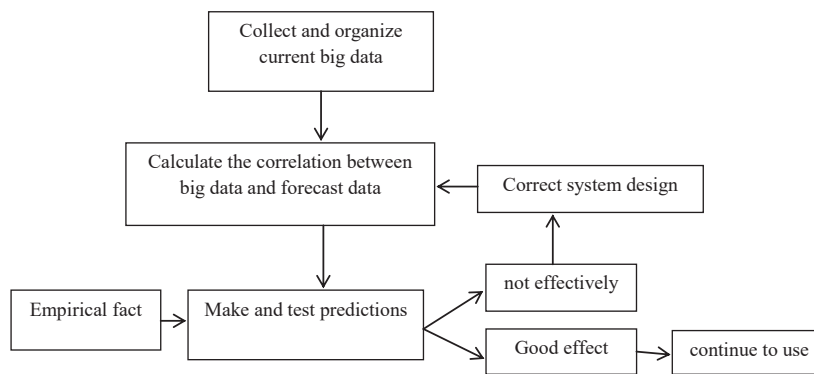


Figure 2
The Big Data Forecasting Process Now

In the era of big data, group prediction ability breaks the limitation of time and space, group prediction is improved in accuracy and convenience, and the prediction method is improved. It can be seen that once the prediction result is biased, it can be corrected in the system for storing data. By changing the correlation coefficient, we can find the data that meets the requirements. In the past data prediction, we not only need to redesign the experiment, but also need to call the people to collect the data, which will bring a lot of work.

3. THE ACTUAL CASE ANALYSIS OF BIG DATA FORECASTING

The stock market is usually in a turbulent uncertainty. In the past, people used the power of the group to buy stocks. They usually gathered together to make stock purchase decisions based on information such as the trend of the stocks some time ago, and predict which stocks will be in the future. With the development of science and technology, people began to use big data to analyze stock trends. Research by the Warwick School of Business in the United Kingdom and the Department of Physics at Boston University in the United States found that the financial keywords that users searched through

Google may guide the direction of the financial market. From the big data of people’s attention, we can see some stock trends and use these big data. The corresponding investment strategy income is as high as 326%. Prior to this, some experts tried to predict stock market volatility through the sentiment of public platforms like Twitter. At present, many hedge funds in the United States have invested in big data technology and have gained a lot. Some funds in China also use big data to predict the harvest. Although this model is similar to the traditional quantitative investment, it also relies on the model, but the data variables in the model show geometric growth. On the basis of the original financial structured data, the unstructured social speech, geographic information, satellite monitoring, etc. Data, and quantify these unstructured data, so that the model can absorb more data for prediction, greatly improving the accuracy of data prediction.

Due to the turmoil in the stock market, there are many uncertainties. There is no uniform standard for its trend prediction model. In the past literature, the multi-level decomposition of wavelet transform is used to process stock price data, bringing accidental factors and macro-mutation factors into the nerve. Network training is a typical stock forecasting model.

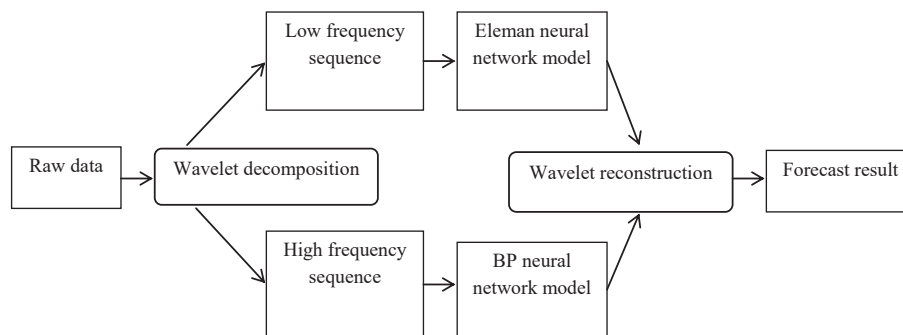


Figure 3 Forecasting Model

4. CHALLENGES FACED BY GROUP FORECASTING IN THE ERA OF BIG DATA

However, while the accompanying large amount of data arrives, various problems are also generated, making the group prediction results unsatisfactory.

4.1 Privacy Security Is Difficult to Protect

“There is no dignity in places where there is no privacy.” Anonymity is an important guarantee of dignity. However, with the advent of the era of big data, the big data forecasting system separates and identifies real data and intentions in huge amounts of structured and unstructured data, and gradually achieves “any form of privacy is as long as there is reasonable commercial motivation

to drive the data mining process, it is impossible to algorithmically”(Wu, Zhu, Wu, Ding, & Engineering, 2013). While the convenience of obtaining data is improved, the privacy of users has been infringed more and more seriously. Under the condition that they have not obtained their own permission, a large amount of relevant data information is leaked, and the majority of users are caught at the expense of privacy disclosure to obtain convenient personality. The embarrassing situation of the service has impaired dignity.

4.2 The Result of the Pre-Judgment Is Limited

The vast array of data and mining-related analysis tools have made the breadth and precision of data prediction applications unprecedented, including many uses of data to predict stock market, event outcomes, user behavior

and more. Gradually formed a smart society, people are more and more respected for such a society. As data-driven decision-making models in areas such as residents' lives, business trade, and political activities become more mature, we are entering an era of "just need to know what, not need to know why"(Raghupathi, Raghupathi, & Systems, 2014). People are increasingly relying on such a pre-judgment result to bring their own behavioral guidance, and internal innovation and independent thinking are gradually becoming less. This huge expansion reduces the courage of individual free exploration and deprives enterprises of the opportunity of independent innovation. . Limiting individuals to various computational models surrounded by data, forbidding individuals to seek a free soul.

4.3 Information Monopoly Is Unfair

"Disclosure is usually not a critical piece of confidential data but ordinary information." Big data prediction can integrate fragmented information, and the data information is related, and the future trends contained in the data are accurately understood. The use of predictive information for subsequent goal planning completely overturns the traditional model. Many companies are also aware that data information offers more possibilities for improving user retention and precise marketing, and is conducive to creating business value. As the forecasting market continues to intensify and the forecasting party's ability to control continues to increase, the Internet giants who monopolize massive high-value metadata use excuses for security and privacy considerations, and quietly complete the gorgeous turn from free to charge, from cheap to expensive. These commercial tycoons destroy the industry environment of orderly competition and monopolize the data information in their hands. Although big data predictions lead humans infinitely close to the ultimate dream of controlling the future, the uneven distribution of information and monopoly will break the fairness of society as a whole.

CONCLUSION

The arrival of the era of big data is irreversible. We must only make full use of the role of big data to make it easier for us to bring convenience to our learning and life. We also need to face up to the problems and challenges that big data brings to us. Positive confrontation and resolution can make the development of technology work for us.

REFERENCES

- Anagnostopoulos, I., Zeadally, S., & Exposito, E. (2016). Handling big data: Research challenges and future directions. *The Journal of Supercomputing*, 72(4), 1-23.
- Asir, D., Appavu, S., & Jebamalar, E. (2016). Literature review on feature selection methods for high-dimensional data. *International Journal of Computer Applications*, 136(1).
- Bolón-Canedo, V., Sánchez-Marroño, N., & Alonso-Betanzos, A. (2016). Feature selection for high-dimensional data. *Progress in Artificial Intelligence*, 5(2), 1-11.
- Demirkan, H., & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems*, 55(1), 412-421.
- Erlenkamp, S., & Becker, F. (2013). A new approach to the use of gestures in aphasia rehabilitation: stimulating spontaneous gesturing and body language. *Procedia - Social and Behavioral Sciences*, 94(1), 265-266.
- Garcia, C., Roberti, W., Schreiber, D., Paludo, M., Malucelli, A., & Reinehr, S. (2013). *The implementation experience of an advanced service repository for supporting service-oriented architecture*. Paper presented at the Informatica.
- Gautam, J. V., Prajapati, H. B., Dabhi, V. K., & Chaudhary, S. (2015). *A survey on job scheduling algorithms in big data processing*. Paper presented at the IEEE International Conference on Electrical, Computer and Communication Technologies.
- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing. *Information Systems*, 47(C), 98-115.
- Long, S. Q., Zhao, Y. L., & Chen, W. (2014). A three-phase energy-saving strategy for cloud storage systems. *Journal of Systems and Software*, 87(1), 38-47.
- Page, S. E. (2008). The difference: How the power of diversity creates better groups, firms, schools, and societies. *J. I. C.*, 88(4), 270-278.
- Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1), 3.
- Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. Doubleday.
- Wu, X. D., Zhu, X. Q., Wu, G. Q., & Ding, W. (2013). Data Mining with Big Data. *IEEE Transactions on Knowledge and Data Engineering*, 26(1), 97-107.