Expert Systems with Applications 40 (2013) 2435–2446

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A hybrid OLAP-association rule mining based quality management system for extracting defect patterns in the garment industry

C.K.H. Lee [a], K.L. Choy [a,*], G.T.S. Ho [a], K.S. Chin [b], K.M.Y. Law [a], Y.K. Tse [c]

[a] Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong
[b] Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong
[c] The York Management School, University of York, United Kingdom

ARTICLE INFO

Keywords:
Quality management
Garment industry
Garment defect
Association rule mining
OLAP

ABSTRACT

In today's garment industry, garment defects have to be minimized so as to fulfill the expectations of demanding customers who seek products of high quality but low cost. However, without any data mining tools to manage massive data related to quality, it is difficult to investigate the hidden patterns among defects which are important information for improving the quality of garments. This paper presents a hybrid OLAP-association rule mining based quality management system (HQMS) to extract defect patterns in the garment industry. The mined results indicate the relationship between defects which serves as a reference for defect prediction, root cause identification and the formulation of proactive measures for quality improvement. Because real-time access to desirable information is crucial for survival under the severe competition, the system is equipped with Online Analytical Processing (OLAP) features so that manufacturers are able to explore the required data in a timely manner. The integration of OLAP and association rule mining allows data mining to be applied on a multidimensional basis. A pilot run of the HQMS is undertaken in a garment manufacturing company to demonstrate how OLAP and association rule mining are effective in discovering patterns among product defects. The results indicate that the HQMS contributes significantly to the formulation of quality improvement in the industry.

1. Introduction

Nowadays, customers are seeking products of high quality and low cost. Manufacturers are urged to achieve better quality in their products so as to stay competitive in the industry. Unfortunately, variance in product quality is unavoidable as it can be induced by many factors during production. One of these critical factors is the workmanship which commonly exists in labor-intensive industries where many production processes are performed manually. In the garment industry, human factors such as different levels of skill, years of experience and human errors may result in garment defects in some circumstances. In order to produce high-quality and low-cost products, it is important to achieve quality improvement while at the same time to identify any product defects at an early stage. Unfortunately, it is challenging to maintain the quality of garments which are processed manually. It is thus necessary to inspect products carefully so as to ensure they are of good quality. Traditionally, garment defects are identified by human inspectors who treat each defect individually without being aware of the relationship between different defects, thus making causal analysis and defect prediction difficult.

Fig. 1 depicts the existing problems in handling quality problems in the garment industry. There are various departments responsible for different tasks along the production workflow from product design to final product inspection. Hence, it is difficult for manufacturers to identify the department to which a particular garment defect should be attributed, and the root causes of the defects. This shows that there is a lack of information to tackle product quality problems. Manufacturers do not have timely information to analyze defect causes and individual departments fail to be aware of any possible defects that they might be causing. In addition, owing to the complexity of garment manufacturing processes, there are numerous defects which can be found on a single garment. Without any tools to manage massive relevant data and identify the hidden pattern among defects, manufacturers are unable to discover any correlations between defects, or the reasons for different defects. This indicates the lack of a mechanism for investigating defect patterns which could be useful in defect prediction and defect diagnosis. The problems outlined above will certainly lead to bad consequences for the garment industry, such as failure in achieving quality improvement, low customer satisfaction, high rework cost and long production cycle times. With the aim of tackling these problems, this paper presents an intelligent system, namely hybrid OLAP-association rule mining based quality management system (HQMS), to extract garment defect patterns in

* Corresponding author. Tel.: +852 27666597; fax: +852 23625267.
E-mail address: mfklchoy@inet.polyu.edu.hk (K.L. Choy).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.10.057

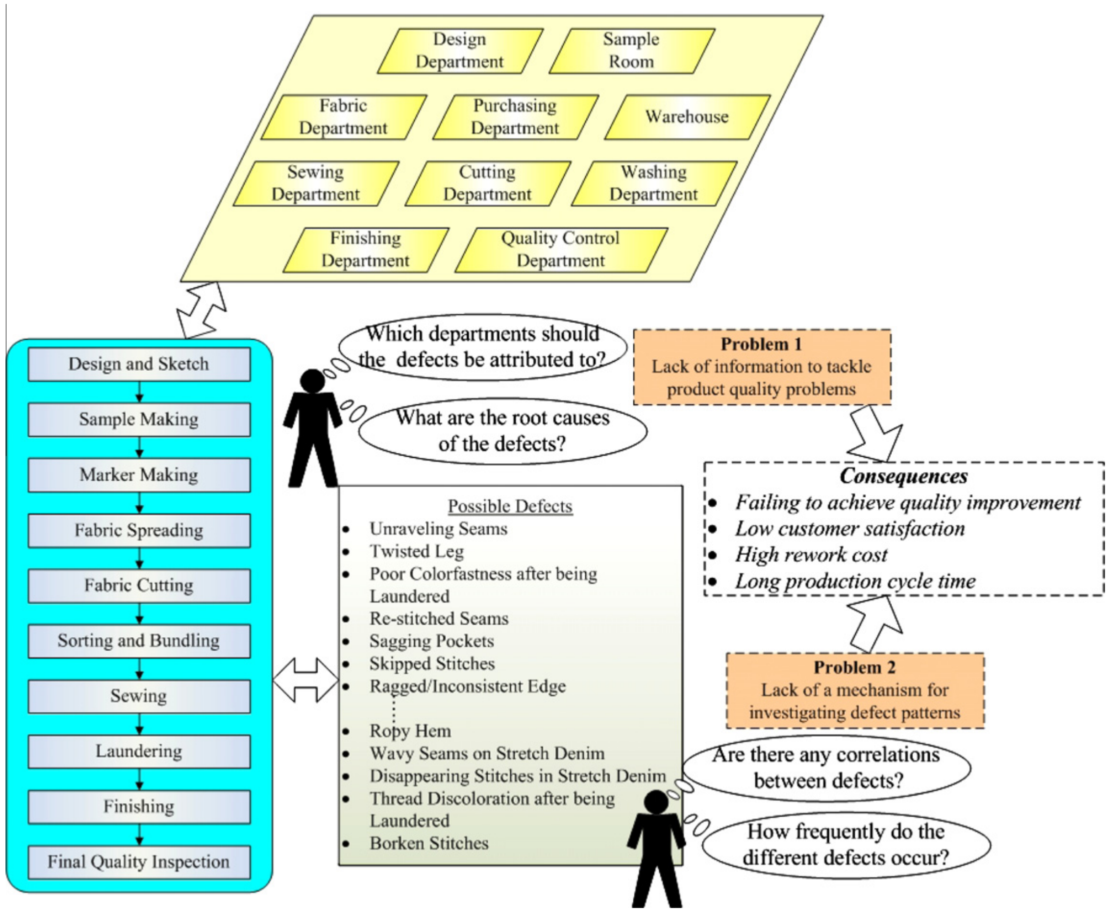


Fig. 1. Existing problems in handling quality problems in the garment industry.

the form of association rules so as to formulate useful quality improvement plans. To provide manufacturers with the ability to explore different kinds of desired data effectively and mine data at different levels, Online Analytical Processing (OLAP) is also applied in the system that is developed.

This paper is organized as follows: Section 2 is a literature review related to this study. In Section 3, the architecture of the HQMS is proposed. Section 4 contains a case study where a pilot run of the system is conducted in a case company. In Section 5, the discussion of the HQMS is presented. Finally, Section 6 is the conclusion.

2. Related studies

Owing to increased economic and environmental pressures in recent years, customers now have higher expectations of products. They are now seeking products which can fulfill environmental requirements, are of high quality but low in cost. These changes present a formidable challenge to product quality improvement. In order to respond to the market changes, manufacturing firms have shifted their attention to quality control and improvement of their products. In many manufacturing sectors, product defects can be eliminated with better machine and process settings. This arouses the interest of many researchers to solve product quality problems by dealing with operation parameters. Ferreiro, Sierra, Irigoien, and Gorritxategi (2011) developed a model for burr prediction during drilling processes by taking machine settings such as that of the drill bit and drilling velocity into account. Lau, Ho, Chu, Ho, and Lee (2009) proposed a methodology for quality management with knowledge discovery based on quantitative process values including the temperature setting of machines, the thickness of the product and the time spent in cleaning. Lou and Huang (2003) developed an intelligent decision support system for defect reduction in automotive coating by recommending changes in process parameters such as the booth air temperature, humidity and the viscosity of the paint. However, in labor-intensive industries, adjustment of machine and process parameters may not be fully applicable to the resolution of their product quality problems which could be caused by workmanship rather than by machines.

Because of the error-prone nature of labor-intensive manufacturing processes, inspection of semi-finished products and finished products is critical in labor-intensive industries. In particular, inspection of garment products usually solely relies on human effort (Yuen, Wong, Quan, Chan, & Fung, 2009), resulting in biased inspection results (Wong, Yuen, Fan, Chan, & Fung, 2009). With the aim of achieving better quality control and improvement of garment products, researchers have started investigating the possibilities of automatic detection of defects which exist in the textile and garment industry. Mak, Peng, and Yiu (2009) proposed a novel defect detection scheme to facilitate automated inspection of woven fabrics. Wong et al. (2009) combined wavelet transformation and a neural network to detect and classify stitching defects. In a similar vein, Yuen et al. (2009) presented a novel hybrid model combining genetic algorithms and neural networks to detect stitching defects. It is found that existing works related to fabric or garment defects focus mainly on automated inspection systems. Yet, little effort has been paid to garment defect diagnosis, such as identification of root causes and corrective action to bring about quality improvement.

To perform quality diagnosis, various types of knowledge such as knowledge of the defect problems are required (Deslandres & Pierreval, 1997). Due to the ease of collecting relevant data, performing the analysis and interpreting the results, data mining applications are commonly employed to provide useful feedback to corrective actions for quality improvement (Baykasoglu, Özbakir, & Kulluk, 2011; Köksal, Batmaz, & Testik, 2011). Milne, Drummond, and Renoux (1998) extracted process data and past faults by data mining so as to discover patterns which may cause paper defects. Hsu and Chien (2007) presented a hybrid data mining approach to extract defect patterns from wafer bin maps and identify root causes of specific patterns. Cheng and Leu (2011) proposed a defect feedback system using a clustering algorithm to analyze bridge construction defects and classify them into different groups. Tsai (2012) integrated clustering and decision tree techniques to identify the soldering defect patterns and classify soldering quality. Since there are numerous possible types of defects which could be found on garment products, one of the critical aspects of planning for quality improvement is to discover the pattern among defects. Ur-Rahman and Harding (2012) also pointed out that product or service quality can be improved through the discovery of hidden knowledge. Chougule, Rajpathak, and Bandyopadhyay (2011) conducted root cause analysis of anomalies using association rule mining to improve service and repair in an automotive domain. Chang, Chu, and Yeh (2009) used association rules to discover software defect patterns for causal analysis and defect prediction. Among data mining techniques, association rules, which are useful in finding the correlation between items (Altuntas & Selim, 2012), could be an appropriate approach to providing knowledge support in root cause analysis of garment defects.

An association rule is represented in the form of "X → Y", where X and Y are defined as sets of attributes or items representing the "If" part and the "Then" part of the rule respectively. When garment defects are treated as items, an association rule is able to discover an "If–Then" relationship among defects, indicating which defects are likely occur if particular defects exist. The Apriori algorithm proposed by Agrawal and Srikant (1994) is considered a classical algorithm, effective for generating association rules between items in large databases (Chung & Tseng, 2012; Lim, Lee, & Raman, 2012). Two values, namely support and confidence, can be used to describe an association rule as follows:

$$\text{Support}(X \Rightarrow Y) = P(X \cup Y) \tag{1}$$

$$\text{Confidence}(X \Rightarrow Y) = P(Y|X) \tag{2}$$

The support for a rule is the percentage of transactions in the databases containing both X and Y while the confidence for the rule is the percentage of transactions in the database containing X that also contain Y (Martínez-de-Pisón, Sanz, Martínez-de-Pisón, Jiménez, & Conti, 2012). According to Yang, Mabu, Shimada, and Hirasawa (2011), association rule mining may generate a large number of rules. Some of them could be contradictory or irrelevant, thus only a reduced number of significant rules are valid for efficient decision making (Moreno, Ramos, García, & Toro, 2008). A rule is considered as significant if its support and confidence are greater than the user's-defined threshold values (Demiriz, Ertek, Atan, & Kula, 2011). Both threshold values for support and confidence which are defined by users will thus affect the quality of association rules (Lim et al., 2012). In view of that, in any association rule mining models, it is important to ensure that user-specified minimum support and confidence are appropriate to generate significant rules.

In addition, data warehousing is always viewed as a vital preprocessing step for data mining by organizing, cleaning and preparing data for data mining (Kusiak & Smith, 2007). In order to analyze the data effectively, there are various applications available, among which the most popular is Online Analytical Processing (OLAP) which allows users to navigate through multidimensional structures so as to access data in a more natural manner (Pardillo, Mazón, & Trujillo, 2010). OLAP is designed to provide summary views of the basis for the aggregation of data stored in the data warehouse according to user-defined business dimensions. For decision making purposes, data warehouse and OLAP are usually inter-connected, where the former is responsible for data storage and handling, and the latter captures the stored data to form meaningful information for decision making (Chow, Choy, Lee, & Chan, 2005; Lau, Wong, Hui, & Pun, 2003). The data warehouse firstly extracts, transforms and loads (ETL) data from different data sources such as operational databases. After that, decision makers perform interactive analyses using OLAP tools to find solutions for their decision tasks (Thalhammer, Schrefl, & Mohania, 2001). Though interconnecting data warehouses and OLAP has been performing well in decision support functions, discovery of knowledge based on implicit and unknown patterns often provides enterprises with more insights into their businesses (Jukić & Nestorov, 2006). This highlights a need to integrate OLAP and data mining methods. The integration of OLAP and data mining facilitates mining on diverse subsets of data and at different levels by such OLAP functions as drilling, pivoting, filtering, dicing, and slicing on the OLAP data cube. In this context, decision makers can gain more in-depth knowledge for their decision-making in defect diagnosis.

In summary, there are a number of approaches and systems which have been applied to improve product quality in different industries. However, very few of them were designed in the garment industry to effectively discover hidden patterns in defects. These patterns are difficult for humans to detect through simple inspection but they are important for quality improvement. However, data mining techniques have been well proven to be effective in conducting defect diagnosis. In this study, association rule mining is attempted in order to extract garment defect patterns which will then be used in root cause identification and analysis. As data mining is applied to a large quantity of data, a data warehouse and the OLAP technology are embedded into the system for storing and accessing data effectively. To ensure the validity of the generated rules and facilitate effective decision making, it is found that it is necessary to review both the minimum support and confidence thresholds regularly as this will facilitate the generation of significant rules.

3. Hybrid OLAP-association rule mining based quality management system (HQMS)

The HQMS is designed to extract quality related data from different sources within the organization and convert them into knowledge in terms of association rules which indicate the hidden relationship between different garment defects. The knowledge gained allows users to predict potential defects and analyze the possible causes of defects so that proactive measures can be taken in corresponding manufacturing departments. There are three key modules constituting the HQMS, namely the Data Warehouse Module, the OLAP Module, and the Association Rule Mining Module, as shown in Fig. 2. In the Data Warehouse Module, there is a centralized data warehouse responsible for data storage. In order to ensure that only the data in the standard format can be stored in the data warehouse, a process called extract, transform and load (ETL) is undertaken to preprocess data collected from different sources such as internal databases. To speed up the data extraction process, the data warehouse is constructed by aggregating data marts of different departments. In the OLAP Module, data stored

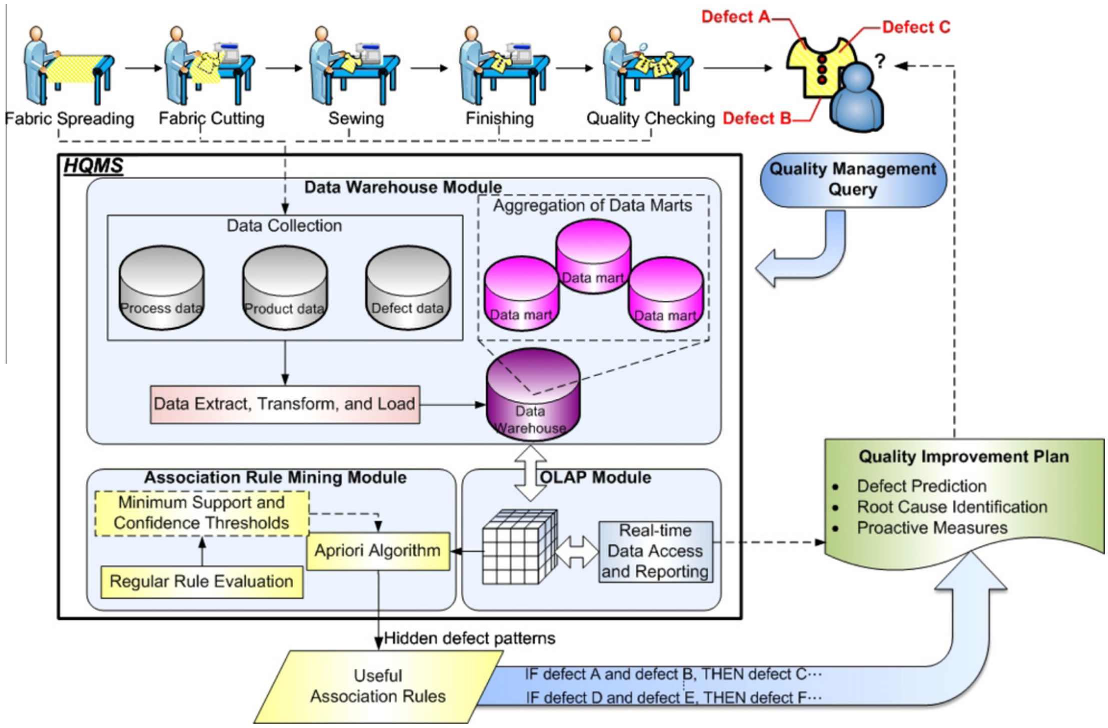


Fig. 2. Hybrid OLAP-association rule mining based quality management system.

in the data warehouse can be viewed in various ways using such functions as slice, dice, roll up, drill down and pivot supported by the OLAP software. This allows users to access data in a desirable way and to report in real-time. Meanwhile, OLAP searches and extracts relevant data from the data warehouse and brings it to the Association Rule Mining Module which is responsible for finding correlations between combinations of garment defects in terms of association rules. Useful association rules obtained are used to formulate quality improvement plans. Details of each module are discussed in the following section.

3.1. Data Warehouse Module

In the Data Warehouse Module, a centralized data warehouse is employed to store and organize data collected from various sources. Since relevant data collected from different departments such as the Fabric Spreading and Cutting Department, the Sewing Department, the Finishing Department and the Quality Control (QC) Department may have different formats, they have to be preprocessed through ETL processes before being stored in the centralized data warehouse. Data are extracted from different departments, and data errors are minimized by verifying data accuracy, correcting spelling errors, completing missing or incomplete entries. They are then transformed to a standard data format so as to prevent problems from occurring in data retrieval and data updating processes. After that, they are loaded into the data warehouse where data are stored in multiple relational tables. A fact table is a table which stores the values of measures or facts. It contains at least one column storing the measure while other columns store facts. These facts are aggregated based on the dimensions of interest which are presented in dimension tables. In order to increase the efficiency of data extraction, individual data marts are created and applied for storing and sharing data within each department. Each data mart has the same structure as the data warehouse, but its stored data are organized according to the corresponding department. The centralized data warehouse is an aggregation of these data marts so that it performs functions of integrating data from various departments and improving data integrity within the company. After the aggregation of the data is completed, users are able to catalogue and directly transfer data in a desirable manner into the OLAP Module.

3.2. OLAP module

In the OLAP module, OLAP is designed as an information system technology for efficiently accessing, viewing and analyzing the data stored in the data warehouse. It acts like a bridge between the data warehouse and the Association Rule Mining Module. There is an OLAP data cube which is n-dimensional where n is the number of dimensions. The measures in the OLAP cube originate from the fact table, while the OLAP dimensions come from the dimension tables. As the dimension tables may contain hierarchical data, different dimension levels are defined and used on different views of the OLAP data cube. Users can view data by drilling up and down in the OLAP cube to suit their needs. In addition, to speed up the query time, some aggregations in the dimension hierarchies are pre-calculated. The multidimensional view of the data cube provides a clear picture to users and allows effective browsing and calculation of the data. Functions of OLAP such as roll up, drill down, slice and dice as well as pivot, make it possible for users to access and turn the stored data into useful information on a real-time basis. However, information displayed by OLAP does not have the ability to give effective suggestions on quality improvement. Hence, the information generated from OLAP have to be transferred to the Association Rule Mining Module for further analysis, which helps to find the correlation between garment defects so as to formulate appropriate quality improvement plans.

3.3. Association Rule Mining Module

In the Association Rule Mining Module, association rule mining techniques are employed to identify groups of items that frequently occur together so as to discover the hidden correlations among items. In the proposed HQMS, an item refers to a garment defect while a frequent itemset refers to a group of garment defects which will frequently occur at the same time. If the generated association rule in this module is "$X \rightarrow Y$", this implies that the garment will contain defect Y if it contains defect X. The rule is said to be useful when both of its support and confidence are greater than the specified minimum thresholds. These thresholds are defined by users based on their needs and the size of the data. In general, the support threshold is set around 30% and the confidence threshold is over 80%. If the minimum support threshold is too low, the number of frequent itemsets and the number of rules discovered will be increased. To avoid trivial rules and inexplicable rules, an evaluation of the rules should be conducted regularly so that the minimum support and confidence thresholds can be adjusted according to the actual production situations.

Useful association rules extracted from the Association Rule Mining Module can be used to formulate quality improvement plans by discovering the hidden patterns of garment defects from the stored data based on the co-occurrence of defects. With the use of the information, managers can perform causal analysis and defect prediction more effectively as well as take proactive measures to achieve better quality of their products.

4. Case study

4.1. Company background

A case study was conducted to demonstrate the practical implementation and validation of the hybrid OLAP-association rule mining based quality management system (HQMS). The case company, Company X, is a Hong Kong-based garment manufacturing company founded in 1978. In recent decades, in order to leverage both the lower land cost and labor cost, Company X has shifted their production-related activities to other low-cost countries while having its headquarters in Hong Kong. Today, its manufacturing capacities are spread all over the world, in countries such as China, Malaysia, Thailand, the Philippines, Vietnam and Bangladesh, producing more than fourteen million pieces of garments annually. However, in order to survive in today's industry, products are expected not only to be low cost, but also to be of good quality. Therefore, the company is striving for quality improvement in its products. Unfortunately, there are two major problems in (i) managing useful quality related data, and (ii) identifying root causes of repeated defects, that hinder the company from improving the quality of its products. Fig. 3 shows three examples of product defects commonly found. The company decided to tackle the problems by having a pilot run of the HQMS in one of its manufacturing plants located in Shenzhen, China.

4.2. Existing problems faced by the company

Company site visits and employee interviews were conducted so as to understand the current practices in the manufacturing plant. Two main problems were observed.

4.2.1. Lack of an integrated system for sharing useful quality-related data

In the company, each department stores data in its own database which is not integrated with databases of other departments. In the usual practice of the company, defect reports, prepared in Microsoft Office Excel, are stored in the database of the QC department after the final inspections of the products. As only limited information is stored in the defect reports, extra effort is always required to identify useful knowledge related to defect problems such as to note which operations created particular defects. These actions consume both resources and time, thus decreasing the efficiency in transforming data into meaningful information for solving problems related to quality.

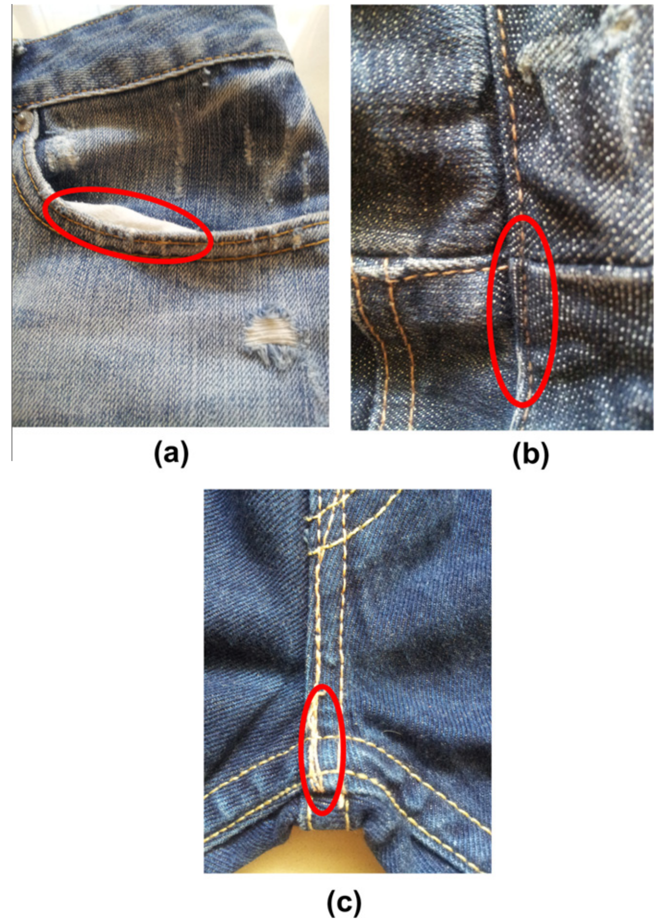


Fig. 3. (a) Sagging pocket. (b) Disappearing stitches, (c) Re-stitched seam.

4.2.2. Lack of a systematic approach to identifying root causes of defect problems

Non-conforming final products may be sent back to appropriate manufacturing departments for reworking. However, rework tasks are performed without any reference to defect reports. Management and individual workers do not have a full picture about the occurrence and hidden patterns of defects, thus fail to prevent the same defect problems from occurring in the future. In some cases, the same quality problems occur frequently as the root causes of defects have not been identified or tackled. Without a systematic approach to analyzing the sources of errors and root causes of defects, it is a challenging task to formulate appropriate strategies for quality improvement.

4.3. Implementation of the HQMS

In order to put the HQMS to the test on a real data set collected from the company, a four-phase implementation was undertaken. The four phases involved in the implementation are (i) Deployment of the Data Warehouse Module, (ii) Deployment of the OLAP Module, (iii) Deployment of the Association Rule Mining Module, and (iv) Formulation of a Quality Improvement Plan. In the first three phases the designed HQMS is constructed while in the fourth phase the output information generated from the HQMS is used to support quality improvement of the final products.

4.3.1. Deployment of the Data Warehouse Module

In the HQMS, Microsoft Access was employed to build the data warehouse. All data related to manufacturing processes and product quality were firstly identified and extracted from different databases. Since different departments may use different data formats in their systems, ETL processes were undertaken to preprocess the data before the data is stored into the warehouse in relational tables as depicted in Fig. 4. Data types, data lengths and primary keys were checked and modified before linking data to establish foreign keys in multiple tables. As shown in Fig. 4, a snowflake schema is designed in the data warehouse. The *Defect Records* fact table contains keys to each of the three dimensions, namely *Product*, *Date*, and *Defect*, with two measures: *Defect Weighting* and *Rework Cost*. *Defect weighting* indicates the seriousness of a defect. The seriousness is classified into three levels: major, moderate and minor, having a weighting of 5, 3 and 1, respectively. *Rework Cost* is the estimated cost for rework activities to tackle the defect. In addition, the *Product* dimension table is normalized by splitting the data into additional tables like the *Order* table. Similarly, the *Defect* dimension table is normalized into the *Cause* table. Keeping the dimension tables in a normalized form can help reduce redundancies and the tables are easier to maintain. In this case study, records on jeans' defects from September 2011 to February 2012 were collected from the QC Department of Company X as the set of sample data.

4.3.2. Deployment of the OLAP module

Microsoft SQL Server 2000 was employed to build the OLAP cube of the HQMS by connecting it to the data warehouse. After being loaded into the data warehouse, data which are crucial for analysis were selected from the data warehouse and imported into the OLAP. The OLAP data cube was built in the OLAP server for real-time analysis and reporting. As the dimension tables contain hierarchical data, data are presented in different hierarchies so as to facilitate a multidimensional view of the OLAP. As all calculations are pre-computed in the OLAP server, users can analyze the data or create reports in real-time, depending on what they need, by choosing the desired dimensions. Fig. 5 shows the preview of the OLAP data cube which allows users to view the aggregated values of the two measures, *Defect Weighting* and *Rework Cost* from all defects stored in the data warehouse. Users are also able to perform roll-up and drill-down functions so as to view the briefly summarized and the most detailed data respectively. For instance, they are allowed to drill down further to view the two measures of each defect type as shown in Figs 5 and 6 show a 3D OLAP cube in which each cell contains the value of the selected measure, *Rework Cost*, corresponding to three dimensions, *Defect*, *Cause* and *Date*. Different values of the measure are marked by different colors, thus it

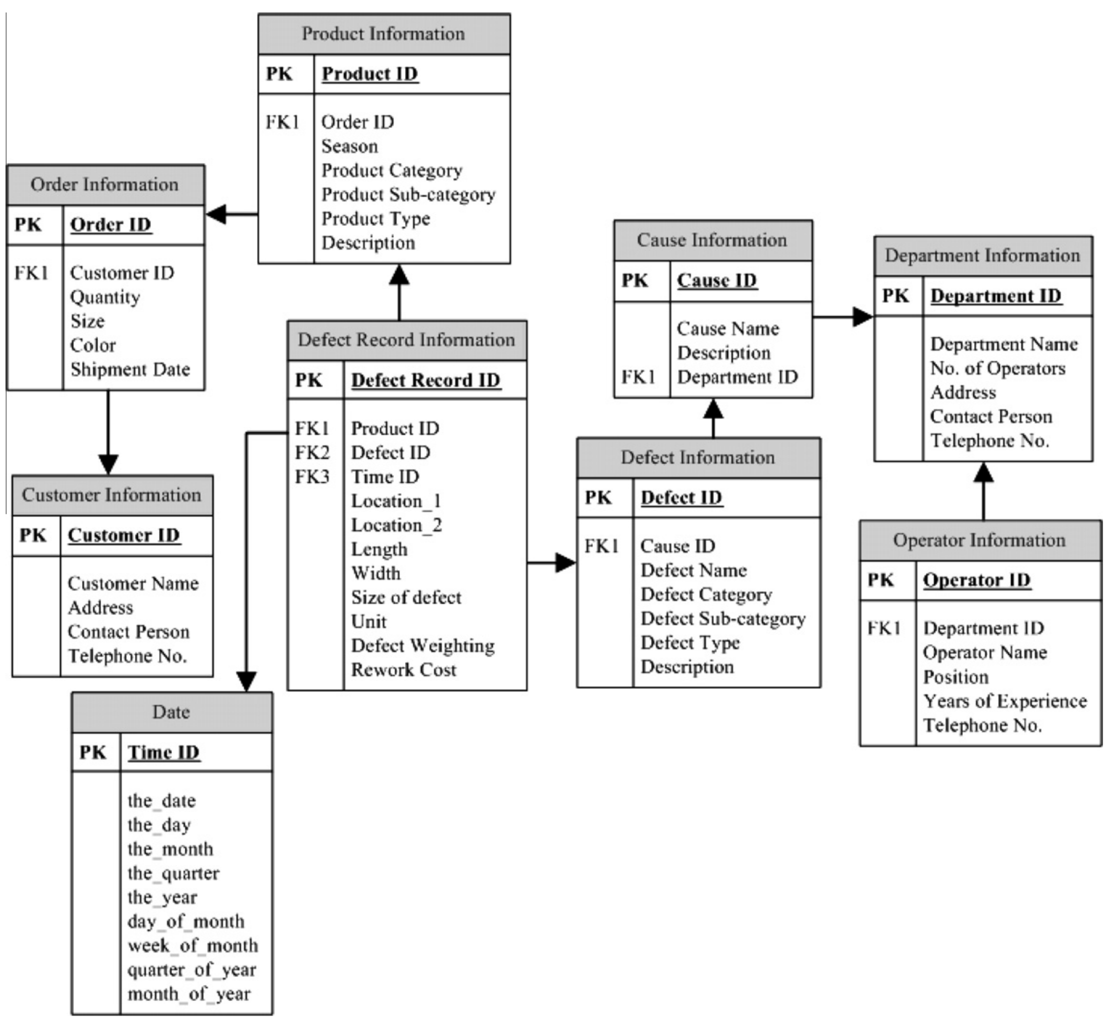


Fig. 4. Relational data warehouse structure.

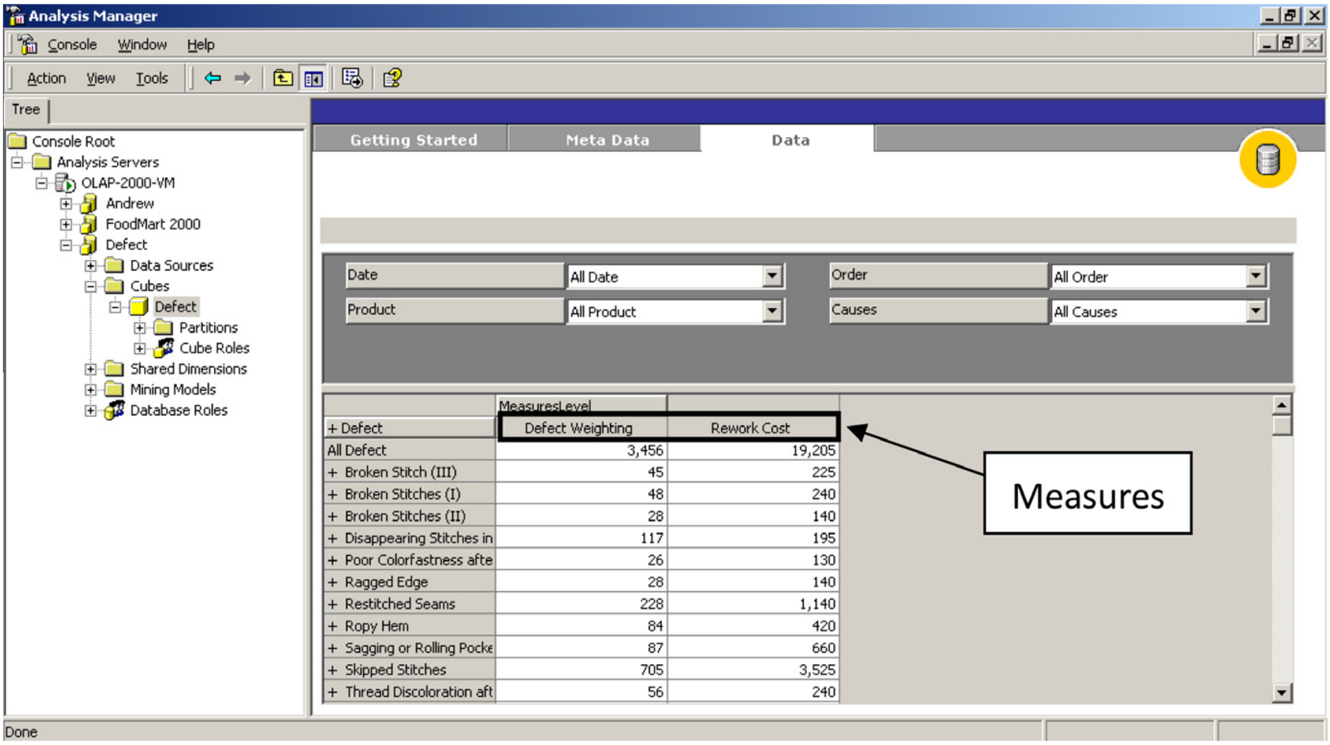


Fig. 5. A print screen of the OLAP cube (preview data cube).

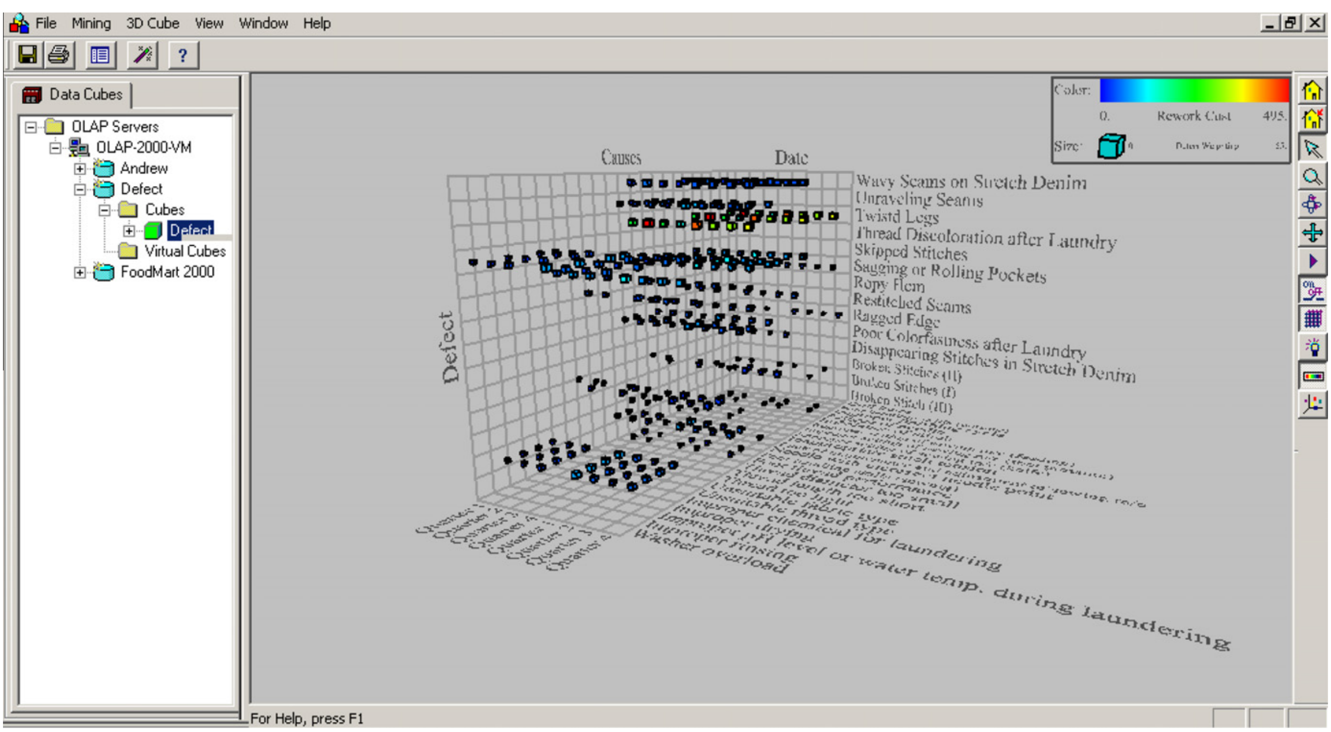


Fig. 6. OLAP 3D cube.

is easy for users to identify any high rework costs from the 3D OLAP cube.

4.3.3. Deployment of the Association Rule Mining Module

The purpose of the deployment of the Association Rule Mining Module is to support quality improvement by identifying the hidden patterns of product defects. In the previous implementation phases, dimensions have been created in the OLAP cube for storing the relevant information of product defects. In this phase, dimensions in the OLAP cube were directly applied to the association rule mining. By choosing the suitable dimension levels, useful association rules can be extracted for quality improvement. In the HQMS,

Table 1
The symbols of different types of defects.

Symbol	Defect
A	Broken stitches
B	Unraveling seams
C	Twisted leg
D	Poor colorfastness after being laundered
E	Sagging pockets
F	Re-stitched seams

Table 2
Extracted defect records for illustration purposes.

Defect record	Product ID	*A*	*B*	*C*	*D*	*E*	*F*
1	005	*A*				*E*	
2	028	*A*	*B*				*F*
3	032	*A*	*B*				*F*
4	058		*B*	*C*		*E*	
5	098						
6	105		*B*	*C*		*E*	
7	110	*A*	*B*	*C*		*E*	*F*
8	153			*C*	*D*		*F*
9	170	*A*	*B*				*F*
10	190		*B*	*C*		*E*	
11	197	*A*			*D*		
12	199	*A*	*B*				*F*

Table 3
Support count and support value of items.

Item	Support count	Support value (%)
A	7	58.3
B	8	66.7
C	5	41.7
D	2	16.7
E	5	41.7
F	6	50.0

Table 4
Support count and support value of 2-itemsets.

2-itemset	Support count	Support value %
AB	5	41.7
AC	1	8.3
AE	1	8.3
AF	5	41.7
BC	4	33.3
BE	4	33.3
BF	5	41.7
CE	4	33.3
CF	1	8.3
EF	1	8.3

Table 5
Support count and support value of 3-itemsets.

3-Itemset	Support count	Support value (%)
ABF	5	41.7
ABC	1	8.3
BCE	4	33.3
BCF	1	8.3
BEF	1	8.3
ABE	1	8.3

Table 6
Support value of the conditions and results of all frequent 3-itemsets.

Itemset	Condition IF	Result THEN	Support	Confidence
ABF	*AB*	*F*	41.70%	100.00%
	AF	*B*	41.70%	100.00%
	BF	*A*	41.70%	100.00%
	A	*BF*	41.70%	71.53%
	B	*AF*	41.70%	62.52%
	F	*AB*	41.70%	83.40%
BCE	*BC*	*E*	33.30%	100.00%
	BE	*C*	33.30%	100.00%
	CE	*B*	33.30%	100.00%
	B	*CE*	33.30%	49.93%
	C	*BE*	33.30%	79.86%
	E	*BC*	33.30%	79.86%

DBMiner 2.0 was used to perform association rule mining. To import the data into the DBMiner software, the OLAP Server was connected to the DBMiner so that the data inside the OLAP were transmitted to DBMiner for the mining function. Appropriate dimension levels can then be selected in accordance with the objectives of the data analysis. In addition, minimum support threshold and minimum confidence threshold levels need to be set for the association rule mining.

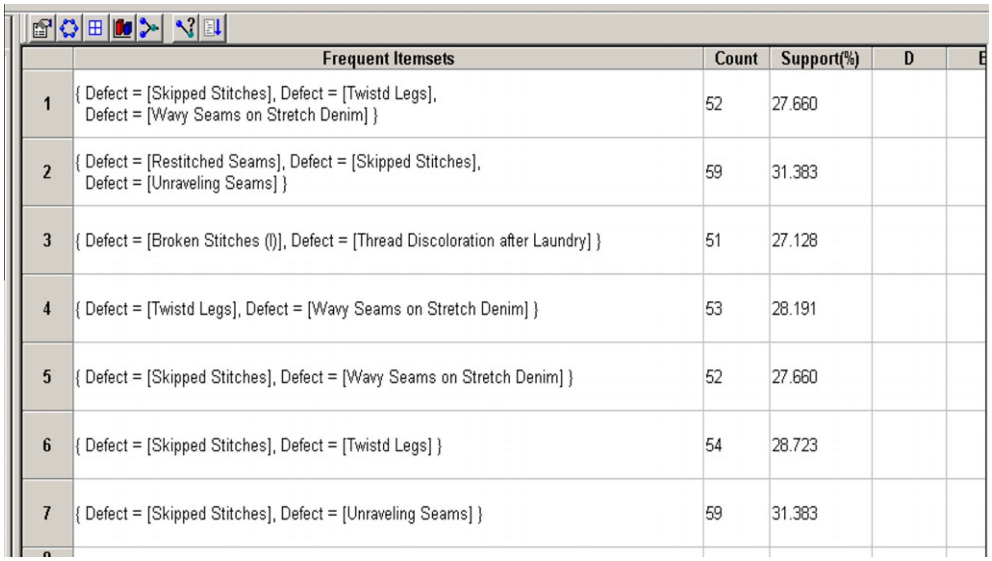
To illustrate the operation mechanism of using association rule mining in HQMS, an example is discussed here to demonstrate how an Apriori algorithm extracts garment defect patterns. Table 1 shows different symbols of various defects while Table 2 shows twelve defect records. In Table 2, each record contains defects found on products during the final inspection done by the QC Department. In the HQMS, only rules with support and confidence levels equal to or greater than the two minimum thresholds are extracted to derive suggestions for quality improvement. Therefore, it is necessary to set the minimum support threshold and minimum confident threshold before association rule mining. They are initially set to 25% and 90% respectively.

In order to discover the hidden association between garment defects, seven steps are followed.

Step 1 Count the frequency of occurrence, i.e., support count, of different items in the record. The support count and the support value of all items are shown in Table 3.

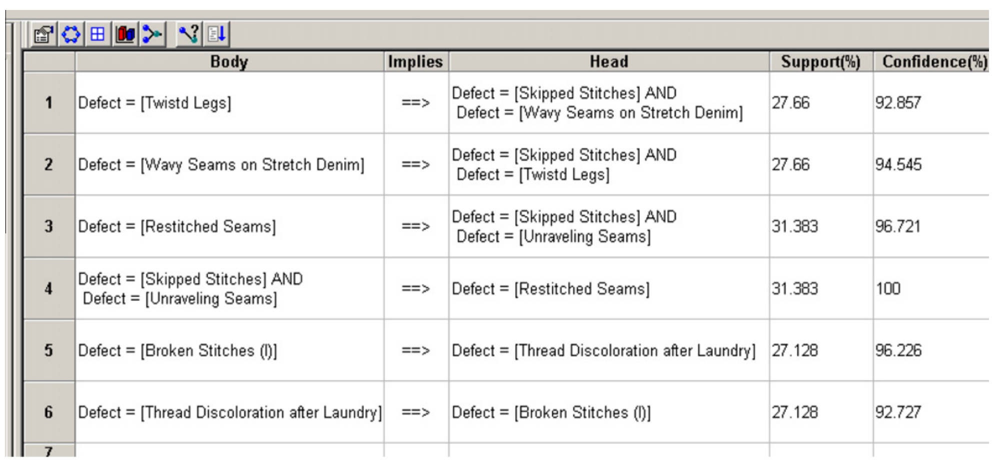
Table 7
Useful association rules in statement form.

Rule 1	
IF	Product defect = Broken stitches AND Product defect = Unraveling seams
THEN	Product defect = Re-stitched seams
Rule 2	
IF	Product defect = Broken stitches AND Product defect = Re-stitched seams
THEN	Product defect = Unraveling seams
Rule 3	
IF	Product defect = Unraveling seams AND Product defect = Re-stitched seams
THEN	Product defect = Broken stitches
Rule 4	
IF	Product defect = Unraveling seams AND Product defect = Twisted leg
THEN	Product defect = Sagging pockets
Rule 5	
IF	Product defect = Unraveling seams AND Product defect = Sagging pockets
THEN	Product defect = Twisted legs
Rule 6	
IF	Product defect = Twisted legs AND Product defect = Sagging pockets
THEN	Product defect = Unraveling seams



	Frequent Itemsets	Count	Support(%)	D	E
1	{ Defect = [Skipped Stitches], Defect = [Twistd Legs], Defect = [Wavy Seams on Stretch Denim] }	52	27.660		
2	{ Defect = [Restitched Seams], Defect = [Skipped Stitches], Defect = [Unraveling Seams] }	59	31.383		
3	{ Defect = [Broken Stitches (I)], Defect = [Thread Discoloration after Laundry] }	51	27.128		
4	{ Defect = [Twistd Legs], Defect = [Wavy Seams on Stretch Denim] }	53	28.191		
5	{ Defect = [Skipped Stitches], Defect = [Wavy Seams on Stretch Denim] }	52	27.660		
6	{ Defect = [Skipped Stitches], Defect = [Twistd Legs] }	54	28.723		
7	{ Defect = [Skipped Stitches], Defect = [Unraveling Seams] }	59	31.383		

Fig. 7. Support counts of the frequent itemsets.



	Body	Implies	Head	Support(%)	Confidence(%)
1	Defect = [Twistd Legs]	==>	Defect = [Skipped Stitches] AND Defect = [Wavy Seams on Stretch Denim]	27.66	92.857
2	Defect = [Wavy Seams on Stretch Denim]	==>	Defect = [Skipped Stitches] AND Defect = [Twistd Legs]	27.66	94.545
3	Defect = [Restitched Seams]	==>	Defect = [Skipped Stitches] AND Defect = [Unraveling Seams]	31.383	96.721
4	Defect = [Skipped Stitches] AND Defect = [Unraveling Seams]	==>	Defect = [Restitched Seams]	31.383	100
5	Defect = [Broken Stitches (I)]	==>	Defect = [Thread Discoloration after Laundry]	27.128	96.226
6	Defect = [Thread Discoloration after Laundry]	==>	Defect = [Broken Stitches (I)]	27.128	92.727

Fig. 8. Useful association rules.

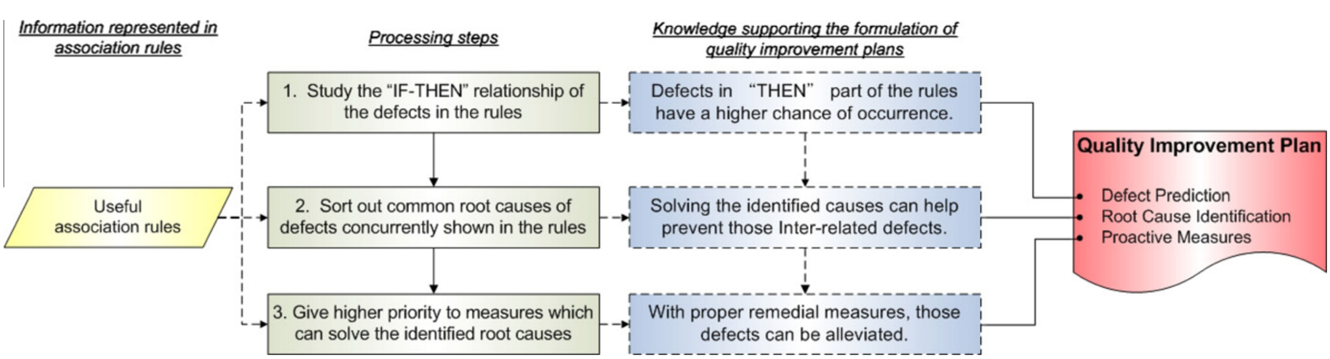


Fig. 9. Procedures for the formulation of a quality improvement plan.

Step 2 Only those items with support value greater than or equal to the minimum support threshold are of interest, and thus others are discarded. In this example, item *D* is pruned off because its support value is less than the minimum support threshold. Items which are not pruned off are called frequent items which are highlighted in Table 3.

Step 3 Frequent items are merged together to generate 2-itemsets by taking pairs. Each support count and the support value are then counted as shown in Table 4.

Step 4 Similar to Steps 2–3, only those itemsets with support value greater than the minimum support threshold are of interest and the remaining itemsets are discarded.

Table 8
Potential causes of the defects.

Defect	Potential causes	Cause ID
Unraveling seams	Poor workmanship (sewing)	001
	Poor thread performance	003
	Lack of maintenance and of adjustment of sewing machines	014
Re-stitched seams	Poor workmanship (sewing)	001
	Poor thread performance	003
	Lack of maintenance and of adjustment of sewing machines	014
Skipped stitches	Poor workmanship (sewing)	001
	Inappropriate stitch tension	008
	Unsuitable thread type	009
	Inappropriate setting of sewing machine (feeding)	015
Wavy seams on stretch denim	Poor workmanship (sewing)	001
	Inappropriate setting of sewing machine (feeding)	015
	Inappropriate setting of sewing machine (foot pressure)	017
Twisted legs	Poor workmanship (sewing)	001
	Poor workmanship (cutting)	002
	Inappropriate setting of sewing machine (feeding)	015
Thread discoloration after being laundered	Unsuitable thread type	009
	Inappropriate pH level or water temperature during laundering	018
	Inappropriate chemicals for laundering	019
	Washer overload	020
Broken stitches (I)	Poor thread performance	003
	Thread diameter too small	004
	Inappropriate pH level or water temperature during laundering	018
	Inappropriate chemical for laundering	019
	Inappropriate rinsing	021
	Inappropriate drying	022

Frequent 2-itemsets, as highlighted in Table 4 are then merged together to generate 3-itemsets. 3-itemsets which contain the pruned parent 2-itemsets will be pruned off automatically due to their low support counts. Table 5 shows the support count and support value of the 3-itemsets. Only two of the 3-itemsets, namely *ABF* and *BCE*, have support values greater than the threshold.

Step 5 After generating the possible garment patterns from the frequent 3-itemsets, possible association rules are formed by combining items in the itemsets in different sequences. There are 12 possible association rules generated as shown in Table 6 where the support and confidence values of each possible rule are also calculated.

Step 6 If both the calculated support and confidence of a rule are equal to or larger than the specified minimum support threshold and the minimum confidence threshold, the corresponding rule is extracted and viewed as a useful rule. In this example, minimum support threshold and the minimum confidence threshold are defined as 25% and 90% respectively. As a result, six useful association rules, as highlighted in Table 6, are obtained.

Step 7 Useful association rules are decoded and expressed in a statement form for ease of understanding. This allows users, especially those lacking data mining knowledge, to understand the meaning of useful association rules. The expression statements of useful rules are shown in Table 7. For example, Rule 1 states that if a product contains *broken stitches* and *unraveling seams* defects, then it will also contain the *re-stitched seams* defect.

In the developed HQMS, the above steps are done by using DBMiner 2.0. *Defect* and *Product* dimensions are selected to do the mining analysis. Items are grouped by the *Product ID* of the

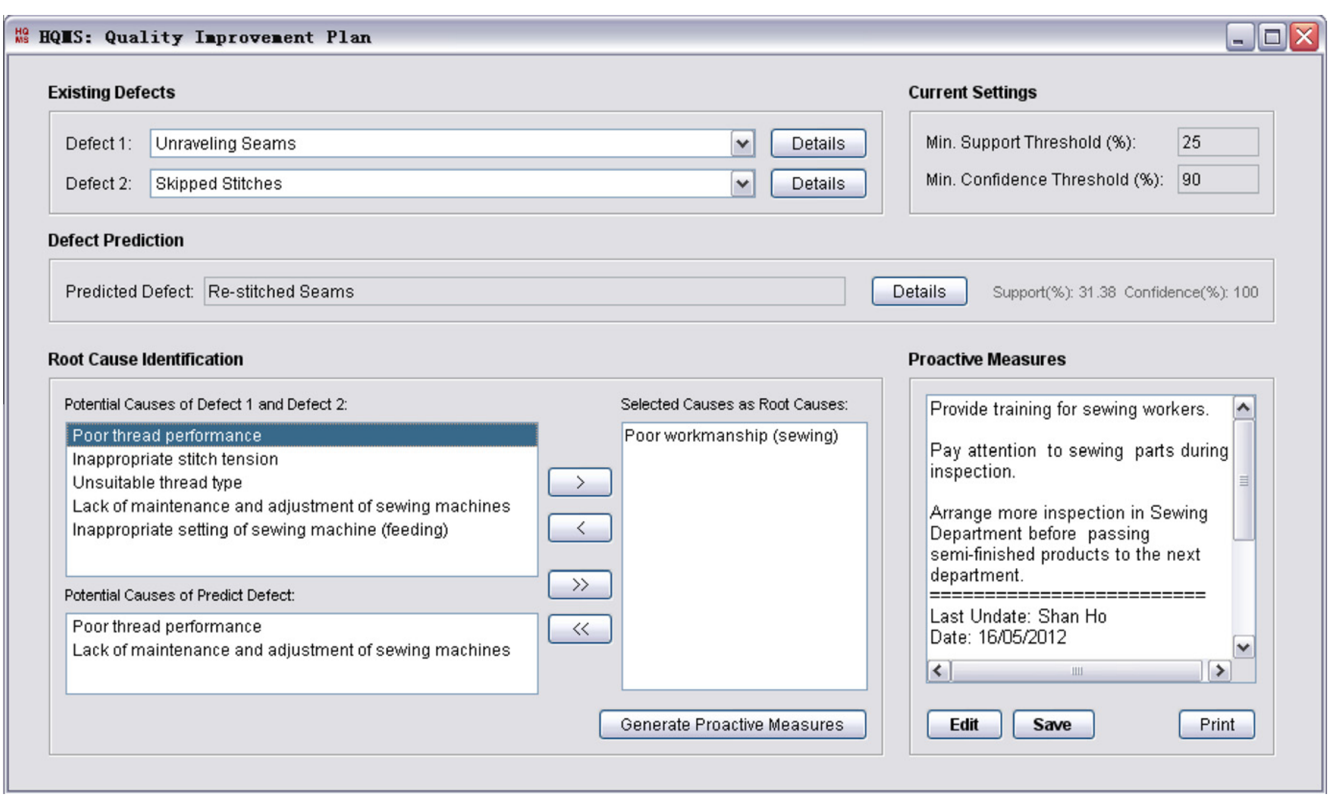


Fig. 10. HQMS interface.

Product dimension to study whether the defects occur concurrently on the same product. The support counts of the frequent itemsets are shown in Fig. 7 and the useful association rules discovered are shown in Fig. 8.

4.3.4. Formulation of a quality improvement plan

The ultimate purpose of using HQMS is to formulate a quality improvement plan. In order to formulate an appropriate improvement plan, those mined association rules generated from the HQMS are used. Suggestions for quality improvement are generated from the defect patterns represented in the rules. There are three formulation procedures, as depicted in Fig. 9, responsible for the transformation of information represented in the association rules into useful knowledge so as to support the formulation of an effective improvement plan.

4.3.4.1. Study the "IF–THEN" relationship of the defects in the rules. In the HQMS, a mined association rule indicates that if certain defects exist, then a particular defect is likely to occur. Therefore, by studying the "IF–THEN" relationship of defects represented in the rules, the Quality Management Team is able to predict potential defects effectively when certain defects have occurred during production. This defect prediction forms an important part of a quality improvement plan. As some garment defects are caused by poor workmanship, it is difficult to predict the resulting defects. Fortunately, the generated association rules can provide knowledge support for defect prediction.

4.3.4.2. Sort out common root causes of defects concurrently shown in the rules. As defects that appear in a single rule are inter-related defects which frequently occur concurrently, the Quality Management Team can sort out their common causes for root cause identification. It is expected that solving the identified causes can prevent such inter-related defects. Table 8 lists the potential causes of seven defects. For instance, according to the mined results in DBMiner 2.0, if the product contains *Unraveling Seams* and *Skipped Stitches*, it is likely to contain *Re-stitched Seams*. With reference to their corresponding possible causes, their common potential cause, namely poor workmanship of sewing workers (Cause ID: 001) is identified as one of the root causes contributing to those defects.

4.3.4.3. Give higher priority to measures which can solve the identified root causes. After root cause identification, a higher priority should be given to proactive measures for tackling the identified root causes. The Quality Management Team can assign more human and equipment resources to solve the identified causes such as the poor workmanship of sewing workers. With proper remedial measures, including the provision of more training for sewing workers and additional zeal on sewing inspection, those targeted defects can be alleviated more effectively. With the knowledge derived from the association rules, it is believed that it is possible to tackle the identified causes and achieve significant improvements in product quality.

Knowledge discovered from the association rules are displayed on the system interface as shown in Fig. 10. The Quality Management Team inputs the existing defects into their query via the interface. Knowledge required for the formulation of quality improvement plans is then suggested. After the implementation of plans, the Quality Management Team evaluates the plans by studying the OLAP real-time reports and observing the improvement in product quality achieved. Feedback is used to confirm if the mined rules reflect the actual production situation, such as the actual garment defect patterns. In order to improve the usefulness of the rules, the minimum support and confidence threshold values are adjusted in accordance with the result of the evaluation. For example, if the rules are found to be trivial or inexplicable, there is a need to fine-tune the existing settings. Since the evaluation is done periodically to keep re-fining the threshold settings, continuous improvement of the result of association rule mining is achieved.

5. Discussion of the HQMS

Conventional approaches to solving quality problems usually require consolidation and analysis of data from diverse sources. This is time-consuming and there is a good chance that human errors will occur. HQMS is designed in such a way that it effectively transforms defect related data into useful knowledge based on the concepts of OLAP and association rule mining. In the HQMS, OLAP functions are supported by the data warehouse which stores data in a multidimensional format. The OLAP is designed for ease of data access and data retrieval allowing users effectively to access data that originated from different sources. In general, enterprises can have both general and detailed views of stored data so as to perform real-time reporting. These reports are useful to managerial issues such as performance evaluation of departments and production lines. In such a time-sensitive industry as the garment industry, functions of real-time access of data are of extreme importance to operational efficiency as well as to survival in the industry. In addition, the mined association rules, output of the HQMS, are used to provide Quality Management Teams with the hidden patterns of defects for identification of potential defects. By identifying the common root causes which are responsible for inter-related defects, the Quality Management Team can give a higher priority to proactive measures of avoiding the identified root causes in their quality improvement plans. The decision support functionality provided by the HQMS can significantly save human and equipment resources and the related planning time as the mined association rules provide clues for the formulation of quality improvement plans. Considering that the minimum support threshold and the minimum confidence threshold will affect the result of association rule mining, mined rules are evaluated regularly to provide feedback on minimum thresholds. In addition, they can be adjusted if necessary to meet the actual manufacturing situation. It is also noticed that the result of association rule mining becomes more reliable when the HQMS are operated for a longer period of time due to the regular rule evaluation for fine-tuning minimum support and confidence thresholds.

6. Conclusion

This paper presents an intelligent system for quality improvement with the integration of data warehousing, OLAP and association rule mining for extracting garment defect patterns. In the usual practice in the garment industry, individual garment defects are solely identified by human inspection without any references to the correlation between defects. Hence, it is difficult to predict defects and take proactive measures for quality improvement. In the HQMS, the data warehouse is used to manage and store the data for data mining while the connection between OLAP and the association rule mining model allows data mining to be applied on a multidimensional basis so that users can gain more in-depth knowledge for defect diagnosis. The garment industry is considered a good example of a labor-intensive industry where product quality is affected by workmanship, making it difficult to predict product defects. Defects which commonly happen concurrently are believed to have certain hidden correlations. Through the implementation of the HQMS, the results show that the proposed methodology effectively extracts hidden relationships among defects based on their co-occurrence, which in turn tell managers which defects are most likely to occur. This allows easy identification of significant root causes of defects, prediction of their occurrence and provides knowledge support in the formulation of effective quality improvement plans. This research study makes a significant contribution by making use of quality-related data to predict product defects and to perform causal analysis in the garment industry. Although every industry has its unique product characteristics, the HQMS' structure presented in this paper, besides being of use in the garment industry, is also applicable to most manufacturing industries, in particular, to some labor-intensive industries where product quality is extremely difficult to maintain. Future research might include real-time capturing of product quality data upon inspection, in order to enhance the efficiency of the system.

Acknowledgment

The authors would like to thank the Research Office of the Hong Kong Polytechnic University for supporting the current project (Project Code: RPXV).

References

Agrawal, R., & Srikant, R. (1994), Fast algorithms for mining association rules in large databases. In *Proceedings of 20th international conference on very large databases*, Santiago de Chile (pp. 487–489).

Altuntas, S., & Selim, H. (2012). Facility layout using weighted association rule-based data mining algorithms: Evaluation with simulation. *Expert Systems with Applications, 39*(1), 3–13.

Baykasoglu, A., Özbakir, L., & Kulluk, S. (2011). Classifying defect factors in fabric production via DIFACONN-miner: A case study. *Expert Systems with Applications, 38*(9), 11321–11328.

Chang, C. P., Chu, C. P., & Yeh, Y. F. (2009). Integrating in-process software defect prediction with association mining to discover defect pattern. *Information and Software Technology, 51*(2), 375–384.

Cheng, Y. M., & Leu, S. S. (2011). Integrating data mining with KJ method to classify bridge construction defects. *Expert Systems with Application, 38*(6), 7143–7150.

Chougule, R., Rajpathak, D., & Bandyopadhyay, P. (2011). An integrated framework for effective service and repair in the automotive domain: An application of association mining and case-based-reasoning. *Computers in Industry, 62*(7), 742–754.

Chow, H. K. H., Choy, K. L., Lee, W. B., & Chan, F. T. S. (2005). Design of a knowledge-based logistics strategy system. *Expert Systems with Applications, 29*(2), 272–290.

Chung, W., & Tseng, T. L. B. (2012). Discovering business intelligence from online product reviews: A rule-induction framework. *Expert Systems with Applications, 39*(12), 11870–11879 (Bill).

Demiriz, A., Ertek, G., Atan, T., & Kula, U. (2011). Re-mining item associations: Methodology and a case study in apparel retailing. *Decision Support Systems, 52*(1), 284–293.

Deslandres, V., & Pierreval, H. (1997). Knowledge acquisition issues in the design of decision support systems in quality control. *European Journal of Operational Research, 103*(2), 296–311.

Ferreiro, S., Sierra, B., Irigoien, I., & Gorritxategi, E. (2011). Data mining for quality control: Burr detection in the drilling process. *Computers and Industrial Engineering, 60*(4), 801–810.

Hsu, S. C., & Chien, C. F. (2007). Hybrid data mining approach for pattern extraction from wafer bin map to improve yield in semiconductor manufacturing. *International Journal of Production Economics, 107*(1), 88–103.

Jukić, N., & Nestorov, S. (2006). Comprehensive data warehouse exploration with qualified association-rule mining. *Decision Support Systems, 42*(6), 859–878.

Köksal, G., Batmaz, İ., & Testik, M. C. (2011). A review of data mining applications for quality improvement in manufacturing industry. *Expert Systems with Applications, 38*(10), 13448–13467.

Kusiak, A., & Smith, M. (2007). Data mining in design of products and production systems. *Annual Reviews in Control, 31*(1), 147–156.

Lau, H. C. W., Ho, G. T. S., Chu, K. F., Ho, W., & Lee, C. K. M. (2009). Development of an intelligent quality management system using fuzzy association rules. *Expert Systems with Applications, 36*(2), 1801–1815.

Lau, H. C. W., Wong, C. W. Y., Hui, I. K., & Pun, K. F. (2003). Design and implementation of an integrated knowledge system. *Knowledge-Based System, 16*(2), 69–76.

Lim, A. H. L., Lee, C. S., & Raman, M. (2012). Hybrid genetic algorithm and association rules for mining workflow best practices. *Expert Systems with Applications, 39*(12), 10544–10551.

Lou, H. H., & Huang, Y. L. (2003). Hierarchical decision making for proactive quality control: System development for defect reduction in automotive coating operations. *Engineering Applications of Artificial Intelligence, 16*(3), 237–250.

Mak, K. L., Peng, P., & Yiu, K. F. C. (2009). Fabric defect detection using morphological filters. *Image and Vision Computing, 27*(2), 1582–1592.

Martínez-de-Pisón, F. J., Sanz, A., Martínez-de-Pisón, E., Jiménez, E., & Conti, D. (2012). Mining association rules from time-series to explain failures in a hot-dip galvanizing steel line. *Computers and Industrial Engineering, 63*(1), 22–36.

Milne, R., Drummond, M., & Renoux, P. (1998). Predicting paper making defects on-line using data mining. *Knowledge-Based Systems, 11*(5–6), 331–338.

Moreno, M. N., Ramos, I., García, F. J., & Toro, M. (2008). An association rule mining method for estimating the impact of project management policies on software quality, development time and effort. *Expert Systems with Applications, 34*(1), 522–529.

Pardillo, J., Mazón, J. N., & Trujillo, J. (2010). Extending OCL for OLAP querying on conceptual multidimensional models of data warehouses. *Information Sciences, 180*(5), 584–601.

Thalhammer, T., Schrefl, M., & Mohania, M. (2001). Active data warehouses: Complementing OLAP with analysis rules. *Data and Knowledge Engineering, 39*(3), 241–269.

Tsai, T. N. (2012). Development of a soldering quality classifier system using a hybrid data mining approach. *Expert Systems with Applications, 39*(5), 5727–5738.

Ur-Rahman, N., & Harding, J. A. (2012). Textual data mining for industrial knowledge management and text classification: A business oriented approach. *Expert Systems with Applications, 39*(5), 4729–4739.

Wong, W. K., Yuen, C. W. M., Fan, D. D., Chan, L. K., & Fung, E. H. K. (2009). Stitching defect detection and classification using wavelet transform and BP neural network. *Expert Systems with Applications, 36*(2), 3845–3856.

Yang, G., Mabu, S., Shimada, K., & Hirasawa, K. (2011). A novel evolutionary method to search interesting association rules by keywords. *Expert Systems with Applications, 38*(10), 13378–13385.

Yuen, C. W. M., Wong, W. K., Quan, S. Q., Chan, L. K., & Fung, E. H. K. (2009). A hybrid model using genetic algorithm and neural network for classifying garment defects. *Expert Systems with Applications, 36*(2), 2037–2047.