2011 International Conference on Computer Science and Information Technology (ICCSIT 2011)
IPCSIT vol. 51 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2012.V51.58

The Clothing Quality Forecast Based on Rough Set

Tao Wu [a+], YuanShu Du[b], Yin Wang[b], Shousan Jiang[a] and Xianbo Wang[a]

[a]Department of Computer Science,Xi'an Polytechnic University,Xi'an, 710048, China
[b] Shandong Ruyi Group,Ji'ning, 273500, China

Abstract. In this paper, towards the massive enterprise production data, a clothing quality forecasting system has been established to solve the problem of the repair rate in garment production process remaining high. It can predict the various factors that affect the failure rates. Through the application of a garment enterprise, it is obvious that the system can effectively reduce the failure rates and enhance the clothing quality and production benefit.

Keywords: rough set theory; discretization of data; attribute reductionstyle; repair rate; rule set;

1. Introduction

The increase of people's consumption level to clothing and the increasingly fierce international competition greatly improves the quality requirements of garment processing enterprise. However, the complicated process and human factors of garment processing keep the repair rate of clothing enterprise products and the failure rate of customer inspection high[1].

At present, the methods of testing the quality are different in the clothing enterprises. At this time, the most garment production models with manual operation lead to uncertainty in the quality of clothing. With the development of information technology, more and more clothing enterprises established information systems for themselves, such as clothing production tracking system. The systems can track the clothing order on the garment production lines. Meanwhile, it was used to record some clothing quality issues. Then there are abundant data storing in the system.

2. Rough Set Theory

Rough set theory is a tool of mathematics describing the incomplete and uncertainty. It can effectively analyze and deal with several kinds of incomplete information, such as inaccurate, inconsistent and incomplete data, and discover hidden knowledge to reveal the potential law [2]. The main idea is to output the decision-making classification rule by knowledge reduction keeping the classification ability. It regards the knowledge as the partition to the domain [3].

The basic principles of information processing based on rough set are following: Inputting the corresponding information into system S, while outputting the rules we need. Information system S=(U, A, V, f) , U is the non-empty finite set in the decision table records, A is non-empty finite set of attributes, V is attribute domain set, f is the information function.

The data mining process using rough set is as follows:

(1) Preparing and preprocessing data. Mining-related data will be collected and then eliminated the inconsistent and incomplete problems in these data. At last, the continuous data will be discreted so as to get decision table suitable for rough set.

[+] Corresponding author. Tel.: +086-15596866380
E-mail address: wutao1976@163.com.

(2) Attribute reduction. Attribute reduction is a very important step in mining process by deleting unnecessary attribute in decision table to simplify the data mining operation. Because there are a huge number of attributes and examples in the information system with massive data sets, the efficiency of attribute reduction algorithm is particularly important[4,5].

(3) Rule generation. We can obtain the required rule through value reduction to the data which has been attribute reduced.

(4) Decision support. Using obtain rule to Analyze and judge the actual data.

3. The Process and Model of Clothing Production Quality Forecast

3.1. Factors that Affecting the Quality of Clothing

Through the investigation of clothing enterprise, we found the following factors which have different effects to clothing production. Including:

(1) Difficulty of process(DP)

All orders have different requirements to design, while the technology difficulty coefficients also have difference. The style of big technology difficulty coefficients has higher technical requirements and larger quality problems arising possibility.

(2) The balance degree of assembly line(BDAL)

The balance degree of assembly line also has great effect to the clothing quality. Compared with the balanced line, the irregular assembly line results in employees' lower work efficiency and larger possibility in quality problems.

(3) Employees' technical proficiency(ETP)

Employees' technical proficiency greatly influences processing quality and efficiency.

(4) Employees' work attitude

Garment production is mostly completed by artificial. So the employees' work attitude as a kind of subjective factor has a certain influence.

(5) Weather(W)

The weather has certain influence to the employees' mood. High temperature often leads to high failure rates.

(6) Characteristics of fabrics and materials(CFM)

Each order defines the fabrics and materials which needed to use. Some characteristics of fabrics and materials have a great impact on clothing production processing, such as fabric thickness, raw materials and the availability of flexibility.

(7) Season of manufacture(SM)

The clothing production is related to the season. Most of the time garment enterprises produce anti-season clothing, so there are differences between low season and peak season. Generally speaking, the peak season has lower repair rate.

(8) Other factors(OF)

In addition to the factors mentioned above, other uncertain factors also impact on the quality, just like the production duration. The orders with short duration easily lead to high repair rate.

According to the survey and analysis, there is a definite relation between various factors. As figure 1 shows.

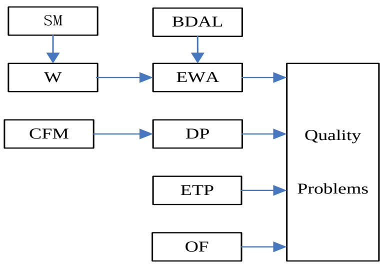

Fig.1. Factors affecting the quality and their contact

3.2. The Process of Forecasting

According to the requirement analysis, the process of forecasting is as figure 2 shows.

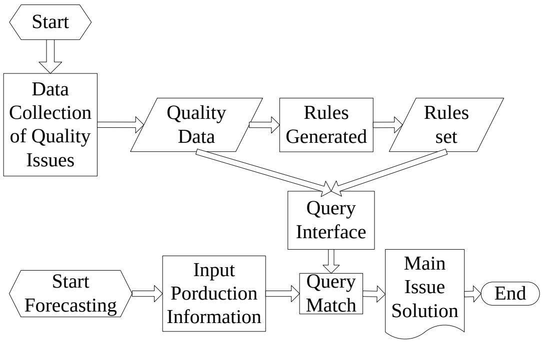

Fig.2. the process of forecasting

The data is the relevant of clothing quality issues which collected through the actual research on clothing enterprise, including factors which affect the quality of clothing, quality issues data caused by these factors, and the data about the relationship between production processes and quality issues. The Relevant rule which can predict clothing quality problem is obtained by using rough set data mining software Rosetta to mine relevant data.

3.3. The Model of Quality Forecasting

The quality prediction model can predict the quality problems which may appear in the production process of this kind and the most likely reason for the problem, according to the input data related to the clothing production, such as fabric, materials, etc. To achieve this module, the first step is to find the key factors that lead to quality issues and dig out the quality issues related rules. Then forecast the clothing quality problems for a particular garment production. Preprocessing the data includes the following steps:

(1) Preparing data

The related database will be established based on the above analysis of the various factors which effect clothing quality. The data used in the system has been stored in the actual production of a certain clothing production enterprise. Preprocessing the data includes the following steps:

① Data completing

There are many ways to complete the data. The principle of the method used in this paper is to combine the incomplete record with all possible records of decision table, but the information filled in the table depends on the decision attributes. This paper uses the conditioned combination completion data supplement method of mining software ROSETTA based on rough set to fulfill the decision table. According to this method, the factors affecting the quality problem has been defined as A1, A2,⋯⋯ An. At the same time, the decision attribute defines as D. The missed attribute has been completed according to the decision attribute D.

For example: one quality problem recorded as x, the value of attribute Ai is missed, and the value of its decision attributes D is 10205.Then if the decision attributes value of one record is 10205, fulfilling Ai with the value of this record.

② Discretization of data

Assuming that the data type of all condition attribute *A* are numeric. For each condition attribute *a*, get the following sort results by sorting its attribute value v_a: $v_a^1 < \cdots < v_a^i < \cdots < v_a^{|v_a|}$. x_a^i is the set of record that its value of attribute *a* is v_a^i. Δ_a^i is the set of decision attribute related to x_a^i. *Ca* is the cut sets about attribute *a* which including the middle value of two testing attribute values. Through distinguish the records with the same decision attribute values, the obvious redundant breakpoint will be deleted from the sets.

$$x_a^i = \{x \in U \mid a(x) = v_a^i\}$$

$$\Delta_a^i = \{v \in v_d \mid \exists x \in x_a^i, d(x) = v\}$$

$$Ca = \{\frac{v_a^i + v_a^{i+1}}{2} \| \Delta_a^i | > 1 or | \Delta_a^{i+1} | > 1 or \Delta_a^i \neq \Delta_a^{i+1}\}$$

According to the principle of data discretization, the completed attribute named process difficulty in the decision table will be discretized. The results of discretization are shown as set, which include the left endpoints and exclude the right endpoints. Some results are shown in table 1. The corresponding decision table has shown the differences in table 2.

Table 1. Part of the discreted process difficulty data

NO	*Left value*	*Right value*
1	0	2.770
2	2.770	2.8775
3	2.8775	2.985
4	2.985	3.116
5	3.116	3.216
6	3.216	3.254
7	3.254	3.2865
8	3.2865	3.401
9	3.401	3.576

Table 2. Before and after discretization the corresponding decision table

	Fabric	*Lining*	*Difficulty of process*
Before	Thin wool	whole stick	[3.918,*)
	Thick wool	Half stick	[3.287,3.4)
	Thin wool	whole stick	[3.287,3.4)
After	*Fabric*	*Lining*	*Difficulty of process*
	Thin wool	whole stick	3.922
	Thick wool	Half stick	3.300
	Thin wool	whole stick	3.300

(2) Attribute reduction

Normally, reduction algorithm does not need to calculate all reduction for information system. It only obtains the reduction that user interested or available. Common calculation methods are Greedy method and the discernibility matrix method[7,8]. This paper adopts Johnson's algorithm of reduction.

Define information system $S=(U,A,V,f)$.

Universe $U=\{x_1,x_2,\cdots,x_n\},|U|=n$, $\forall\alpha\in A$, $\forall x_i,x_j\in U$.

The difference variable related to attribute α between object x_i,x_j is:

$$\alpha(x_i,x_j)=\{\alpha\,|\,(\forall\alpha\in A)\wedge(f_\alpha(x_i)\neq f_\alpha(x_j))\}.$$

Then define

$$\sum\alpha(x_i,x_j)=\begin{cases}\alpha_1\vee\alpha_2\vee\cdots\vee\alpha_k,\alpha(x_i,x_j)=\{\alpha_1,\alpha_2,\cdots,\alpha_k\},\\ 1, \qquad\qquad \alpha(x_i,x_j)=\phi\end{cases}$$

Define discernibility function as follows:

$$\Delta=\prod_{\forall\ (x_i,x_j)\in U\times U}\sum\alpha(x_i,x_j),i,j=1,2,\cdots,n.$$

According to the discernibility function, we should calculate the difference corresponding to all condition attributes between the record and others in the system, which is defined as $f(i)$. i is the order of the record in the information system and count from 0. The set which needed to reduce is defined as B. S' is the set of attribute combination by calculating the discernibility function Δ of information system S. Defining $w(S')=1$ for each combination in S' to calculate the importance of attributes.

Steps:

①define B=ϕ;

②choose attribute a in S' that the $\sum w(S')$ is the biggest;

③add attribute a to B;

④remove the combination that includes a from S';

⑤if $S'=\phi$, return B; otherwise, go back to step 2.

By the attribute reduction following the above steps, the core attribute of the decision table shown in table 2 is shown as table 3. The attribute set is obtained after attribute reduction. Length is the number of the attributes after reduction.

Table 3. Attributes reduction result

reduction	*Support*	*Length*
（operation_dif, job_attitude, Weather,pipeline,proficiency）	100	5

From table 3, we can see that the main factors affecting the clothing quality are process difficult, employees work attitude, weather, the balance degree of assembly line and employees technical proficiency. Because other factors have little effect on the clothing quality in production process, they can not be considered.

(3) Rule generating

Generating rules is based on the resulting set of attribute reduction. Sequentially scan the record of decision table, and add the corresponding attribute values and decision attribute values to the rule set. If the rule set has the same record, the appropriate support should be changed. Otherwise, the record will be added to the rule set as a new rule.

In the rule set, support after if expresses the number of records which meet the condition attribute value, coverage express the proportion of such records in the whole decision table record. Decision support expresses the number of records that appear the decision attribute value when meet the previous conditions. Accuracy expresses the proportion of appearing the decision attribute value in this rule. Coverage expresses the number of appearing the decision attribute values on this condition versus the number of appearing the decision attribute values in the decision table.

According to the decision table shown as table 2and the reduce results shown as table 3, we can generate the rule shown as figure 3.

```
<rule>
    <if support="1250" coverage="0.00970874">
        <and>
            <descriptor attribute="operation_dif" value="[3.401,3.576)"/>
            <descriptor attribute="job_attitude" value="Serious"/>
            <descriptor attribute="weather" value="Clear"/>
            <descriptor attribute="pipeline" value="Smooth"/>
            <descriptor attribute="proficiency" value="Skilled"/>
        </and>
    </if>
    <then>
        <or>
            <decision support="339" accuracy="0.2712" coverage="0.2">
                <descriptor attribute="qltiss_id" value="20113"/>
            </decision>
            <decision support="128" accuracy="0.1024" coverage="0.125">
                <descriptor attribute="qltiss_id" value="20207"/>
            </decision>
            <decision support="783" accuracy="0.6264" coverage="0.25">
                <descriptor attribute="qltiss_id" value="20213"/>
            </decision>
        </or>
    </then>
</rule>
```

Fig.3. Decision Rules

The rule expresses that quality problem number 20113, 20207 or 20213 will happen when the process difficulty between 3.401 to 3.576, employees with hard-working attitude, good weather, assembly line in an orderly way and skilled workers. Because the quality problem number 20213 account for 783 records in 1250 records as a proportion of 0.6264, the problem most likely happens at the time of predicting the problem on this condition.

(4) Selecting rules

Some of the generated rules are not universal, so we should filter the rule set by deleting the rule which dose not meet the condition of accuracy and confidence to get the new universally applicable rules.

(5) Systematic training

The generated rules need to be tested after the mining. Some of the data are used to generate the rules, and the others for testing the system in order to modify the unsuitable rules so as to achieve the best forecast effect.

3.4. Forecast quality issues

According to the rules, the problem can be predicted after determining the input information. The prediction interface of system is shown in figure 4. The enterprise can take appropriate measures to prevent possible problems when obtain the prediction results.

According to the rules shown in figure 3, we can see that the corresponding decision attribute values are not unique when the condition attribute with a certain value. At this time, the next issues should be analyzed according to the times, the accuracy and the coverage to the purpose of quality problems prediction.

Fig.4. the quality forecasting interface

4. Application analysis

By tracking the system trial situation of clothing enterprise, the repair rate goes down gradually in using rough set to predict clothing quality. The result is shown in figure 5.

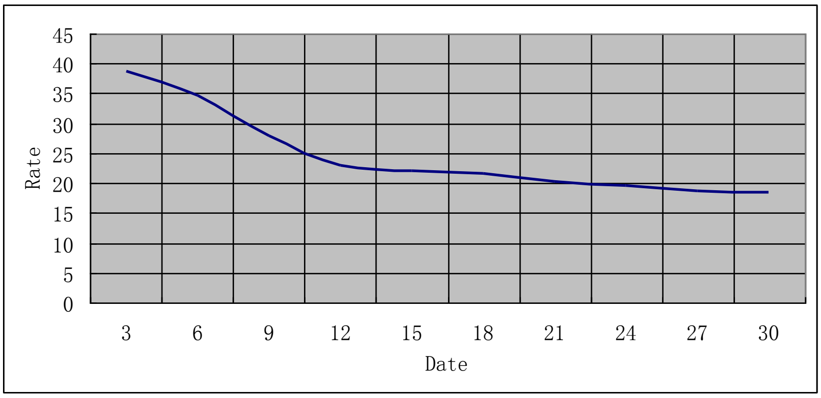

Fig.5. changes in repair rates chart

With this system, according to the status of production, it can be predicted that the quality problem which may appear in an order and the focus process caused the problem. And then, according to the experts' advice and the production experience, enterprise can get the measure of preventing and improving so as to reduce the repair rate.

5. Conclusion

For clothing companies, the clothing quality is directly related to the efficiency of enterprises and Corporate Image. Through the prediction of clothing quality problems, managers can take appropriate measures to prevent them so as to reduce repair rate and improve the production efficiency. The relationship between various factors that affect the clothing quality is often uncertainty. Using the data mining method which based on rough set, this paper analyzes the data in clothing enterprise tracking system and digs out the relationship between the factors that hide in the data and their impact on quality. Then get the law that affect the clothing quality issues and establish the clothing quality prediction system. The result of practical application shows that the system effectively reduces the repair rate and improves the clothing enterprise competitiveness.

6. References

[1] Wang Yishen. Test Production of Clothing in the Position [J]. *Shandong Textile Economy*. 2009, 1: 120 - 121.

[2] JI Bin. Research on Rough Set Theory and Its Progress on Applications [J]. *Computer Technology and Development*. 2007，17(3): 69-76.

[3] ZHAO XI-lin, LIU Hui. An Analysis of Problem about Rough Set Theory in Fault Diagnosis [J]. *Computer Technology and Development*. 2008，18(1): 132-135.

[4] NIE Lin-di. The Study on Attribute Reduction Algorithm Based on the Rough Set Theory [J]. *Computer Knowledge and Technology*. 2009，5(9): 2428-2429.

[5] Ding, Shifei; Ding, Hao; Research and development of attribute reduction algorithm based on rough set , *2010 Chinese Control and Decision Conference, CCDC 2010*, p 648-653.

[6] LI Hui, YAN De-qin, HAN Li. Novel algorithm for discretization of continuous attributes based on rough sets theory [J]. *Application Research of Computers*. 2010，27(1) : 77-78.

[7] LIU Yong, XIONG Rong, CHU Jian. Quick Attribute Reduction Algorithm with Hash [J]. *Chinese Journal of Computers*. 2009,8.

[8] Xu, E.; Shao, Liangshan; Qiao, Zhu; Cao, Guanghui; Qiu, Feng; Attribute reduction method of uncertain information, *Advanced Materials Research*, v 108-111, n 1, p568-573, 2010, Progress in Measurement and Testing.

Appendix A. About the author and this study

The first author is a PhD of Aerospace Manufacturing, Research in information engineering and data mining. The study mainly focuses on the unpredictability of clothing production quality, which results in the problem of high failure rates. By studying the use of rough set theory and methods, the relationship of the production process and conditions were obtained. This makes the garment before production can better predict the impact on the quality of the key points. Strengthen the management of these critical points, we can reduce failure rates.This study has been supported by Key Project of the National Eleventh-Five Year Research Program of China. The system has been run in a large clothing company, and achieved very good results.