ATLANTIS PRESS
Advances in Computer Science Research, volume 74
2nd International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2017)

The Simple Analysis of the Application of Data Mining in Mathematical Modeling

Jirong Lv[1], Shihu Wang[2], Haicui Lv[1]

[1]Haojing College of Shaanxi University of Science & Technology, Faculty of Science; Xi'an 710000 China

[2]Xixian New Area Big Data; Xi'an 710000 China

Keywords: Data mining ,Mathematical modeling, Application.

Abstract. The 21th century has been a high informative era, and computer science and artificial intelligence have quickly developed and penetrated into all walks of life day by day. Same, database technology also has been widely applied, and varieties of database system also get deep exploration. However, faced with the overwhelming data, database system can only make simple operation, for example, management, and it cannot make the deeper level analysis. Based on this background, data mining technology comes into being to meet the demand of handling data in the current times. This paper will also start from the perspective of data mining technology and simply analyze the application in mathematical modeling.

1. Introduction

The 21th century has been a high informative era, and computer science and artificial intelligence have quickly developed and penetrated into all walks of life day by day. In the period of 12th Five-year plan, the information infrastructure of all domains are constantly imperfect and information construction level has increased steadily. Its achievement generally expresses in the following several aspects:

(1) The significant status of information gets the high attention

Material determines consciousness, and consciousness reversely influences material. Before the 12th five-year plan, most people think that the main reason that hindering the information development is the negative act of leaders and lacking finance support, etc. But during the period of 12th five-year plan, this condition has greater improvement. In the current stage, most leaders begin to pay more attention to the information construction, and they have realized that information is the significant method to comprehensively deepen reform and enhance the independent innovation ability, and all the related sectors also increase the input capital of information project.

(2) Information technology has been gradually applied to all domains

Abstract the information development experience all over the world, and our domestic information development experienced the four great development process of point, line, surface and solid. The application of all levels and all domains are gradually perfected, and information management is gradually strengthening and make the support layout of "all-round, multi-level and three dimensional".

(3) Information has began to transform to strategy drive

At the beginning of information construction, due to the defect of cognition, the development mode of information construction is mainly through the technology drive development. As time goes by, information has deeply integrated into all the domains and formed the new pattern that regards the information technology as its support and the coordination around all the walks of life and five principles as its target.

However, there also are some giant challenge, many problems and shortcomings in the information construction: the deficiency of the whole structure design, the ignorance of software application and analysis of data and system, etc. Therefore, this paper made the simple discussion and analysis with the problem of ignoring data analysis, with the background of data mining technology and the example of mathematical modeling.

2. The Development of Data Mining Technology

Along with the wide application of database technology, varieties of database system get deep exploration and large amount of data push into the computer with the overwhelming attitude, and people can make analysis and research based on these data in the computer. However, database system has its congenital defect—-it cannot make deep-level analysis and usage aiming at the hidden data information, thus, people hope that there will be one kind of technology that assists handling many complex data information and abstracts the useful message to apply to the practice. Under this background, data mining technology comes into being. And the main purpose of data mining technology is to abstract useful information and hidden relation, find the potential relation among all the data and make deep-level analysis through making deep mining analysis with the information in the database. According to the related report review of *Technology Review* in American MIT, that "data mining" technology will become one of the ten great emerging technologies that bring human being significant change.

The development history of data mining is as shown in table1.

Table 1. The development history of data mining

Time	Field	Achievement
The end of 18th century	Statistics	Bayes theorem
The early 20th century	Statistics	Regression analysis
1920s	Statistics	Maximum likelihood estimation
1940s	Artificial intelligence	Neural network
1950s	Artificial intelligence	Perceptron
1950s	Statistics	Re-sampling
1960s	Artificial intelligence	Machine learning
1960s	Statistics	Decision tree
1960s	Information retrieval	Similarity measure
1960s	Database	Relational data model
1970s	Artificial intelligence	Genetic algorithm
1970s	Statistics	The estimation of incomplete data, etc.
1980s	Artificial intelligence	Kohonen self-organizing map
1980s	Artificial intelligence	ID3 algorithm in decision tree
1980s	Artificial intelligence	Bayesian network
1990s	Database	Association rules algorithm、search engine, etc.
1990s	Database	Data warehouse

3. Mathematical Modeling from the Perspective of Data Mining Technology

The mathematical model is defined by choosing a mathematical language, for a particular object of study or a mathematical relationship, that can be used to characterize the mathematical relationship and structure of a particular study object by means of mathematical notation. There are commonly two kinds of mathematical modeling methods, which are statistical analysis and mechanism analysis. The mechanism analysis refers to analyze its inherent relationship, to find the necessary mechanism,

and to establish a mathematical model can be expressed with practical significance based on the characteristics of the reality of the object. The statistical analysis refers to make detailed records of experimental data for the object to be studied, to find its inherent trends through a large number of data comparison analysis, and then to build a mathematical model which may reflect the rule of data variation by mathematical language. However, the established mathematical model will change in accordance with the changes of data, due to its own real-time variability of statistical analysis. For example, an assessment on the degree of water contamination in a given area depends on the initial sample data collected in the particular river. The different rivers have their particularities and need to be progressively adjusted on the initial sample data so as to apply to the specific situation of a certain river. This reflects the dynamic nature of the statistical analysis. Due to this feature, statistical analysis has been more widely used in mathematical modeling process.

With the rapid development of information technology and Internet technology, computers are more applied to various fields. At the same time, in the face of amount of data in each application area, it is necessary to adopt more scientific data mining technology, which can establish the mathematical model closer to research object, to explore the hidden relationship behind the data.

The essence of data mining technology is mathematical modeling, and the mathematical modeling through data mining technology is similar to that of statistical analysis. Therefore, data mining technology has been regarded as modern statistics. The modeling by statistical analysis method is based on classical mathematical statistic. In the current stage, in the face of the increasing amount of data, the traditional statistical analysis method is obviously unable to meet the current need. The data mining technology, belonging to modern statistical analysis method, under certain conditions can be more scientific to solve practical problems, thus avoiding its limitations brought about by the traditional statistical analysis method. To sum up, data mining technology is not only the inheritance of traditional statistical analysis method, but also the development and extension of traditional statistical analysis method.

4. The Application Practice of Data Mining Establishing Mathematical Modeling

There are two important parts through data mining to establish the mathematical modeling, and the first one is to transform the data of hidden and unclear data relationship to the mathematical problem; the second one is to transform unclear data relationship to obvious data relationship through mathematical modeling. Mathematical modeling is the key point of data mining, and data mining is a rather popular research hot project in the current stage. Next, choose the example of machine process and make the specific explanation and analysis.

4.1 Raise of Problem

In the process of machine process, the error produced by human factor is inevitable. Thus, in the finished parts that have been produced, it will find that the quality of some parts is qualified, but the quality of some parts is not up to the standard. Aiming at these defective goods, how many goods can be used after the process and recover? Aiming at this question, establish the mathematical modeling, analyze the reason of defective goods and raise the repair suggestions through making data mining, applying the probability theory and statistics principle.

4.2 Preparation of Data

This paper applied the example of ream processing a group of holes, and the specific size parameters are that the diameter is 20mm, and the upper and lower deviation are 0.01mm respectively. The reaming selects the reamer whose diameter size is 20mm, and the usage of cutting remains unchanged. The measured hole size after the reaming and process is near the normal distribution, and its X_i value respectively is: 20.60, 20.080, 20.085, 20.062, 20.070 ， …, 20.068(unit mm).

4.3 Data Analysis

4.3.1 Compute the Average Size of Hole

x, and the formula is as following,

$$x = \frac{\sum_{i=1}^{n} xi}{n} \tag{1}$$

Acquire the value based on the prepared data： x=20.080mm。

4.3.2 Calculate the Constant System Error

The meaning of constant system error is that comparing with the theory size, the actual measured offset of arithmetic mean value can be expressed as:

$$\Delta s = x - x_m \tag{2}$$

Among, x_m—— the ideal size of workpiece.

It can calculate that the constant system error of this example is 0.080mm, and the constant system error usually can be eliminated or reduced through making proper adjustment to the craft system.

4.3.3 Calculate Random Error

Based on the background of part processing error near the normal distribution, the dispersion degree of a group of parts size usually is applied to measure the size of random error, namely:

$$\Delta_R = 6\sigma = 6\sqrt{\sum_{i=1}^{n}(xi - x)^2 / n} \tag{3}$$

It can be calculated from the above formula：

$$\Delta_R = 6 \times 0.0067 = 0.04\text{mm}$$

What random error reflects is making the average size as the middle value and the positive and negative symmetric distribution. Its size determines the shape of normal distribution curve.

4.3.4 Conform the Process Capability Coefficient

Use the process capability coefficient C_P to express the size of process capability. The meaning of process capability coefficient is shown as the figure 1.

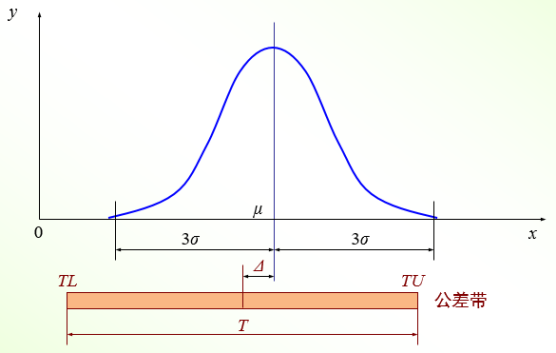

Figure 1. The figure of process capability coefficient meaning

When the part processing error is relatively conformed with normal distribution, the calculation of process capability coefficient can refer to the following formula:

$$C_P = T / 6\sigma \tag{4}$$

What is needed to be noticed: when $C_P > 1$, it shows that the process capability is enough; otherwise, process capability is not enough.

4.3.5 Calculate Unqualified Rate

Based on the actual condition of machine process, the calculation of unqualified rate not only needs to consider the constant system error, but also cannot ignore the random error. Taking the above hole process requirement as an example, if after the improvement and adjustment, then processing this group of parts can get:

x=20.005mm, σ=0.0067mm

Through the change, it can be acquired that:

$$\frac{x_u - x}{\sigma} = \frac{20.01 - 20.005}{0.0067} = 0.75 \tag{5}$$

$$\frac{x_l - x}{\sigma} = \frac{20.005 - 19.99}{0.0067} = 2.24 \tag{6}$$

According to the standard normal distribution probability table in the formula (5), the unqualified rate of oversizing can be solved as:

$$0.5 - 0.2730 = 0.2270$$

According to the standard normal distribution probability table in the formula (6), the unqualified rate of undersized can be solved as: ：

$$0.5 - 0.4874 = 0.0126$$

Therefore, the total unqualified rate is: $0.2270 + 0.0126 = 0.2396$, namely, 23.96%.

4.4 Put Forward the Solution to Solve Problems

The size distribution figure and tolerance zone requirement are shown as the figure 2

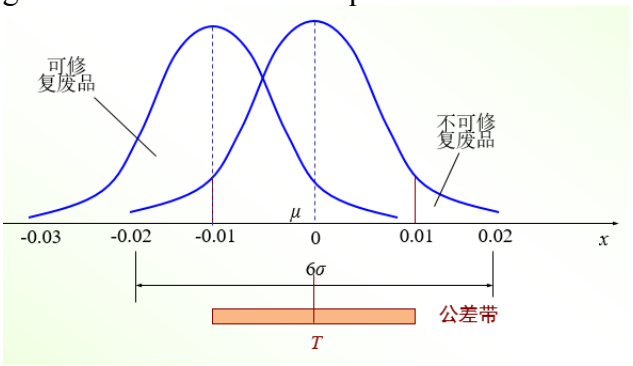

Figure 2. Mathematical modeling of size distribution

According to the mathematical modeling of size distribution shown in the figure 2, there are several effective methods to reduce the unqualified rate as following:

(1)Eliminate the constant system error

The process size measure of holes selected by this paper is $\phi 20 \pm 0.01$mm, namely, the tolerance of this hole is: T=0.02mm. According to the above measured data and the later mathematical calculation, the ream random error of this example can be acquired that: $6\sigma = 6 \times 0.067 = 0.04$mm. Thus, the constant system error of this hole has reached 0.08mm, namely, the unqualified rte of the reaming process hole is 100%.

In order to reduce the unqualified rate of part process, the elimination of constant system error should be considered at first, and combining the above measured data, lessening the reamer diameter 0.08mm, and then it can realize the goal of zero constant system error. But random error is still bigger than the tolerance of parts, and the relation in the mathematical modeling can be expressed as: $6\sigma > T$. Therefore, there will still be the condition of unqualified goods.

Calculate as the formula (5) and (6), the total unqualified rate can be obtained as 13.36%.

(2)Enhance the process capacity

Based on the problems reflected from the established mathematical modeling, it needs to adopt the specific machine process craft method to make improvement, for example, select the more accurate machine tool can effectively reduce the random error of craft, 6σ, so that enhancing the process capacity and reducing the unqualified rate.

(3)Increase the repairability of unqualified goods

According to the size distribution curve shown in the figure 2, if lessening the reamer diameter 0.01mm on the basis of eliminating the constant system error, it is equal to making the distribution center adjust to being smaller 0.01mm than the tolerance zone center, shown in the left one distribution curve in the figure 2. After such adjustment, the probability of occurring the unrepairable goods is zero, and the repairable unqualified rate known presently is 50%.

5. Conclusion

Data mining is widely been regarded by the research scholars as one of the ten great emerging technologies in the next several decades, and this paper simply analyzed the development history of data mining, the mathematical modeling from the perspective of data mining technology and the construction of application practice of data modeling through data mining. In the prior of 13th Five Year Plan, data mining technology still is a strong potential stock, and it is bound to be more widely applied to all the fields, so the nation should master the development trend of information construction, comprehensively deepen the reform, eliminate the traditional mindset, enhance the independent innovation ability and try the best to improve the shortcomings in the information construction, for example, increasing the application of software, making deep analysis of data mining. Though the simple analysis of this paper, the writer hopes to provide some references to the research fellow in the process of studying the related project.

Acknowledgments

Special research project in Shaanxi province department of education：“Data mining is applied in colleges and universities”(Serial number:15JK2024).

References

[1]. Ping Xue, Liping Chen, Zhipeng Jiang. The application of mathematical statistics method reduces the process error of inner hole of eccentric bearing [J]. Haerbin bearing, 2013,(02):51-52+62.
[2]. Dedong Zhang. The increment mining algorithm research faced with incomplete data [D]. Ludong University, 2010.
[3]. Nan Wu. The sequential Pattern mining algorithm research based on cluster sub region [D]. Hefei University of Technology, 2009.
[4]. Yan Pan. The analysis of application of data mining in the enterprise custom relationship management [J]. Technology information, 2009, (07):148+166.
[5]. Dongmei Cui. The statistic of application research of data mining in CRM [D]. Jiangsu University, 2007.
[6]. Xiang Wang. The development current situation and trend analysis of data mining [J]. Technology Outlook, 2017, (07):13.
[7]. Huizhong Wang, Anqun Peng. The research current situation and development trend of data mining [J]. Industry and Mine Automation, 2011, (02):29-32.
[8]. Chunhua Zhang, Yang Wang. The application and development trend of data mining technology [J]. Modern Information, 2003, (04):47-48+50.
[9]. Xaiobin Huang, Aizhen Deng. The deepening of modern information management———the development trend of data mining and knowledge finding [J]. New Technology of Library and Information Service, 2003, (04):1-3+16.