

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Quality Improvement using Data Mining in Manufacturing Processes

Shu-guang He<sup>1</sup>, Zhen He<sup>1</sup>, G. Alan Wang<sup>2</sup> and Li Li<sup>3</sup>

<sup>1</sup>*School of Management, Tianjin University, Tianjin,*

<sup>2</sup>*Pamplin College of Business, Virginia Polytechnic Institute and State University,*

<sup>3</sup>*School of Electronics and Information Engineering, Tianjin Professional College, Tianjin,*

<sup>1,3</sup>*P.R.China*

<sup>2</sup>*U.S.A.*

## 1. Introduction

Nowadays, manufacturing enterprises have to stay competitive in order to survive the competition in the global market. Quality, cost and cycle time are considered as decisive factors when a manufacturing enterprise competes against its peers. Among them, quality is viewed as the more critical for getting long-term competitive advantages. The development of information technology and sensor technology has enabled large-scale data collection when monitoring the manufacturing processes. Those data could be potentially useful when learning patterns and knowledge for the purpose of quality improvement in manufacturing processes. However, due to the large amount of data, it can be difficult to discover the knowledge hidden in the data without proper tools.

Data mining provides a set of techniques to study patterns in data “that can be sought automatically, identified, validated, and used for prediction” (Witten and Frank 2005). Typical data mining techniques include clustering, association rule mining, classification, and regression. In recent years data mining began to be applied to quality diagnosis and quality improvement in complicated manufacturing processes, such as semiconductor manufacturing and steel making. It has become an emerging topic in the field of quality engineering. Andrew Kusiak (2001) used a decision tree algorithm to identify the cause of soldering defects on circuit board. The rules derived from the decision tree greatly simplified the process of quality diagnosis. Shao-Chuang Hsu (2007) and Chen-Fu Chien (2006 and 2007) demonstrated the use of data mining on semiconductor yield improvement. Data mining has also been applied to product development process (Bakesh Menon, 2004) and assembly lines (Sébastien Gebus, 2007). Some researchers combined data mining and traditional statistical methods and applied to quality improvement. Examples are the use of MSPC (multivariate statistical control charts) and neural networks in detergent-making company (Seyed Taghi Akhavan Niaki, 2005; Tai-Yue Wang, 2002), the combination of automated decision system and six sigma in the General Electric financial Assurance businesses (Angie Patterson, 2005), the combined use of decision tree and SPC with data from Holmes and Mergen (Ruey-Shiang Guh, 2008), the use of SVR (support vector regression) and control charts (Ben Khediri ISSam, 2008), the use of ANN (artificial neural

Source: Data Mining and Knowledge Discovery in Real Life Applications, Book edited by: Julio Ponce and Adem Karahoca, ISBN 978-3-902613-53-0, pp. 438, February 2009, I-Tech, Vienna, Austria

network), SA ( simulated annealing) and Taguchi experiment design (Hsu-Hwa Chang, 2008). Giovanni C Porzio(2003) has presented a methods for visually mining off-line data with combination of ANN and  $T^2$  control chart and to identify the assignable variation automatically.

Although techniques are available to learn numerous patterns and knowledge hidden in mass data, it can be difficult to identify those patterns that can be directly applicable. Kaidi Zhao (2005) developed a visualization tool, named as Opportunity Map, which can help identify useful and actionable knowledge quickly. Mu-Chen Chen (2007) proposed a method that ranks the rules learned by data mining techniques using DEA (Data Envelopment Analysis).

This work focuses on using data mining for quality improvement in manufacturing processes. The chapter is organized as follows. In section 2, we introduce a knowledge-based continuous quality improvement model in comparison with DMAIC for six-sigma. In section 3, we explain parameter optimization, quality diagnosis and service data analysis for quality improvement. In section 4, we propose a system framework for quality improvement in manufacturing processes. Finally, section 5 concludes the chapter and highlights the areas for further work.

**2. Knowledge-based continuous quality improvement in manufacturing processes**

Continuous quality improvement is an important concept in the field of manufacturing quality management. DMAIC (Define-Measure-Analyze-Improve-Control) for six-sigma is the most commonly used model of continuous quality improvement. Fig.1 illustrates the processes of DMAIC.

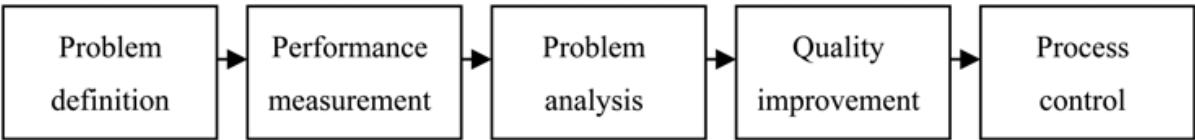


Fig. 1. The processes of DMAIC

We can observe from Fig. 1 that DMAIC is a problem driven approach. The entire process begins with locating a problem. However, in a complicated manufacturing process, such as semiconductor manufacturing and steelmaking, it may not be easy to identify and define a proper problem. Moreover, in high-speed manufacturing processes quality problems must be quickly identified and eliminated. Otherwise it may lead to a large amount of loss in both cost and productivity. Therefore, we propose a knowledge-based quality improvement model (see Fig. 2). Different from DMAIC, this model is a goal-driven process. The central idea of the knowledge-based quality improvement mode is to mine the mass data collected from a manufacturing process using automated data mining techniques. The goal is to improve the quality performance of manufacturing processes by quickly identifying and eliminating quality problems.

In the knowledge-based quality improvement model, the first step is to define the goal. The goal here may be defect elimination, efficiency improvement, or yield improvement. Data mining is used to analyze the quality related data for finding the knowledge between the goal and the factors such as machinery parameters, operators, and material vendors. After

the knowledge has been verified, opportunities of quality improvement can be identified using the knowledge and patterns learned by data mining techniques. The scope of the problem can be broad across different phases of a manufacturing process. In the following sections, we explained how to apply the model to parameter optimization, quality diagnosis and service data analysis.

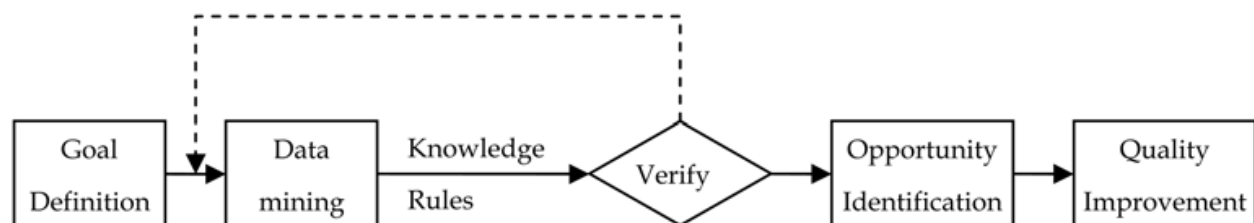


Fig. 2. The knowledge-based quality improvement mode

### 3. Using data mining to improve quality in manufacturing processes

There are three important processes in manufacturing enterprises, including design, manufacturing and service. Data mining techniques can be used in all three processes to improve the quality of manufacturing processes and final products.

#### 3.1 Parameter optimization with data mining

Parameter optimization is critical to quality improvement in manufacturing processes. There are various methods proposed for parameter optimization. DOE (Design Of Experiments) is considered as the most important one (Myers R H, 1985). RSM (Response Surface Methodology) is a technique developed based on DOE and has been most widely used. The main idea of DOE and RSM is to build a function between the inputs (factors) and outputs (responses) of a process. The parameters of the function can be optimized using mathematical methods.

Although DOE is valuable for parameter optimization, there are also drawbacks. DOE is a static method in which the parameters are optimized in a certain situation and the parameters cannot be adjusted during the real manufacturing process that is often dynamic in all kinds of situations. In addition the DOE model only considers a small number of factors with the constriction of cost.

To overcome the problems of DOE, we propose a technique that combines DOE and data mining. DOE is used for parameter optimization for certain static modes. Data mining is employed to analyze actual quality measures with different parameters settings. It aims to learn the patterns between the parameters settings and quality outcome. The pattern can then be used to dynamically adjust the parameters during the manufacturing process so that undesired outcome can be avoided. In the following, we use an injection modeling process in semiconductor assembly as an example to illustrate the process.

In the molding process of LED packaging, there can be many factors affecting the quality of final products. We chose four most important factors with screening experiments. The factors are explained below.

1. Mold temperature ( $^{\circ}\text{C}$ ). The mold temperature will affect the shape of final products, which is a critical quality characteristic. If the temperature is too high, the number of products with defects will increase.

Standard Order	Run Order	Mold temperature(°C)	Warm-up temperature(°C)	Screw pressure(N)	Screw duration(s)	Defect rate (%)
16	1	0	1	0	-1	2.25
21	2	-1	0	0	-1	2.24
9	3	0	0	-1	-1	2.96
24	4	0	-1	0	1	4.95
8	5	0	1	1	0	1.40
13	6	-1	1	0	0	10.05
4	7	0	0	1	1	4.78
12	8	0	0	-1	1	5.50
20	9	0	0	0	0	1.16
7	10	0	-1	0	-1	2.62
26	11	0	0	0	0	0.69
1	12	0	-1	1	0	7.17
22	13	1	-1	0	0	4.67
2	14	1	0	0	-1	4.17
17	15	-1	0	-1	0	10.62
6	16	-1	-1	0	0	11.20
11	17	0	1	-1	0	7.47
5	18	0	-1	-1	0	1.67
10	19	-1	0	0	1	2.15
19	20	0	0	1	-1	6.85
18	21	1	1	0	0	0.76
3	22	1	0	-1	0	2.35
23	23	1	0	1	0	0.69
15	24	0	1	0	1	2.24
25	25	-1	0	1	0	2.77
14	26	0	0	0	0	3.96

Table 1. The RSM table for parameter optimization

- 2. Warm-up temperature (°C). In injection molding process, epoxy compound, which is a kind of material, must be softened before being used. The warming up temperature is to ensure the material be softened in a proper temperature. The warm-up temperature is also very important to the process.
- 3. Screw pressure (Newton). If the screw pressure is too high, the wire attached to the dice and the PCB will be skewed. On the other hand, if the screw pressure is too low, the epoxy compound will be cooled and also lead to quality problems.
- 4. Screw duration (Second.). Screw duration is the time intervals between the beginning and the end of molding processes. The epoxy compound will be cooled with too long screw duration. Screw duration is also a vital parameter of this process.

Experimental data are collected in a RSM table (Table 1) and analyzed using the Minitab© software. And the optimized manufacturing parameters are found using statistical methods. The mold temperature is 142.3 °C, the warm-up temperature is 65.7 °C, the screw pressure is 22.4 Newton and the screw duration is 40 seconds.

It must be pointed out that these optimized parameters are obtained in a controlled experiment. The parameters may not achieve the optimal results in real manufacturing

conditions. Other uncontrollable factors may also have an effect on the quality of final products. Thus the process should be monitored and optimized continuously using real quality related data. Data mining can be used to serve that purpose by automatically obtaining knowledge and patterns about the manufacturing process.

We collect 1000 records randomly from the molding processes. Each record consists of values of the four factors. The records were classified based on their defect rates. Those records whose defect rates were higher than 3.0% were categorized into a negative class and labeled with *L*. All other records belonged to a positive class and were labeled with *H*. We use the C5.0 decision tree algorithm to analyze the data. Fig. 3 shows the result of the decision tree. We can observe that the right branch of the tree achieves a better performance than the left branch. The path indicated by the circled nodes provides us guidance for parameter optimization.

The decision tree shown in Fig. 3 can also be presented by a set of rules. We have identified five rules. Two of them lead to the classification of the positive class (*H*) where the other three predict the negative class (*L*).

- 1) Rule 1 for *H* (144; 0.973)  
If Mold\_Tem > 135.345 and  
Warm\_Tem > 66.762 and  
Pressure > 21.235  
Then *H*.
- 2) Rule 2 for *H* (127; 0.961)  
If Mold\_Tem ≤ 135.746 and  
Warm\_Tem > 62.106 and  
Warm\_Tem ≤ 66.762 and  
Pressure > 20.733  
Then *H*.
- 3) Rule 1 for *L* (44; 0.848)  
If Mold\_Tem ≤ 135.345 and  
Warm\_Tem > 66.761  
Then *L*.
- 4) Rule 2 for *L* (390; 0.625)  
If Pressure ≤ 21.235  
Then *L*.
- 5) Rule 3 for *L* (790; 0.598)  
If Warm\_Tem ≤ 66.762  
Then *L*.

Although a decision tree does not provide precisely optimized parameters like what the RSM method does, decision tree can analyze a very large amount of quality related data with noise which is still a constriction in DOE. The combination of these two methods can provide more feasible results than using DOE only.

### 3.2 Quality diagnosis with data mining

During a manufacturing process, product quality can be affected by two types of variation: random variations and assignable variations. Random variations are caused by the intrinsic characteristics of a manufacturing process and cannot be eliminated completely. Assignable variations are often produced by a flawed manufacturing setup that involves machinery,



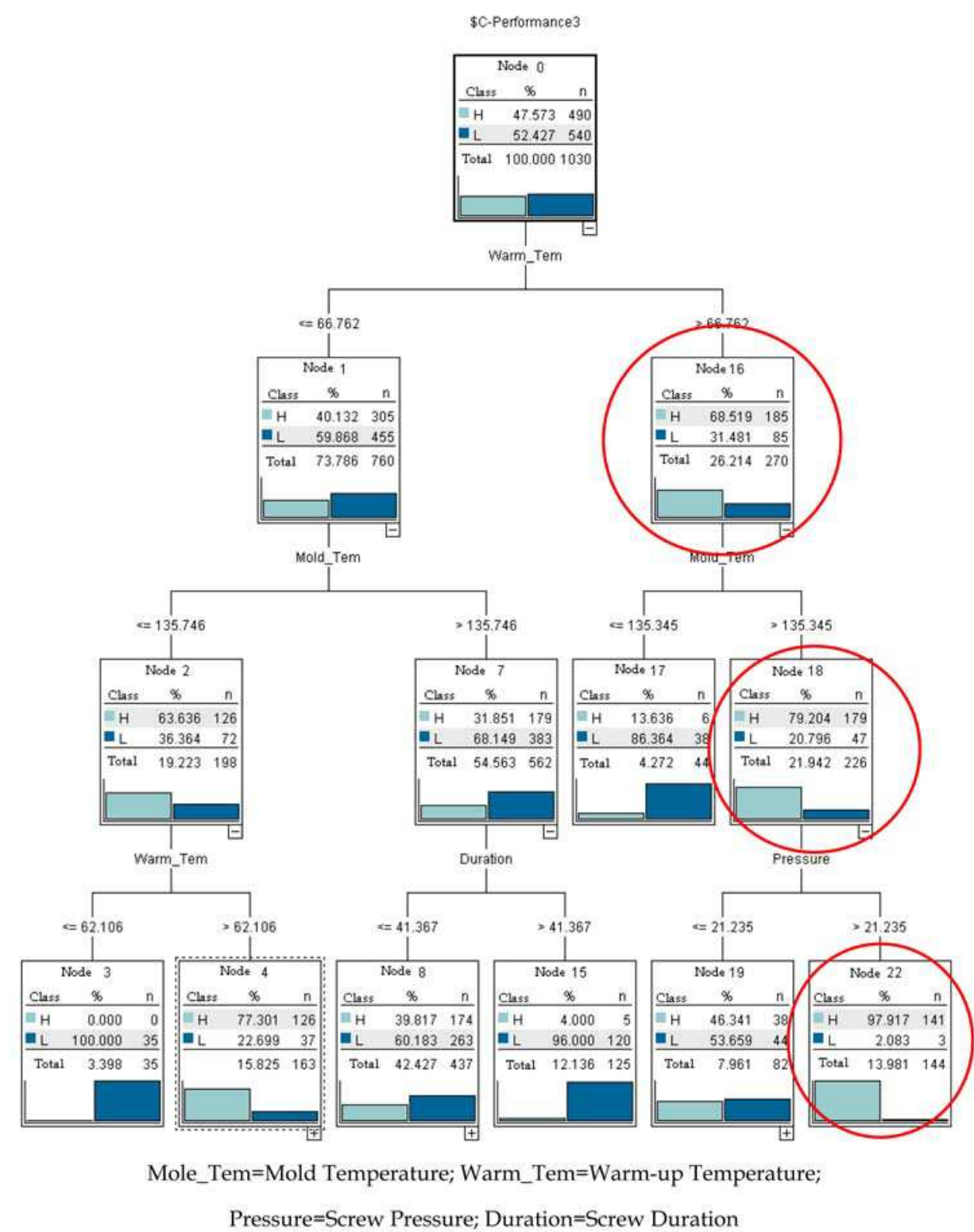


Fig. 3. The decision tree of manufacturing data analysis

operators, materials and environment. Assignable variations are predictable and such can be eliminated as soon as we identify them.

SPC (Statistical Process Control) and MSPC (Multivariate SPC) are the most widely used tools in manufacturing for finding assignable variations. Although they can effectively

detect assignable variations in manufacturing processes, they give no clue to identifying the root causes of the assignable variations. Data mining techniques can again be employed in this case to provide insights for quality diagnosis.

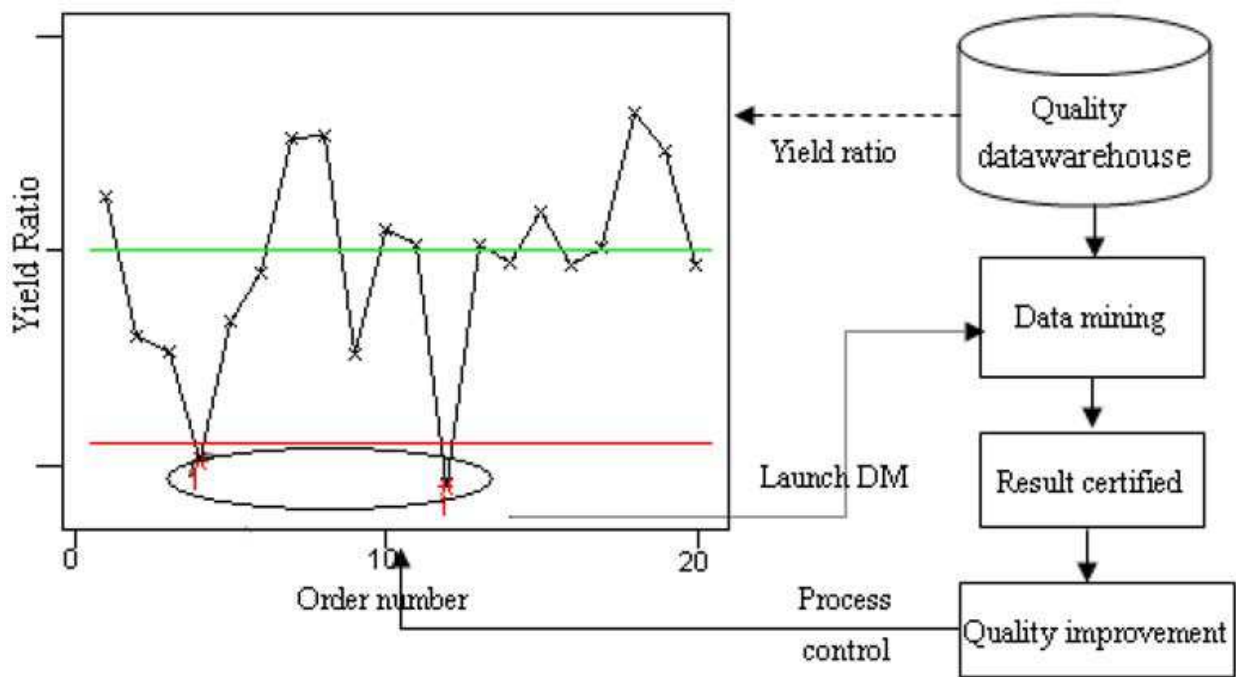


Fig. 4. The combination of SPC and data mining

In the model shown in Fig. 4, the yield ratio of a product is defined as the index of the quality performance of a manufacturing process. The chart on the left is a control chart. When the chat shows alarming signals, i.e., points located beyond the control limit, the data mining process will be engaged. Data related to quality are stored in a data warehouse. Data mining techniques such as a decision tree and association rule mining can be applied to the data to identify the causes of the alarming signals.

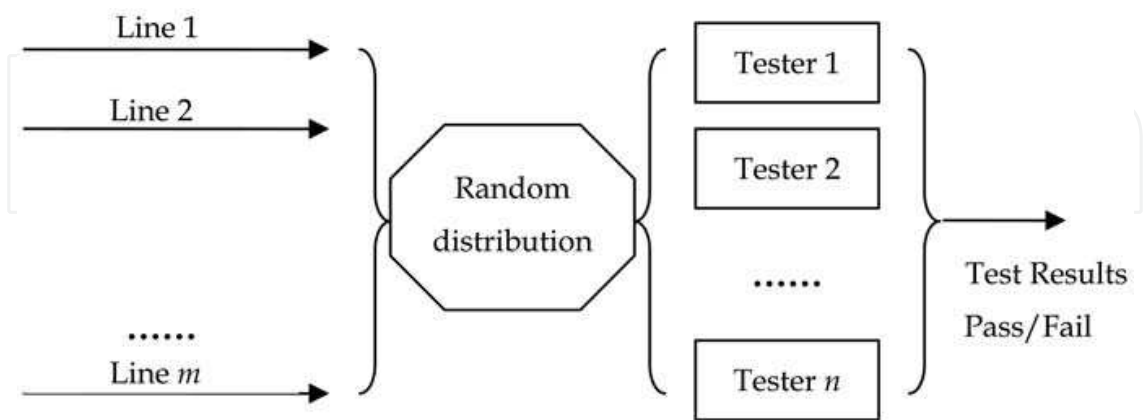


Fig. 5. A mobile phone assembly line model

We use a mobile phone assembly line as an example to explain the process described above. Electronic test is an important process in a mobile phone assembly lien. During electronic test, all the products will be tested using an electronic test instrument. Quality related data



can be collected off the test instrument and stored in a database. We can then analyze the data using data mining in order to find ways to improve the quality of the manufacturing process. A mobile phone assembly line model is presented in Fig. 5. There are  $m$  assembly lines while products assembled in these lines are randomly distributed to  $n$  testers. The test results are classified into two classes: **Pass** and **Fail**.

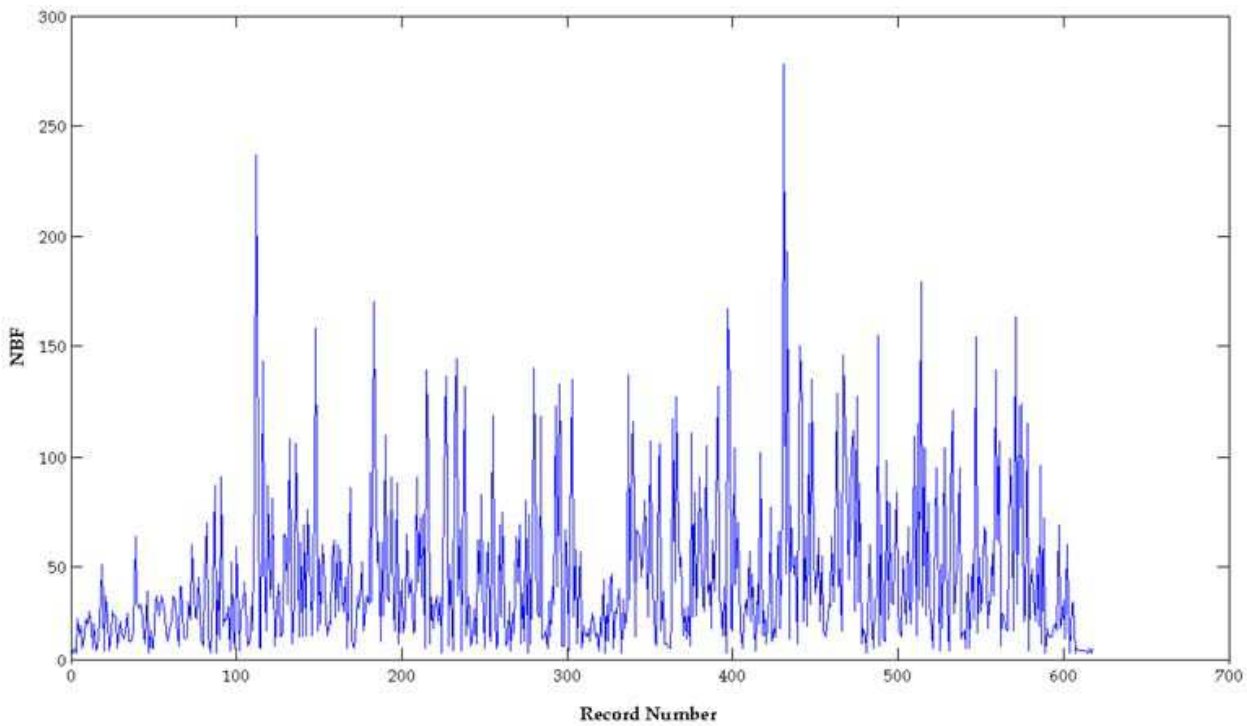


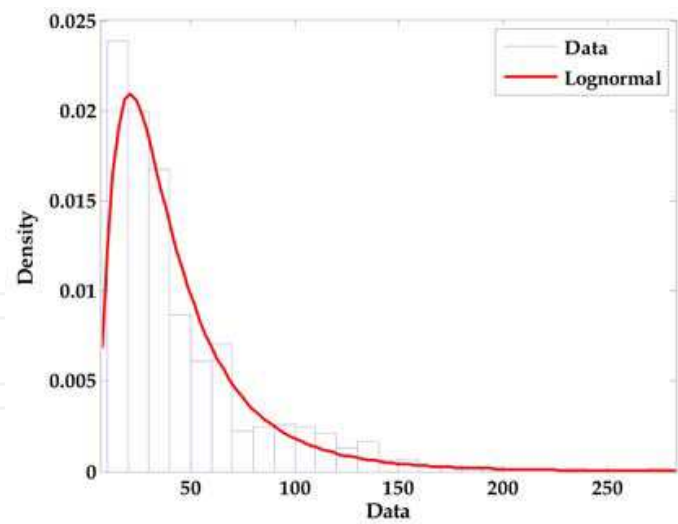
Fig. 6. The distribution of NBF values

In a mobile phone assembly line, there is a very important process named electronic test. In this process, all the products will be tested with an electronic test instrument. And the data are collected and stored in a database. We can analyze the data with data mining to find the opportunity to improve the quality of manufacturing process. The mobile phone assembly line is illustrated in fig. 5. There are  $m$  assembly lines and the products assembled in these lines are randomly distributed to  $n$  testers. The test results are classified into two classes, marked with **Pass** and **Fail**. Furthermore, there are also 27 numerical attributes in the test results, but they are not concerned in this study.

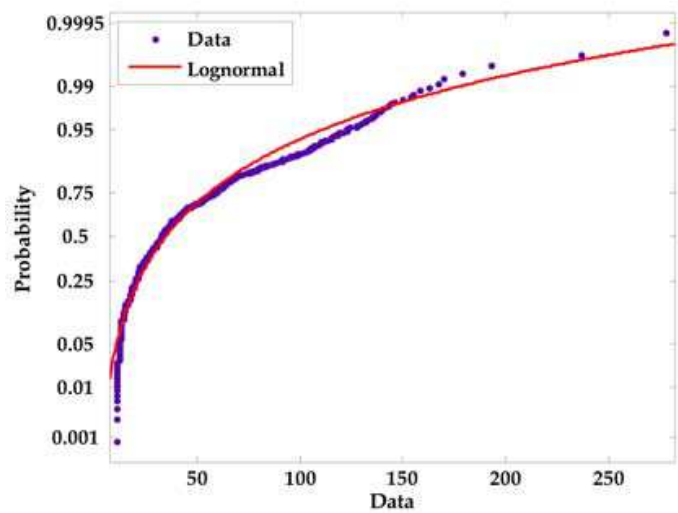
Testers are assumed to work under the same condition with same precision. The number between failures (NBF), i.e., the number of products that passed the test between two successive failure tests, is used to measure the quality performance of the manufacturing process. A larger NBF denotes a more stable manufacturing process.

In the experiment we collected 62,592 records from the electronic test process of the mobile phone assembly line. The NBF values of the records are shown in Fig. 6 that provides a concise description of the data.

We used Matlab software to analyze the distribution of NBF values. Fig. 7 shows that the data can fit to a lognormal distribution. Fig. 7(a) shows the probability density function plot while Fig. 7(b) draws the probability plot of the NBF values and the lognormal distribution curve. The parameter estimation of the lognormal distribution is presented in Table 6.



(a) The PDF plot



(b) The probability plot

Fig. 6. Distribution fitting of the NBF data

Parameter	Estimate	Standard Error
Mean	3.54087	0.0284031
Sigma	0.70609	0.0201084

Table 6. The estimated parameters of the lognormal distribution

The standard error of both the mean and the standard deviation of the NBF data are really small. Therefore we can say that the NBF data can approximately fit a lognormal distribution with mean=3.54087 and sigma=0.70609. With the estimated parameters, we can get the 0.0027 fractile, which is a vital parameter in SPC, of the lognormal distribution is 4.78. This means that if the NBF is smaller than 5, we can make the decision that there are assignable variations in the process and the expected probability of an error decision is less

than 0.27%. Thus we can set the CL=5 as a control line of the SPC chart for the NBF data. If there are points distributed beyond the control limits, the data mining methods will be used for quality diagnosis.

Next an association rule mining tool is used to find the root cause of the assignable variations of the process. Attributes such as tester ID, assembly line number, and test results are supplied into our association rule mining tool with test results being the consequent variable. We used the Apriori algorithm as the association rules analysis method. The minimum antecedent support was set to 0.65% while the minimum rule confidence was 80%. The obtained association rules are presented in Table 7.

Consequent	Antecedent	Instances	Support %	Confidence %	Rule ID
Fail	Line = L15	3,810	6.087	99.948	1
	Line = L20	4,209	6.725	99.81	2
	Line = L20 and Tester_ID = T106	440	0.703	100	3
	Line = L20 and Tester_ID = T21	429	0.685	100	4
	Line = L20 and Tester_ID = T31	431	0.689	100	5
	Line = L20 and Tester_ID = T73	411	0.657	100	6
	Line = L20 and Tester_ID = T97	425	0.679	98.824	7
	Line = L20 and Tester_ID = T35	410	0.655	100	8
	Line = L20 and Tester_ID = T63	423	0.676	100	9
	Line = L33	3,216	5.138	100	10
	Line = L33 and Tester_ID = T5	409	0.653	100	11
	Line = L36	4,297	6.865	100	12
	Line = L36 and Tester_ID = T10	432	0.69	100	13
	Line = L36 and Tester_ID = T14	503	0.804	100	14
	Line = L36 and Tester_ID = T57	455	0.727	100	15
	Line = L36 and Tester_ID = T70	516	0.824	100	16
	Line = L36 and Tester_ID = T75	416	0.665	100	17
	Line = L36 and Tester_ID = T81	438	0.7	100	18
	Line = L36 and Tester_ID = T89	499	0.797	100	19
	Line = L36 and Tester_ID = T104	472	0.754	100	20
	Tester_ID = T104	585	0.935	83.077	21

Table 7. The obtained association rules

Some conclusions drawn from the association rules are listed as below.

1. There were 3,810 mobile phones assembled by line 15, 99.948% of which failed the test (Rule 1).

2.

There were 4,209 mobile phones assembled by line 20. 99.81% of which failed the test (Rule 2). Rules 3- extended Rule 2 by considering each different tester.
3.

There were 3,216 mobile phones assembled by line 33, all of which failed the test (Rules 10 and 11).
4.

There were 4,297 mobile phones assembled by line 36, all of which failed the test (Rules 12-20).
5.

There were 585 products tested by tester 104, 83.077% of which failed the test (Rule 21).
- In all the 62,592 products, there were 16,905 products that failed the test. The percentage of the failed tests was 27%. The failed products assembled by lines 15, 20, 33, 36 or tested by tester 104 are summed up to 16,117. That means nearly 95% of the products that failed in the tests were caused by the four assembly lines and/or tester 104. This example shows how data mining techniques can be used to identify the root causes of quality problems in a manufacturing process. That kind of knowledge is valuable for quality diagnosis and quality improvement.

3.3 Quality improvement using customer service data

Manufacturing companies have obligations to repair sold products covered by the manufacturer’s warranty. Customer service records can be a valuable information source for quality improvement. Those records usually are multi-dimensional with many attributes. It is a challenging issue for companies to extract useful information from the vast amount of customer service records and provide feedback to product design and manufacturing quality improvement.

We propose a research design for customer service data analysis. It includes four major steps as we discuss below.

Step 1: Data structure design

Customer service data usually includes customer ID, product ID, service items and duration of the owner ship. Fig. 7 shows an example of the relational database used to store the customer service data. The database design follows the guidelines of developing a relational database (Edgar F Codd, 1970).

This is the fact table of customer service data. We can define different dimensions for data analysis. For example, in a mobile phones customer service domain, dimensions are customer, product, service items and the duration of ownerships’.

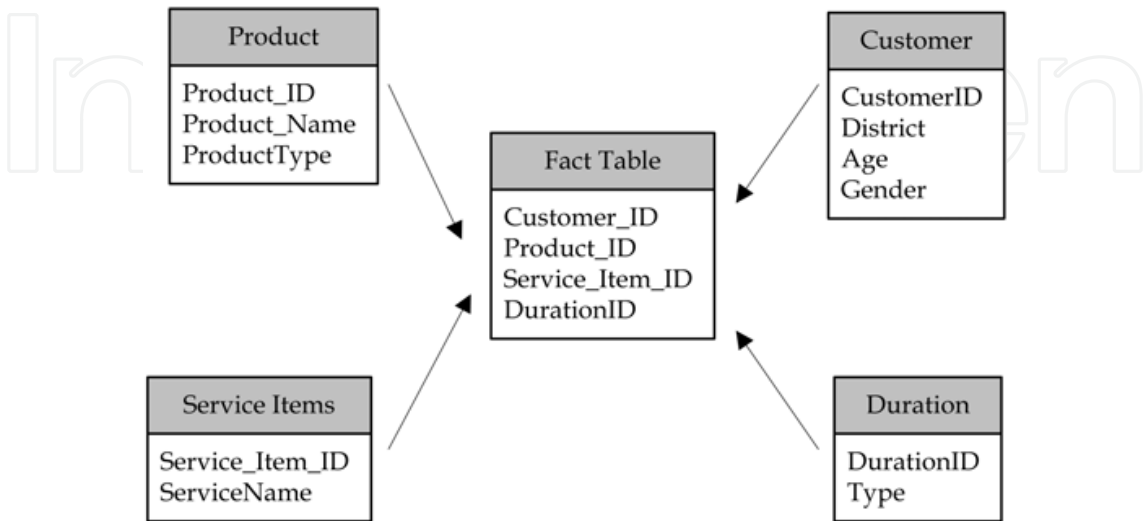


Fig. 7. The database design for the customer service data

**Step 2: MDA (Multi-Dimensional Analysis)**

The stored customer data can be analyzed in different dimensions. And the MDA can give a statistical interpretation of the data. With the drill-down and drill-up operations, we can analyze the data at a detailed level or a summarized level.

**Step 3: Data mining**

With the integrated customer service data, data mining techniques such as decision tree or association rule mining can be used to find the relationship among customers, products, durations and service items. These kinds of relationships will help improve product design.

**Step 4: Rules certification and usage**

The patterns and knowledge obtained by data mining techniques must be verified by domain experts or customer interview before being applied.

We illustrating the methods discussed above using 2000 records extracted from a mobile phones manufacturer's service process. The purpose is to demonstrate how the learned knowledge can improve the quality of products by design.

The data were rearranged to a table with seven columns, including customer ID, Age, District, Gender, Product Type, Service Item and Duration. The customer's age was divided into five intervals of above 55, between 45 and 55, between 35 and 45, between 25 and 35 and below 25. The location of customers was divided into four districts; they are southeast, southwest, northeast and northwest of China. There are four types of products named as series A, series B, series C and series D. The service items were classified into four types including the product appearance, product function, product performance, and others. The duration of ownerships was divided into four intervals: 1-6 months, 7-12 months, 13-18 months and 19-24 months.

To identify the relationship among these factors, we used the CHAID decision tree algorithm (J. A. McCarty, 2007) to analyze the data. Fig. 8 shows the decision tree.

There is an interesting pattern in the decision tree. 48.3% of the products had function defects. If the products were used in the district of southeast of China, the percentage increased to 79.13%. Among them, 69.95% were the product of series C, 21.7% series B, and only 8.3% product of series D. Obviously we can say that if the products of series C are used in southeast of China, they are inclined to have functional defects.

Furthermore, we should analyze the rule deeply. After a technical examination of the repaired products, it was found that the repaired products have almost the same erosion in circuit boards. Compared with the environment conditions, it was found that the reason of the functional defects of the products was the humid atmospheric corrosion. Erosion occurred more easily on the products of series C than other types of products. The defects were eliminated by design improvement. But this kind of problems was not very easy to identify without the deep analysis of the service data.

**4. System framework**

Based on the above discussion, we propose a system infrastructure for quality improvement using data mining techniques (Fig. 9). There are three layers in the model, namely data collection layer, data analysis layer and data view layer. The function of each layer is described below.



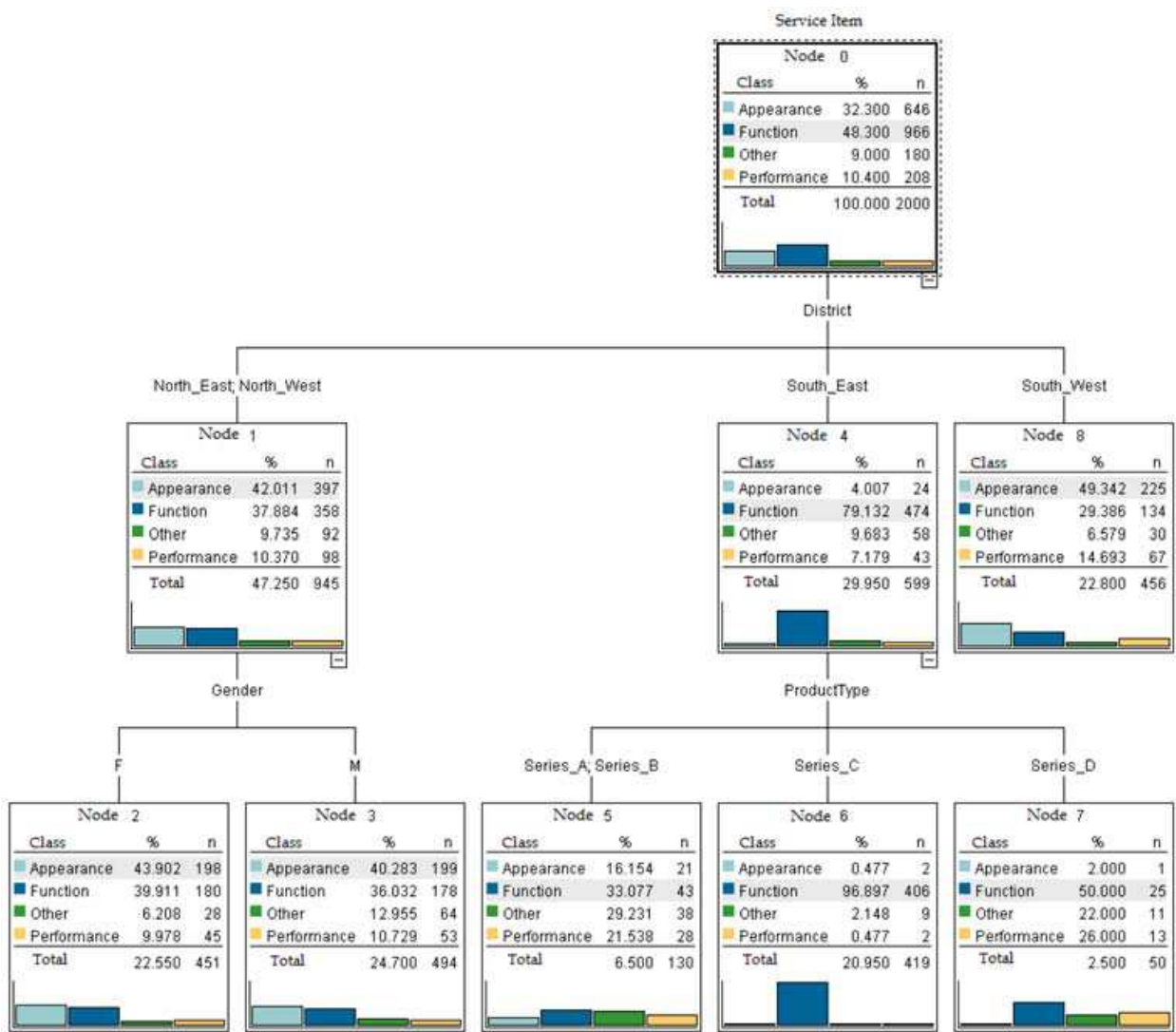


Fig. 8. The decision tree of service data analysis

1. Data collection layer
- Functions in this layer are mainly used by operators for daily operations. They include quality data maintenance, simple data analysis and query, and generating daily reports.
2. Data analysis layer
- Functions of this layer are mainly used by mid-level managers and quality engineers. This layer is the center of the system. Most of the quality data analysis functions are provided in this layer including SPC, statistical analysis, data mining, quality diagnosis, etc.
3. Data view layer
- Functions of this layer are mainly used by top managers for an integrated data view. In this layer, data are often shown in graphs or reports.
- Among the four layers, the data collection layer is the basis of the system. All the quality data are collected and stored by this layer. The data analysis layer is the middle layer, where the raw data are processed and analyzed. Results of the data analysis layer can be transferred to the data view layer where can help high-level managers make quality improvement decisions.



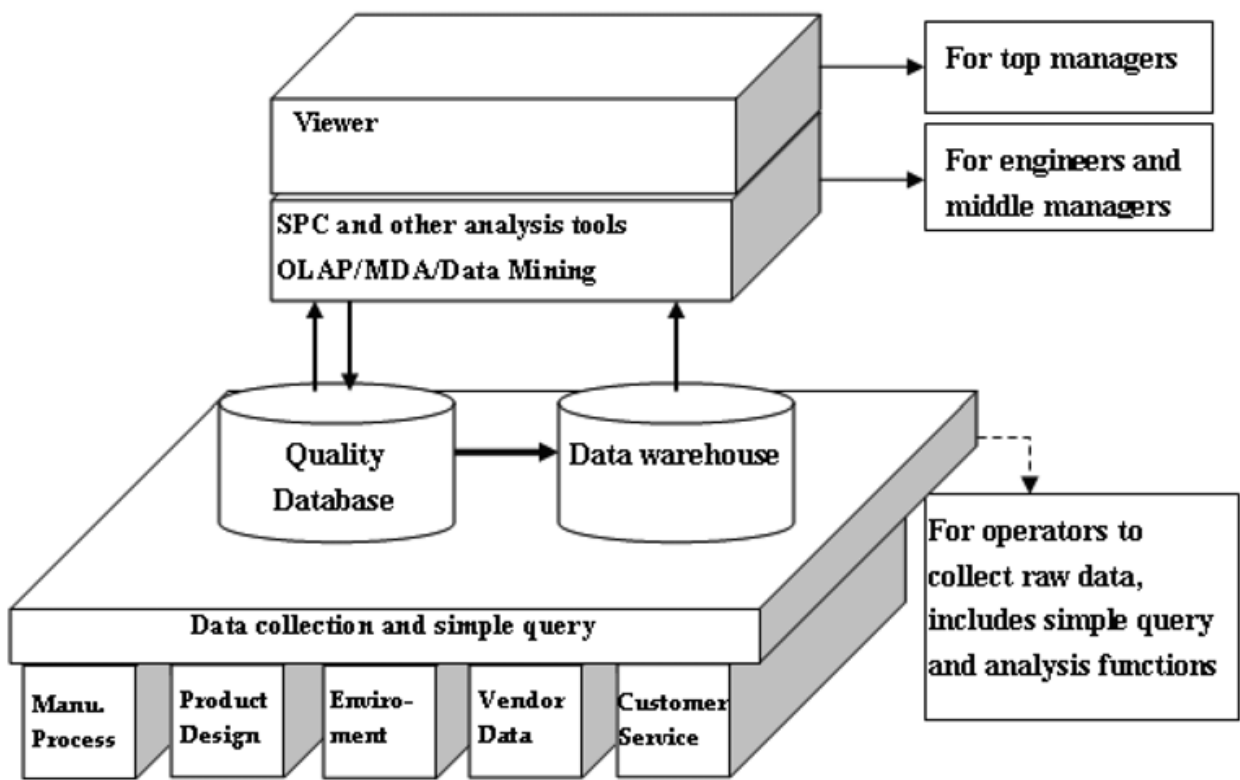


Fig. 9. A system infrastructure for quality improvement using data mining techniques

5. Conclusions and further research

In a competitive global market, manufacturing enterprises must stay agile when making quality improvement decisions. The development of IT and other related technologies makes the collection of quality related data easy and cost-effective. However, it is still an open question on how to leverage the large amount of quality data to improve manufacturing quality. This chapter has approached the problem of quality improvement in manufacturing processes using data mining techniques.

Firstly, we proposed a knowledge based six-sigma model where DMAIC for six-sigma was used along with data mining techniques. The knowledge learned by the data mining techniques was helpful in identifying potential quality problems and assisting quality diagnosis. We also examined the problem of parameter optimization by applying data mining techniques to DOE. A decision tree was build in order to dynamically adjust parameter optimization. In addition, we applied data mining to quality diagnosis where an association rule mining technique was used to analyze the electronic test data. The rules obtained by data mining provided a direct guidance in identifying the root causes of the quality problems. And the findings were beneficial for quality diagnosing that is still a difficult problem in six-sigma. Furthermore, a decision tree was also used in the service data analysis. The findings were valuable to improve product design. Finally, we presented a system infrastructure for quality improvement in manufacturing processes.

On the other hand, there are still a lot of open questions to be studied in this field. Firstly, the manufacturing processes are quite different with each other. The quality of the collected data from each manufacturing process varies significantly. There can be a significant portion of missing values and errors in the raw data. It is still a challenging issue as how to

preprocess raw data for data mining algorithms. Secondly, patterns and knowledge learned by data mining techniques are not always usable. How to ascertain the usable knowledge in a large amount rules and patterns is also a problem that deserves attention. Finally, the learned rules and patterns have to be analyzed by domain experts with their domain knowledge. How to present the domain knowledge and build an automated knowledge ascertain system is also a challenging issue in this field.

## 6. Acknowledgement

The authors of this work would like to thank the National Natural Science Foundation of China (NSFC) who sponsored this research (grant no. 70572044). The authors would also like to thank the anonymous reviewers for their constructive comments on this work.

## 7. References

- Andrew Kusiak, Christian Kurasek. Data mining of printed-circuit board defects, IEEE transactions on robotics and automation, vol.17, No.2, 2001.
- Angie Patterson, Piero Bonissone, Marc Pavese. Six sigma applied throughout the lifecycle of an automated decision system, quality and reliability engineering international, 2005.21: 275-292
- Ben Khediri Issam, Limam Mohamed. Support vector regression based residual MCUSUM control chart for autocorrelated process, Applied mathematics and computation, 2008 (In press)
- Chen-Fu Chien, Wen-Chih Wang, Jen-Chieh Cheng. Data mining for yield enhancement in semiconductor manufacturing and an empirical study, Expert Systems with Applications, 2007.33: 192-198
- Chen-Fu Chien, Huan-Chung Li and Angus Jeang. Data mining for improving the solder bumping process in the semiconductor packaging industry, Intelligent systems in accounting, finance and management, 2006.14: 43-57
- Edgar F. Codd, A Relational Model of Data for Large Shared Data Banks, Communications of the ACM 13(6): 377-387.
- Giovanni C Porzio, and Giancarlo Ragozini. Visually mining off-line data for quality improvement, Quality and reliability engineering international, 2003.19:273-283
- Hsu-Hwa Chang. A data mining approach to dynamic multiple responses in Taguchi experimental design, Expert systems with applications, 2008.35: 1095-1103
- Ian H Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann Publishers, 2005.
- J A McCarty, Manoj Hastak. Segmentation approaches in data-mining: a comparison of RFM, CHAID, and logistic regression, Journal of Business Research, 2007.60: 656-662.
- Kaidi Zhao, Bing Liu, Tomas M Tirpak and Weimin Xiao. A visual data mining frame work for convenient identification of useful knowledge, proceedings of the fifth IEEE international conference on data mining, 2005.
- Myers R H, Montgomery D C. Response surface methodology. John Wiley & Sons, New York, 1985.
- Mu-Chen Chen. Ranking discovered rules from data mining with multiple criteria by data envelopment analysis, Expert systems with application, 2007.33:1110-1116

- Rakesh Menon, Loh Han Tong, S Sathiyakeerthi, Aarnout Brombacher and Christopher Leong. The needs and benefits of applying textual data mining within the product development process, *Quality and reliability engineering international*, 2004.21:1-15
- Ruey-Shiang Guh, Yeou-Ren Shiue. An effective application of decision tree learning for on-line detection of mean shifts in multivariate control charts, *Computers & Industrial Engineering*, 2008 (In Press)
- Seyed Taghi Akhavan Niaki, Babak Abbasi. Fault diagnosis in multivariate control charts using artificial neural networks, *Quality and reliability engineering international*, 2005.21: 825-840
- Sébastien Gebus, Kauko Leiviskä. Knowledge acquisition for decision support systems on an electronic assembly line, *Expert systems with applications*, 2007 (In press).
- Shao-Chuang Hsu, Chen-Fu Chien. Hybrid data mining approach for pattern extraction from wafer bin map to improve yield in semiconductor manufacturing, *Int J Production Economics*, 2007.107: 88-103
- Tai-Yue Wang, Long-hui Chen. Mean shifts detection and classification in multivariate process: a neural-fuzzy approach, *Journal of intelligent manufacturing*, 2002.12: 211-221

IntechOpen



## **Data Mining and Knowledge Discovery in Real Life Applications**

Edited by Julio Ponce and Adem Karahoca

ISBN 978-3-902613-53-0

Hard cover, 436 pages

**Publisher** I-Tech Education and Publishing

**Published online** 01, January, 2009

**Published in print edition** January, 2009

This book presents four different ways of theoretical and practical advances and applications of data mining in different promising areas like Industrialist, Biological, and Social. Twenty six chapters cover different special topics with proposed novel ideas. Each chapter gives an overview of the subjects and some of the chapters have cases with offered data mining solutions. We hope that this book will be a useful aid in showing a right way for the students, researchers and practitioners in their studies.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Shu-guang He, Zhen He, G. Alan Wang and Li Li (2009). Quality Improvement using Data Mining in Manufacturing Processes, Data Mining and Knowledge Discovery in Real Life Applications, Julio Ponce and Adem Karahoca (Ed.), ISBN: 978-3-902613-53-0, InTech, Available from:  
[http://www.intechopen.com/books/data\\_mining\\_and\\_knowledge\\_discovery\\_in\\_real\\_life\\_applications/quality\\_improvement\\_using\\_data\\_mining\\_in\\_manufacturing\\_processes](http://www.intechopen.com/books/data_mining_and_knowledge_discovery_in_real_life_applications/quality_improvement_using_data_mining_in_manufacturing_processes)

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2009 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen