Adel Al-Shayea
Emad Abouel Nasr ⊠
Hisham Al-Mubaid
Abdulrahman Al-Ahmari
Ali K. Kamrani
Husam Kaid
Haitham A. Mahmoud

https://doi.org/10.21278/TOF.454025420
ISSN 1333-1124
eISSN 1849-1391

A NEW ASSOCIATION ANALYSIS-BASED METHOD FOR ENHANCING MAINTENANCE AND REPAIR IN MANUFACTURING

Summary

Maintenance and quality of products are absolutely crucial for any organization to succeed in the industrial and manufacturing engineering. Current research studies have confirmed the presence of a high correlation between these two factors, namely maintenance and quality of products, in industrial organizations. Nevertheless, no extensive research has been conducted in order to study the link between maintenance and the quality of products in manufacturing. In this paper, we conduct a study in this domain and examine the relationship patterns between maintenance and the quality of product using manufacturing data on maintenance and the product quality. Specifically, we employ association analysis and association rule mining with large and extensive sets of product quality, repair, and maintenance data. Our main objective is to discover interesting and non-trivial associations for feature failure resulting in the repair or maintenance of a product with unapproved quality. The results of evaluation are quite interesting. The resulting association rules with high values of confidence and lift suggest some essential associations between the product features and the failure; such findings have not been known and used before. This can help quality engineers and maintenance teams to enhance maintenance and repair operations and lower the overall cost of manufacturing.

Key words: *Association Rule Mining in Manufacturing, Data Mining in Manufacturing, Quality Control, Maintenance and Repair in Manufacturing*

1. Introduction

It has been discovered and confirmed that maintaining a high quality of the product and lowering the cost of production are essential tasks and the main objectives of successful manufacturing organizations. In such large manufacturing organizations, the most crucial factors affecting the success and prosperity of the organization are the maintenance and repair operations. One of the most viable ways to keep the organization up to date and competitive in the market is to adopt efficient techniques for utilizing the data and information collected during the manufacturing process. With the current highly advanced systems used in manufacturing, most organizations can collect volumes of data and information about all aspects of operation and production.

Maintenance can be defined as a combination of activities required to monitor, control and supervise a system to perform its intended functions. The goal is to restore the system to a state in which the predetermined task can be carried out with a reduced risk of equipment breakdown and increased availability of production systems. Maintenance tasks are divided into preventive and corrective categories [1,73,74].

Preventive maintenance is a planned activity (time-line plan) which is performed on the basis of a given schedule. The goal is to decrease the chance of sudden unexpected breakdowns by making sure the machines have a good operational status. The probability of unplanned repair and the need for corrective maintenance is not completely eliminated by preventive maintenance. In order to minimize the probability of breakdowns, the plan implements the required maintenance [2-5]. When there is a machine failure, corrective maintenance is a policy that is carried out. It is unplanned and the goal is to restore the machine functionality and resume production. In Condition-Based Maintenance (CBM), activities are performed as a response to a specific condition. The ability for predicting breakdowns is a major feature of CBM [6, 7]. Physical conditions are determined using sensors or other means (e.g. vibration or temperature) of monitoring [8-11]. To restore the machine to the desired level, maintenance shall be carried out once the machine reaches one or more predetermined levels of condition. CBM is conducted on the basis of particular critical signals before any failure, and therefore it is considered as preventive and not corrective maintenance [12-14].

While production engineers are more interested in the availability of resources, maintenance is treated as a cost centre and is used for the required budgeting by the budgeting office. It is not an easy task to measure maintenance performance and require data that may not always be accessible [15, 16]. The most common indicators of maintenance performance are costs and efficiency. Efficiency is measured in terms of total production time, downtime, and number of failures. For the analysis of its economic value, the total maintenance cost is used [17-20]. Determining the maintenance cost requires the manager to collect information such as expected life, frequency of repairs required, unplanned repair costs, corrective maintenance, and equipment replacement cost [21-23]. In this paper, a plan is to propose a model to calculate the total expected maintenance cost and to apply it for selecting the best policy. Due to the loss of sales and high costs of unplanned repairs, the impact of production system downtime is significant. For example, the estimated downtime costs of US automobile plant assembly lines are as much as \$10,000-15,000 per minute, while the estimated hourly costs on an off-shore platform are \$20,000 to 25,000 per hour [1, 24-26]. The oil and gas industry is continuously in search of "premium" technologies to enhance the life-span of their strategic components. The industry is mainly driven by performances, sustainability, and downtime reduction [27-30, 75-77]. Maintenance is a crucial function for maintaining resources for producing products of high quality [31-33]. Studies have seldom been focused on how the interaction between quality and maintenance is used for the implementation of CBM systems. Most of the research relates to the quality control or total quality management tools, without any direct correlation with the quality and maintenance of the actual product [26, 34-36].

Typically, in a manufacturing organization, the data on product quality, product defects, maintenance, failures, repairs, and breakdowns are continuously collected [37]. To utilize these massive volumes of data collected over several years, the need for computational and data analysis techniques is of extreme significance for the organization to improve its productivity and benefit from it. Data mining can help significantly in this domain [2]. Data mining is a branch of science that includes methods which enable us to work with data and utilize it so that knowledge, interesting patterns, and non-trivial relations can be extracted from it [37 – 43]. As one of vital techniques of data mining, Association Rule Mining (ARM) can be used for extracting relevant, important and interesting relationships from given transaction data [38, 39]. In this paper, we are interested in applying association analysis and association rule mining in the manufacturing domain for product quality and maintenance [40, 41]. In fact, to prosper and remain successful, maintenance and the quality of products are of utmost importance to any successful manufacturing organization. Moreover, the presence of a high correlation between these two aspects of any industrial organization, namely the product quality and maintenance, has been confirmed. However, no extensive research analysing and investigating the relationship between maintenance and the quality of various products in manufacturing has been conducted [42, 43, 71,72].

An integrative preventive maintenance approach that considered the quality loss while a manufacturing system is in operation was presented by Piatetsky-Shapiro [44]. Functional risks in the manufacturing system were considered in the first stage, and later, the degradation path of machine performance was also included. In the end, the optimal maintenance strategy was presented and verified through a numerical example. In another study, conducted by He, Gu, and Chen, an integrative preventive maintenance method was presented; it included both the quality control and mission reliability analysis [45]. The results revealed that the suggested methodology attained approximately 26.04% of cost savings.

To cope with the contemporary intelligent manufacturing systems, a maintenance strategy was proposed in [46]. This methodology has four stages; in the first stage, the operating mechanism was characterised by including mission reliability. In the second stage, a quantitative model which includes maintenance resources and real-time quality data was developed. In the third stage, a decision model for a multi-state manufacturing system was developed with constraints. Finally, in the fourth stage, particle swarm optimization was employed to solve that optimization problem, and the developed methodology was verified using a numerical example. Fault diagnosis is a vital step in the maintenance process, and it is always better to make it at initial stages. A methodology based on fuzzy data envelopment analysis was developed for root causes of infant failure [47]. To verify the efficacy of the proposed methodology, a typical infant failure of the vibration and noise of a washing machine problem was presented. A real-time maintenance strategy was proposed for deteriorating machines [48]. The objective was to select the optimal maintenance level for a machine, and hence reduce the overall maintenance cost. The model considered both the resource cost and the cost incurred due to stoppage.

In this paper, we conduct research into this area to investigate the correlation between the quality of product and maintenance by using data on maintenance and the product quality. We use Association Rule Mining (ARM), specifically, with product quality data in large data sets including repair and maintenance data. The association rule analysis has been applied and proved successful with various problems in the field of manufacturing, see for example [28, 38, 39, 42-44, 49]. However, in the context of maintenance/repair and product quality, we could not find studies related to ARM. Therefore, the main objective of this study is to discover some interesting and non-trivial association rules for feature failure in the presence of a product with unapproved quality that might require repair or maintenance. This objective will significantly improve the maintenance and repair processes and reduce the overall production costs. Thus, the main contribution of this work is twofold:

1. examining the relationships between the quality of product and maintenance, and
2. discovering associations of feature failure, which result in repair and maintenance.

The evaluation results and outcomes of the proposed methodology are very promising in producing highly reliable information about the product quality and maintenance. The resulting association rules suggest some vital associations between the product features causing the failure with fairly high confidence and lift values; such findings have not been known and utilized before. Improving maintenance and repair and reducing manufacturing costs can assist maintenance and quality engineers.

2. Previous Work and Literature Review

The subject of data mining is extremely beneficial in several application areas and in many branches of science [50]. Production processes, control, maintenance, decision support systems, quality improvement, failure detection, maintenance, and engineering design are some examples of the many manufacturing areas in which data mining is used [2, 42]. For instance, data mining has assisted in Computer Aided Design (CAD) elements in search and retrieval operations [51, 52].

Previously, data mining was referred to as knowledge engineering (also, sometimes, knowledge management). It has been an elusive and unusual technology until recently, explored more by theoreticians in the fields of artificial intelligence. In the study [53], data mining is described as a step in the Knowledge Discovery in Database (KDD) process involving the application of computational techniques which, under appropriate computational efficiency constraints, generate a specific list of patterns or models over the data. A more general definition that is currently used by many researchers was presented by Adriaans and Zantinge in [54]. They stated that data mining is a process of searching for unknown patterns or trends through details of data. They highlighted and stressed the importance of possessing an efficient search method in vast amounts of data until a sequence of patterns appears, whether complete or in an acceptable probability.

Very large databases are frequently searched for relationships, trends, and patterns, which are neither known nor detectable prior to the investigation. Engineers and marketers usually assume that these relationships or trends exist. However, a proof can only be provided by the data itself. The new knowledge or information enables the community of users to be better at what it does. An issue that often occurs is that very few facts that are searched for in very large databases will give us the necessary information. Furthermore, when researching a new trend or pattern, the algorithm and the search criteria used in a single database can change and each database may require a different search criterion as well as new algorithms which can be adapted to the new data conditions and problems. Often, understanding and visualizing large data sets proves to be rather difficult to humans. In addition, as Fayyad and Stolorz stated in [55], data can grow in two dimensions defined as the number of fields and the number of cases for each one of these fields. The way they explained this is that human capacity for analysis and visualization does not extend to high dimensions and massive data volumes. Also, the growth rate of data sets far exceeds the conventional rates that 'manual' analytical techniques can cope with. This fact is the second factor that further increases the importance of data mining, making it a necessity. This means that a normal method used by a company to extract knowledge from a database will leave large amounts of data unsearched because the data growth exceeds traditional mining procedures. Consequently, this leads to the need for a technology that will enable people to tackle the problem by using large amounts of data without disregarding or losing valuable information that can help to solve any type of problems involving large data sets. In the study [56] it was stated that “data mining is asking a processing engine to show answers to questions we do not know how to ask”. The author clarified that the aim of data mining is to find similar trends that will somehow address the desired questions raised by the engineers or marketers, instead of asking a direct question about a singular occurrence on a database in the standard query language. If the questions that are asked or relationships that are looked for are too explicit in a database, the process will be more complex and time-consuming. Moreover, several vital relationships may be missed or disregarded. Several enterprises have developed infrastructural databases that contain data about their products and processes. Potentially, these databases were a gold mine containing terabytes of data with much "hidden" information that was not easy to comprehend. Machine-learning methods have been developed thanks to the great advances made by artificial intelligence researchers. Knowledge extraction from large databases has now been made easier and more productive than ever due to genetic algorithms, neural networks, and other appropriate learning methods. Data mining is an iterative process [57]. As the process progresses, new knowledge and new hypotheses should be developed in order to adapt to the quality and content of the data. In order to further elaborate, the quality of the data being studied will determine the time and precision of any given data mining algorithm, and important information about a problem will be found if the flexibility of the algorithm is sufficient, even if the central question has not been answered in full.

Some researchers (e.g. Purarjomandlangrudi [37]) have proposed an approach to data mining that employs an anomaly detection (AD) learning algorithm in order to detect the anomalies in early stages. In the proposed anomaly detection algorithm, Kurtosis and Non-Gaussianity Score (NGS) are the two features used to diagnose any early defects of a system. To lower the total number of maintenance visits, Ullrich, ten Hagen, and Lassig [58] suggested a data mining approach by employing group maintenance visits based on predictive maintenance. In this maintenance strategy, the repair recommendations for an ensemble of pieces of equipment located close to each other are combined into one maintenance visit and consequently, the total number of maintenance visits is reduced compared to a reactive strategy. Two approaches were proposed in [59] for the purpose of modelling the relationship between quality and maintenance. Since maintenance affects the equipment failure pattern, the first approach is based on imperfect maintenance. The second approach is based on an approach by Taguchi. A data-driven software reliability model (DDSRM) was developed by Yang, Li, Xie, et al. in [5]; DDSRM considers the software failure process as a time series process. However, unlike previous DDSRMs, there is a disregard of the non-realistic assumption that a software failure has a strong correlation with the latest failures. A failure prediction decision support model was proposed in [60] for the purpose of predictive maintenance of a critical machine component by the use of the condition monitoring data. Several data mining techniques were tested to identify upcoming failures, and the best technique was identified among them. Moreover, the developed model was compared with the proposed model of Advanced Semiconductor Materials Lithography (ASML) and the developed model was shown to increase the availability of the system and decrease the (associated) costs. In the study [61], visualization and data mining for condition-based maintenance was utilized. In the proposed approach, reoccurring patterns are recognized; this helps decision makers determine evolving technical problems and take proper counter measures. In order to recognize failures of process units, the authors of the study [62] address the application of probabilistic networks by the use of data mining approach and by extracting knowledge from databases. Previously, in order to extract data from data sets, Vaz et al. [57] employed software based on the Principal Component Analysis (PCA), hierarchical classifiers and prototype methods. The authors used hierarchical classifiers to diagnose irregular operations and found that there is a connection between the performance of the method and the extent of the failure and its localization. Additionally, to better define failure types, they developed a hybrid approach which combines hierarchical classifiers and prototypes. In the study listed under [2], co-occurrences were responsible for the proposition of a methodology. The following move was to employ such co-occurrences in order to extract temporal association rules resulting in unusual goal events requiring immediate corrective maintenance actions. Ciflikli and Kahya-Ozyirmidokuz [49] developed a knowledge induction model based on the C4.5 decision tree algorithm, which was used to detect the failure of the process in a carpet manufacturing firm located in Turkey. The authors used pre-processing techniques to enhance the quality of data. There is an application for the Binarization approach to the target data and for a variable reduction attribute relevance analysis with the use of information gain. Sun, Wang, Wang et al. [63] studied the causes of rupture in the water supply networks; they attempted to identify high-risk areas in water supply networks and demonstrated that the data mining of spatial pipe networks while changing the initial parameters can be very beneficial in the analysis of spatial clusters of bursting pipes. Iwata, Tsumoto, and Hirano [64] employed a temporal data mining process (similarity-based visualization approach) for the maintenance and construction of clinical pathway. Through overusing clustering, they were able to define the temporal dimensions of nursing orders and find the missing information in the clinical pathway. Using test data collected from life cycle tests, Li, Wu, Wang, et al. [31] measured the capacity, resistance, and the life cycle of lithium iron phosphate batteries of products. Furthermore, the data mining method was used for pattern recognition, and the battery reliability was estimated using cluster analysis; as a result, a suggestion was put forward for a strategy for the improvement of reliability based on a statistical analysis and the study of the macro mechanism of product failures. Based on a small sample and the data mining of failure-physics, Zhou, Liu, Jin, et al. [65] developed a new reliability estimation method for some special areas such as aerospace and military. In their study, Momentum Wheel (MW) is taken as an example of how to apply the method step by step; the method was verified through the reliability estimation of the MW of a satellite and physical experiments. To develop a predictive maintenance system, Bastos, Lopes, and Pires [66] presented a new conceptual framework for discovering failure patterns with the goal of early detection of faults in machines; it is characterised by autonomy in data collection and the use of the knowledge gathered from data. A study done by He, Han, Gu, and Chen [69] also presents the importance of predictive maintenance as compared to reactive maintenance.

3. Proposed Methods and Techniques

3.1 Background

Data mining is also called Knowledge Discovery in Database (KDD); some consider data mining as an analysis step in KDD, but it is actually the heart of KDD [67]. Data mining has been one of the fastest growing research areas in the last two decades. In general, data mining deals with discovering important knowledge and interesting relations and patterns and extracting them from large data sets in all application areas. Association Rule Mining (ARM) is one of the important areas in data mining which is concerned with extracting useful association rules and patterns from large data sets [49, 63].

A set of transactions T = $\{t_1, t_2, .., t_n\}$ given so that each transaction t_i consists of items chosen from a data set I of all available items $I = \{ i_1, i_2,.. , i_m \}$, where n is the total number of transactions and m is the total number of items, constitutes a model of data.

In data mining, this model of data is commonly known as market basket analysis data [68]. As shown in Table 1, each such transaction t_i may or may not contain a particular item, and so the market basket analysis data can be represented as a binary matrix. The data in Table 1 includes five items (columns) with four transactions (rows). Each item in a transaction data set is represented in a column, and each row represents one transaction t_x with all items in t_x having a value of 1.

For example, in a common market basket analysis task, the data represent shopping transactions at a supermarket where products may be juice, cake, milk, etc. Also, the data can be features of certain products or parts, e.g., spare parts for certain machines, produced by a given manufacturing plant, a cutting machine, or an assembly line in an industrial plant. From the set I of all items, the collection of k items is called k-itemset. Each transaction can include any combination of the k items. In the analysis of association rules, the *support count* $\boldsymbol{\sigma}$ of an itemset is defined as the total number of transactions that contains a particular itemset as follows:

$$\sigma\ (X) = |t_i\ and\ X \in\ t_i\ |$$

In Table 1, for example, the itemset $\{i_1, i_3\}$ occurs three times and therefore

$$\sigma(\{i_1, i_3\}) = 3;$$

$$\text{and } \sigma(\{i_2, i_3\}) = 2.$$

We note here that the itemset {*i1*, *i2*} is called a *2*-itemset and the itemset {i_4} is a *1*-itemset. Thus, we can generate five 1-itemsets (see Algorithm 1) from Table 1. In association rule mining, an association rule is defined in the following expression:

$$X \rightarrow Y,$$

Table 1 Example of transaction and item data for the market basket analysis where the data are in binary format.

Transaction Id	i_1	i_2	i_3	i_4	i_5
t_1	1	0	1	0	1
t_2	0	1	1	0	0
t_3	1	1	1	1	1
t_4	1	0	1	1	0

where the operator → is an *implication* operator indicating that the presence of the itemset *X* in a transaction implies the occurrence of the itemset Y; and X and Y are disjoint sets, that is:

$$\boldsymbol{X} \cap \boldsymbol{Y} = \emptyset$$

In the rule *X* → *Y*, the itemset *X* is called the *antecedent* and *Y* is known as the *consequent.* A simple example of association rule is

$$\{\boldsymbol{i_x}\} \rightarrow \{\boldsymbol{i_y}\}$$

which is interpreted as $\{\boldsymbol{i_x}\}$ $\boldsymbol{implies}$ $\{\boldsymbol{i_y}\}$; it refers to those transactions that have i_x and i_y. In this scenario, one of the main goals is to extract and quantify all important and most significant rules in the transaction data. In the data mining research, we use three main metrics (or measures) to identify significant association rules: *Support*, *Confidence*, and *Lift*. The *Support* (*s*) for the rule *X* → *Y*, denoted as *s(X* → *Y)*, is defined as the fraction of all transactions in the data set that contain the itemset XUY. This means, the fraction of transactions comprising all products from both X and Y is defined as:

$$\boldsymbol{s}(\boldsymbol{X} \rightarrow \boldsymbol{Y}) = \frac{\boldsymbol{\sigma}(\boldsymbol{X} \cup \boldsymbol{Y})}{\boldsymbol{N}} \tag{1}$$

As a matter of fact, we would like to know how important a rule is in a given transaction data set by calculating how often the items of the rule occur in the data set.

Algorithm 1: Determining and generating frequent itemsets:

```
q = 1
Find all frequent 1-itemsets: { x | σ(x)/N > threshold}
repeat
    q = q + 1
    generate all candidate itemsets Cq:
    generate q-itemset from (q-1)-itemsets
    prune q-itemset based on min support threshold mst
    For each transaction t in T do
        Recognize all contenders that belong to t: Ct=subset(Cq, t)
        For each contender itemset c belong to Ct do
            σ(c) = σ(c) + 1
        end for
    end for
    Extract the frequent k-itemset F_q = {c | σ(c) > threshold}
until F_q = ∅
return ∪ F_q
```

For example, in Table 1, the rule $i_1 \rightarrow i_3$ has a *support* of ¾ {*i.e.,* s($i_1 \rightarrow i_3$)= 0.75} because three of the four transactions (*t1*, *t3*, *t4*) contain both i_1 and i_3; moreover, s($i3 \rightarrow i4$) is 0.50, as can be seen from Table 1. Therefore, by *support* only, we conclude/estimate that the rule $i1 \rightarrow i3$ is more significant than $i3 \rightarrow i4$ in this data set. Notice here that the *support* is a symmetric operation: s($X \rightarrow Y$) = s($Y \rightarrow X$). In the above example, the support s($i_1 \rightarrow i_3$) = s($i_3 \rightarrow i_1$) = 0.75.

The *Confidence* $c(X \rightarrow Y)$ of the rule X→Y is the number of transactions containing the itemset X that also contains the itemset $\boldsymbol{Y}$.

In other words, the confidence c(X→Y) calculates from (among) the transactions containing *X* how many transaction also contain *Y*; it is defined as follows:

$$\boldsymbol{c}(\boldsymbol{X} \rightarrow \boldsymbol{Y}) = \frac{\boldsymbol{\sigma}(\boldsymbol{X} \; \boldsymbol{U} \; \boldsymbol{Y})}{\boldsymbol{\sigma}(\boldsymbol{X})} \tag{2}$$

This implies that the *confidence* estimates the rule accuracy in terms of the support count of the itemsets in *X* and *Y* among the transactions containing the itemset X. For example, in Table 1, the rule $i_1 \rightarrow i_2$ confidence is 0.33 whereas the confidence of $i_1 \rightarrow i_3$ is 1.0; thus, the rule $i_1 \rightarrow i_3$ is more significant in terms of *confidence* (has more confidence) than $i_1 \rightarrow i_2$. The *confidence* operation is asymmetric:

$$\mathbf{c}(\boldsymbol{X} \rightarrow \boldsymbol{Y}) \neq \mathbf{c}(\boldsymbol{Y} \rightarrow \boldsymbol{X}).$$

In statistics, the confidence is basically a conditional probability of the consequence conditioned by the antecedent; that is $\boldsymbol{c}(\boldsymbol{X} \rightarrow \boldsymbol{Y}) = \boldsymbol{P}(\boldsymbol{Y}|\boldsymbol{X})$.

All rules which have the support and confidence above certain threshold values are extracted in association rule mining. Another useful and reliable measure for association rules is *Lift* (*l*) which refers to the ratio of the confidence of a rule $X \rightarrow Y$ to the *support* of the consequent defined as:

$$\boldsymbol{l}(\boldsymbol{X} \rightarrow \boldsymbol{Y}) = \frac{\boldsymbol{c}(\boldsymbol{X} \rightarrow \boldsymbol{Y})}{\boldsymbol{s}(\boldsymbol{Y})} \tag{3}$$

This means that the lift estimates the confidence of a rule divided by the expected confidence if the antecedent and consequent are independent; here, the expected confidence is measured in terms of the *support* of the consequent [6]. Notice that the support of *Y* *s(Y)* in statistics is the simple probability of *Y* in the data set:

$$\boldsymbol{s}(\boldsymbol{Y}) = \boldsymbol{P}(\boldsymbol{Y}$$

From Table 1, the confidence c($i_1 \rightarrow i_4$) of the rule {$i_1 \rightarrow i_4$} is 0.67 and the support of the consequent i_4 is 0.50; therefore, the *lift* l($i_1 \rightarrow i_4$) is 0.67/0.50 = 1.34. Let us take another example: for the rule {$i1 \rightarrow i2$} in Table 1, the lift $l(i1 \rightarrow i2)$ is 0.3/0.5 = 0.6. Notice that the *lift* operation is symmetric, the same as the *support* operation:

$$l(X \rightarrow Y) = l\,(Y \rightarrow X)$$

Usually, the *lift* value shows that the occurrence of *X* and *Y* together happens more often than anticipated if the lift $l(X \rightarrow Y) \geq 1.0$. Thus there exists a positive *co-occurrence* relationship between *X* and *Y*, which is useful for identifying significant association rules. Algorithm 1 shows how frequent itemsets are generated [40]. The algorithm starts by determining all frequent 1-itemsets. As explained earlier (section III), Table 1 contains a total of five items and therefore we have a total of five *1*-itemsets and ten *2*-itemsets from Table 1. The 2-itemsets in Table 1 are: {i_1, i_2}, {i_1, i_3}, … , {i_4, i_5}. The transactions in Table 1 have only one 5-itemset: {i_1, i_2, i_3, i_4, i_5}. The algorithm and the methodology are illustrated in Figure 1.

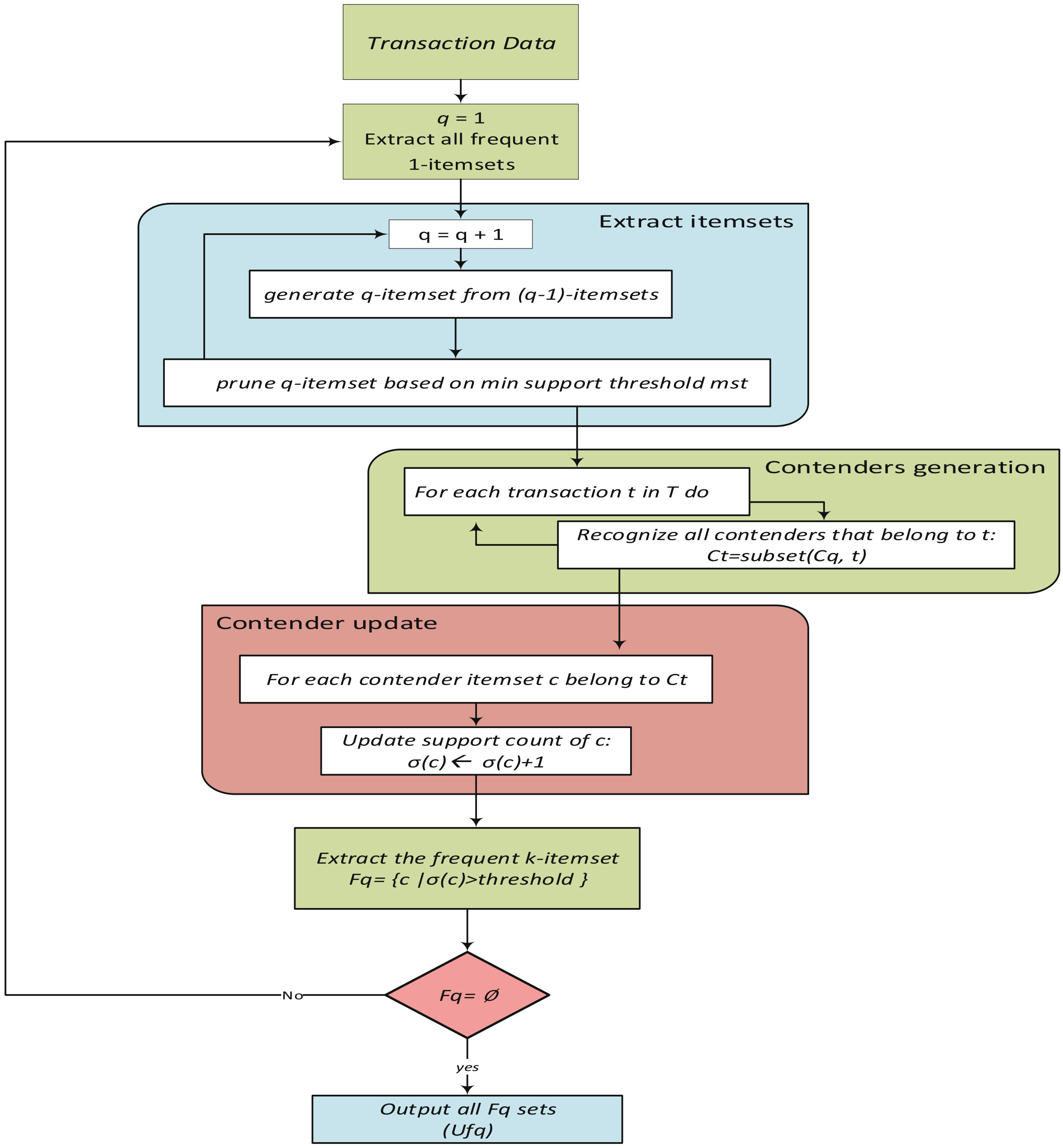

Fig. 1 Illustration of the algorithm and the association rule process.

3.2 Application of Association Rule Mining in Manufacturing

In this study, we want to apply the association rule analysis in the manufacturing and maintenance processes in major manufacturing and industrial organizations. The central aim of association rule mining is to find non-trivial and significant rules and extract them from a given transaction data. The interplay between quality, maintenance, and repair in a manufacturing system is very interesting and it attracts researchers to examine and model these important manufacturing factors. Within a given manufacturing system, there are products, parts, and other items being produced and worked on in daily operations. Such items and products are produced by various machines and have a number of features; these features should be within some (predefined) acceptable range. For example, in the context of cutting machines, e.g. Computerized Numerical Control (CNC) machines, a cutting machine produces and cuts parts that have several dimensions (up to 15 dimensions in some cases). Each one of these dimensions must be within certain acceptable ranges and limits. When any dimension is outside the acceptable range, then a quality issue arises and the product is destroyed and classified as a manufacturing error. Of course, each produced part is subject to certain quality assurance inspection and will be either approved or not approved. This study and the proposed methodology are not limited to CNC and can be applied to any industry where quality control (acceptable or not acceptable) has to be in terms of which are the features causing the quality not approved. For example, in the food industry [78], the (final) product can be acceptable or not acceptable. If it is not acceptable, then there are one or more features that are bad, such as colour, viscosity, taste, texture, and packaging. If the part (the product) passes the quality approval, we denote it by $\boldsymbol{qa}$ (for *quality approved*), otherwise, the part is referred as $\boldsymbol{qna}$ (*quality not approved*). In general, a number of reasons can lead to a *qna* part; one of them can be that one or more of the dimensions are outside the range. For example, one of the dimensions (say dimension 4 or d_4) of some product P_x must be between 0.25 and 0.27 inches (dimension d_4 is represented by feature f_4). If one item of P_x is produced with $d_4 = 0.28$ inch, then the feature f_4 is determined as a bad feature in this transaction. In this context, we can use the dimensions as features of each part; and a *qna* of a part $\boldsymbol{p_j}$ is an indication that one of the dimension $\boldsymbol{d_i}$ of $\boldsymbol{p_j}$ is outside the range. In this case, we assume that the part $\boldsymbol{p_j}$ has one or more bad features.

Assumptions:

1. Each manufactured part is either *qa* or *qna* (that is, each part after being cut will be determined as approved or not approved in its quality).
2. If the part is determined as *qna* (quality not approved), there can be several reasons for that. In this study, we focus on the dimensions only.
3. If the part is *qa* (approved quality) we utilize it in learning good features versus bad ones.

Let the part orders be our transactions, and let the dimensions, or features, of each part be the items. We often use the terms *feature* and *dimension* interchangeably to refer to the same concept. We also use the terms *transactions* and *orders* interchangeably to refer to the same concept. Let us refer to the dimension $\boldsymbol{d_i}$by feature $\boldsymbol{f_i}$. Then, we assign $\boldsymbol{f_i}$=1 if the feature $\boldsymbol{f_i}$ is good; that is, the $\boldsymbol{d_i}$ dimension is within the acceptable range. On the other hand, a feature value of 0 indicates a bad feature, e.g. the dimension is outside the range. A simple example with three transactions and two features is shown in Table 2 (note: the feature in this scenario are items in ARM). In Table 2, the two features shown are *fi* and *fj* and the three transactions are the rows: *tx*, *ty*, *tz*. Each transaction represents one order for cutting/producing a part in accordance with certain dimensions. As we can see in Table 2, the feature *fj* is bad (*fj*=0) in the transaction *tx* and it is good in the other two transactions.

Table 2 Manufacturing data represented in a format appropriate for the association rule process. Each row signifies a transaction while each column signifies a feature.

	....	*fi*	*fj*	....
	. .			.
Tx	…..	1	0	…..
Ty		0	1	
Tz		1	1	
	. .			

We also notice that the feature *fi* in the transaction *ty* is bad, as shown in Table 2. In a data set of bad orders (e.g. representing the data on *qna* parts) we expect each transaction (row) to contain at least one 0 (where 0 represents a bad feature) due to some problem (mechanical problem, electrical problem, software problems, etc.).

From all the transactions of each part in the data, we are most interested in the *qna* ones (*i.e.* transactions involving one or more dimensions outside the range). These orders suggest the need for maintenance, which can be immediate or not immediate. In this study, we utilize ARM with the data of the manufactured parts to extract the association rules and interesting relations and analyse them. We also want to analyse the relationship between bad features and maintenance/repair process features (notice that here a bad feature, like one dimension of the cut part being outside the range, is considered as a manufacturing error or a failure).

Data pre-processing is one of the initial steps in most data mining operations. During the data pre-processing step, all data are cleaned and converted into transaction data. All items and transactions are extracted independently. Then the data are ready for the mining task. Various data mining tasks, such as *classification*, *regression*, *clustering*, and *pattern discovery*, require specific data pre-processing steps. After data pre-processing, rule mining techniques are executed to identify important associations and rules. The main goal is to investigate whether any significant and non-trivial and previously unknown associations exist between features of the cut parts at certain threshold levels. For a manufacturing plant, the resulting relationships and outcomes will have several applications and benefits for the design engineering. For example, if a given part number *i* on the order number *tj* has a bad feature *f5* (*f5=0*) indicating that the dimension d_5 is outside the range, then it would be easier for the design engineer and quality controller to inspect both *d5* and *d9* (i.e. whether f9 is 0 or 1) as these two dimensions (*d5* and *d9*) are outside the range most of the time due to some kind of known mechanical failure at that point in time. In general, we would like to know which features/dimensions are more related and dependent and which features are independent in the existence of failures. The extraction of all possible sets of rules from the data set is a computationally-intensive task. In this study, using association rules mining, we extract interesting, significant and nontrivial rules from the transaction data of part cutting (cutting machines) for a large manufacturing organization.

4. Evaluation and Experiments

4.1 Manufacturing Data

The data set used for the evaluation and discovery of association rules consists of manufacturing data with order transactions from an industrial plant. Each transaction includes a bad part with one or more bad features representing outside-the-range dimensions caused by some electrical, mechanical, or software failure. For the association rule discovery and for extracting the significant association from the data, we pre-processed the data to accommodate the process. The data were collected over a period of more than five years. The collected data include a large number of parts, as shown in Table 3.

Table 3 Details of the manufacturing data.

Total number of orders/transactions	401,000
Period	From June 2004 to October 2009
Duration of data collection	5 years and 4 months
Total number of different parts	1,900
Total number of features/dimensions	15
Number of orders with 10 or more dimensions	2,415
Average number of dimensions	6.3
Average number of orders per part	211

Each transaction represents one order on a part and it is recorded if the part is determined as the *qna* class (quality not approved). The whole data set contains ~401,000 transactions. Every transaction has one or more bad features indicating one or more outside-the-range dimensions. The total number of features (dimensions) in the data set is15; however, some parts have less than that. Of course, these 15 features are all the features needed for this application (note that the manufacturing system collects these 15 features over the period of 5 years). Table 3 contains the details of the first data set. In the data pre-processing phase, we made certain data cleaning and data conversion steps that are needed for the ARM operations. For example, we deleted all the records with many missing values. We also removed every feature that occurs in less than 1000 orders. This step eliminated the last five features $\boldsymbol{f_{11}}$ ... $\boldsymbol{f_{15}}$ leaving us with the remaining 10 features: $\boldsymbol{f_1}$ to $\boldsymbol{f_{10}}$; see Table 5. Initially we conducted experiments with all transactions and the entire data set. We performed the association rule analysis; the outcome included association rules with the three metrics (explained earlier): *support s*, *confidence c*, and *lift l*. We conducted extensive experimentation to derive the best threshold/cutoff for *support* and *confidence*. We used 0.003 as a *cutoff* (experimentally derived) for *support* producing around 1200 rules extracted from the entire data set. The support threshold was set to 0.003, which enabled us to evaluate the data set with the first 9 features (see Table 5). For example, feature *f9* has a support of 0.004, which is higher than the threshold:

$$\boldsymbol{s(f9) = 1744/401000 = 0.004}.$$

This makes feature *f9* included in the association rule mining, and features after *f9* (namely *f10 …f15*) will not be included in the process as their support is below 0.003. Moreover, for the *confidence* threshold, we used 0.510, which means that for the rule $\boldsymbol{fi \rightarrow fj}$, more than half of transactions containing *fi* contain *fj* as well. Table 4 shows the top 10 (most significant) rules which were sorted by *lift* values. These association rules, have high *lift* values (the mean lift value is 20.5), which indicates a significant correlation between the antecedents and consequents of each rule. For example, in the first row in Table 4, the rule $\{\boldsymbol{f_1}, \boldsymbol{f_4}\} \rightarrow \{\boldsymbol{f_7}\}$ has the highest lift value (37.02) indicating that there is an important pattern: when the features *f1* and *f4* are both outside the range, then the feature *f7* will be most likely outside the range as well. Another example from Table 4 shows that if the features *f1* and *f8* are bad, then it is highly likely that *f2* and *f3* are also outside the range. Such information and relations are important for manufacturing and maintenance processes, and may lead to the initiation of certain maintenance or repair tasks that may reduce further failures.

Table 4 The most significant association rules from the entire data set (manufacturing data).

Rule	Antecedent	Consequent	Support	Confidence	Lift
1	*f1, f4*	*f7*	0.0035	0.79	37.02
2	*f1, f8*	*f2, f3*	0.0035	0.53	35.05
3	*f3, f7*	*f8*	0.0046	0.59	26.96
4	*f3, f4*	*f1, f5*	0.0144	0.58	19.88
5	*f4, f5*	*f6*	0.0077	0.66	18.96
6	*f2, f6*	*f5*	0.0083	0.65	17.11
7	*f1, f2, f5*	*f4*	0.0095	0.74	17.11
8	*f5*	*f2, f6*	0.0038	0.7	16.03
9	*f2, f5*	*f3*	0.0495	0.79	12.57
10	*f6*	*f1*	0.0725	0.68	4.17

4.2 Experiments and Feature Analysis

Next, we conducted an extensive analysis of the features in terms of their occurrence with the failure vs without the failure; the results are shown in Table 5. As we can see in Table 5, the feature$\boldsymbol{f_1}$, for example, appeared in transactions more than *320K* times, of which $\boldsymbol{f_1}$was outside the range more than *27K* times (~8.5%). For better visualization, these occurrence and failure details of all the features are shown in Figure_2. Moreover, the percentage of failure of each feature is shown in Table 6. As explained earlier, in subsection 3.2, *qna* means the product (or the part) is not approved (quality not approved: *qna*) because one or more features, which can be f1, f2, and so on, is/are outside the range. Therefore, each final product is determined (and classified) by quality assurance as either *qa* or *qna*. If *qa* is determined, no further action is needed and the final product is within the quality range. On the other hand, if the final product is classified as *qna,* then the bad feature should be determined. We noticed that the feature $\boldsymbol{f_6}$ has the highest percentage of failure, i.e. 49.9%. These feature failure percentages are shown in Figure 3. In that figure, we notice that the feature $\boldsymbol{f_3}$ has the second highest failure percentage of ~27%. So, these two features, $\{\boldsymbol{f_3}$ and $\boldsymbol{f_6}\}$, are closely related to failures and call for further investigation.

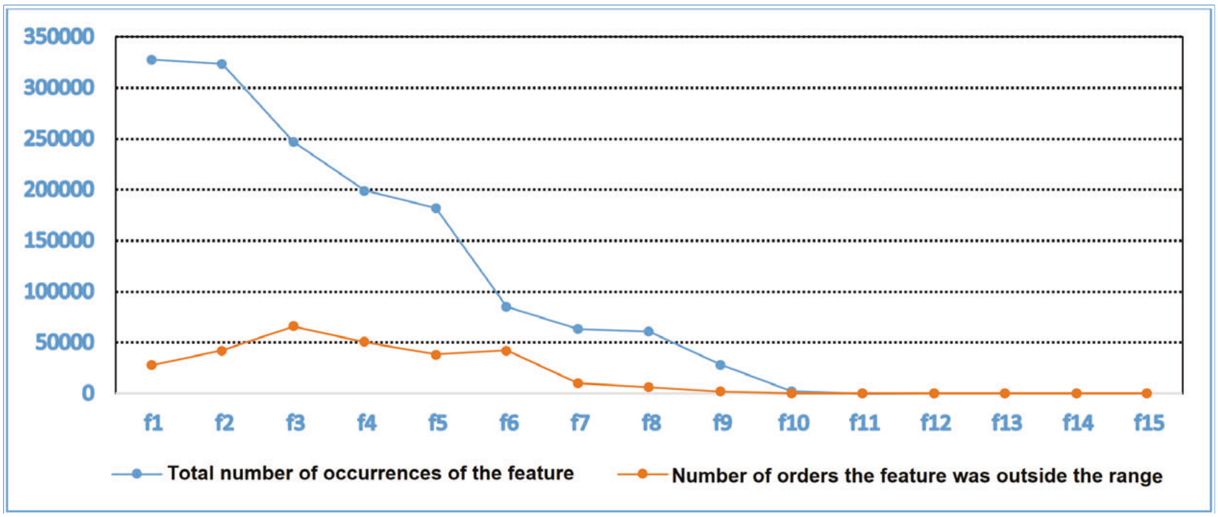

Fig. 2 Distribution of feature occurrences (in how many orders) and the number of occurrences in which the feature is outside the range

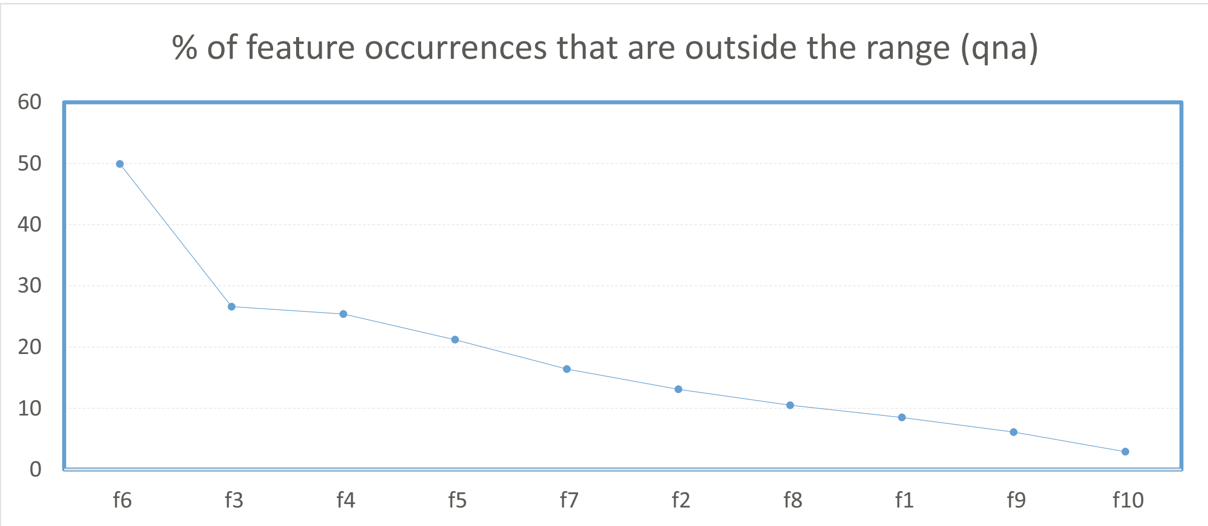

Fig. 3 Illustration of the percentage of *qna* orders for each feature. For example, 50% of the orders containing feature f_6 result in a bad f_6 feature.

We investigated two features, $\{\boldsymbol{f_6}, \boldsymbol{f_3}\}$, and looked further to see how often these two features fail together and we got the following result:

Transactions with failure in:		
$\boldsymbol{f_3}$	$\boldsymbol{f_6}$	$\boldsymbol{f_3}$ and $\boldsymbol{f_6}$
65748	42461	23099

This means that 54% of the failures of $\boldsymbol{f_6}$ are associated with the failures of $\boldsymbol{f_3}$ (in other words, in every two transactions with $\boldsymbol{f_6}$ failures, one of them also includes an $\boldsymbol{f_3}$ failure); in addition, 35% of the time when $\boldsymbol{f_3}$has failure, $\boldsymbol{f_6}$ also has failure, which is statistically significant (*$p < 0.001$*, *hyper geometric test*). Then, we analysed only the top five features with the highest failure percentages in the context of association rule mining to identify the two most co-occurring feature failures. The five features with the highest failure percentages are: $\{\boldsymbol{f_6}, \boldsymbol{f_3}, \boldsymbol{f_4}, \boldsymbol{f_5}, \boldsymbol{f_7}\}$. The results confirm that the pair $\{\boldsymbol{f_3}, \boldsymbol{f_6}\}$ has the highest co-occurring failure percentage, followed by the pair $\{\boldsymbol{f_4}, \boldsymbol{f_6}\}$. We would like to apply these findings in the maintenance and repair processes to improve the overall production and to reduce the cost of the manufacturing system; refer to discussion in Section 5.

Table 5 The top features sorted based on their occurrences in the data set.

Feature	Total number of occurrences of the feature	Number of orders the feature was out of range
f1	328169	27764
f2	323275	42400
f3	246806	65748
f4	199286	50652
f5	181841	38536
f6	84931	42461
f7	63103	10335
f8	60894	6390
f9	28430	1744
f10	2415	72
f11	14	0
f12	0	0
f13	0	0
f14	0	0
f15	0	0

Table 6 The failure percentage of each feature.

Feature	% of feature occurrences that are outside the range (qna)
f1	8.5
f2	13.1
f3	26.6
f4	25.4
f5	21.2
f6	49.9
f7	16.4
f8	10.5
f9	6.1
f10	2.9

Moreover, we then investigated and analysed the two most frequent features, $\{\boldsymbol{f_1}, \boldsymbol{f_2}\}$, in the transaction data in the context of association rules analysis. The results are:

Transactions with failure in:		
f1	***f2***	***f1* and *f2***
27764	42400	10107

This means that in 36% of the transactions when *f1* fails, the feature *f2* fails as well; further, in 24% of the transactions with *f2* failure there are also *f1* failures. Giving this large number of transaction and feature occurrences, these figures are significant with $p < 10^{-5}$ $(hypergeometric\ test)$. In the discussion section, we explain how these results and findings will help the maintenance and repair processes.

5. Discussion

In this paper, we present the results of our methods for extracting the relationships and association rules in the product features and failure data of the manufactured parts. We employed association rule mining as an unsupervised method for extracting interesting and useful patterns in the product quality in relation to the maintenance/repair processes. Association rule mining is one of the established and efficient techniques for data mining and unsupervised learning with wide application in many fields of science [23, 38]. One of the contributions of this paper is the methodology of examining association rule mining in the field of manufacturing for the purpose of identifying the causes and relationships that help the users improve the manufacturing process (e.g. reduce cost).

The data we use include a large number of transactions that include products with quality not approved (*qna*). It is interesting to note that if the features *f1* and *f4* fail, then the feature *f7* also fails with confidence of 0.79 and lift of 37, as shown in Table 4. These are fairly high confidence and lift values (0.79 and 37, respectively) suggesting that this can be a very interesting rule. From industrial engineering and quality control points of view, this rule is meaningful and can be applied to improve the process or reduce maintenance and repair costs. After careful examination of the repair data that include time and date, duration of repair, and type of repair, we noticed that there is some date-time correlation between a particular failure and the repair that was conducted. In this study, we just want to extract significant associations among the features in the presence of failure in more than 400K transactions of failure data. So, this paper serves as a proof of the concept and future research will include extensive experimentation and analysis of the data with the repair and maintenance information. For this application, we should mention that features represent dimensions of a certain part produced by the manufacturing plant. The part is not approved if one of the dimensions, say *dq*, is outside the range, and thus we record that the feature *fq* fails in that particular transaction. Several types of failures (e.g. electrical, mechanical, and software) can occur in machines used in manufacturing. The features and dimensions of the products are standardized according to some norm. For example, two adjacent features, e.g. *f3* and *f4*, fail together most likely because of a mechanical failure. Therefore, association rules will be valuable to design and maintenance engineers for adjusting the maintenance and repair schedules. Sometimes, some of the *qna* parts (parts with bad dimensions) are outside the range due to some external reasons such as the part is not aligned with the cutting/milling machine or the raw material is of low quality. Also, sometimes there are external pollutants/props that cause the part to shift slightly during cutting, causing an outside-the-range dimension and a *qna* part. The association rules can assist manufacturing engineers in determining the cause of the failure. As we mentioned earlier, the maintenance and repair data sets include all the maintenance and repair operations conducted on the machines used in the first data set. This way we can correlate the *qna* with maintenance and repair operations using the data and time fields. In other words, we can use the extracted significant association rules to predict the relationship between bad features and subsequent repairs.

6. Conclusion

In this study we investigate the application of data mining in industrial and manufacturing systems to analyse good features and feature failure patterns so as to improve quality and reduce cost. The main goal is to extract and discover interesting and non-trivial rules and patterns from feature failure data in the presence of products with unapproved quality, in which case repair or maintenance operations might be required. Successful manufacturing and industrial organizations devote special attention to maintenance and the quality of products, which are considered to be of utmost importance to the organization. As reported in the industrial and manufacturing literature, the quality of products and maintenance are highly correlated, and further research is needed in this direction. This paper presents a study in this direction to investigate and discover some interesting association rules and relationship patterns between quality and maintenance using manufacturing (maintenance and product quality) data. Consequently, we use association rule mining, as a methodology, with a large and extensive set of data collected over five years; the set includes the data on the product quality, repair, and maintenance. Our evaluation results and outcomes are very interesting and encouraging. Discovering the features that are linked together and the features that are associated with failures will be very helpful for manufacturers. Repair and maintenance teams can benefit greatly from any information indicating that the feature x is closely related to the failure of type y. With high confidence and lift values, the resulting association rules suggest fairly important associations between product features and the failure, and such findings have not been known and utilized before. These outcomes and findings can assist quality engineers and maintenance teams to improve their overall operations and reduce manufacturing cost.

Acknowledgments

This Project was funded by the National Plan for Science, Technology and Innovation (MAARIFAH), King Abdulaziz City for Science and Technology, Kingdom of Saudi Arabia, Award Number (14-ELE69-02).

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

[1] Baek J-G. An intelligent condition-based maintenance scheduling model. International Journal of Quality & Reliability Management 2007; 24: 312-327. https://doi.org/10.1108/02656710710730898

[2] Sammouri W, Come E, Oukhellou L, et al. Floating train data systems for preventive maintenance: A data mining approach. In: Industrial Engineering and Systems Management (IESM), Proceedings of 2013 International Conference on 2013, pp.1-7. IEEE.

[3] Lewandowski M and Scholz-Reiter B. A Framework for Systematic Design and Operation of Condition-Based Maintenance Systems: Evidence from a Case Study of Fleet Management at Sea Ports. International Journal of Industrial Engineering: Theory, Applications and Practice 2013; 20.

[4] Marseguerra M, Zio E and Podofillini L. Condition-based maintenance optimization by means of genetic algorithms and Monte Carlo simulation. Reliability Engineering & System Safety 2002; 77: 151-165. https://doi.org/10.1016/S0951-8320(02)00043-1

[5] Yang B, Li X, Xie M, et al. A generic data-driven software reliability model with model mining technique. Reliability Engineering & System Safety 2010; 95: 671-678. https://doi.org/10.1016/j.ress.2010.02.006

[6] Rochdi Z, Driss B and Mohamed T. Industrial systems maintenance modelling using Petri nets. Reliability Engineering & System Safety 1999; 65: 119-124. https://doi.org/10.1016/S0951-8320(98)00093-3

[7] Sarker R and Haque A. Optimization of maintenance and spare provisioning policy using simulation. Applied Mathematical Modelling 2000; 24: 751-760. https://doi.org/10.1016/S0307-904X(00)00011-1

[8] Al-Najjar B and Alsyouf I. Enhancing a company's profitability and competitiveness using integrated vibration-based maintenance: A case study. European Journal of Operational Research 2004; 157: 643-657. https://doi.org/10.1016/S0377-2217(03)00258-3

[9] Al-Najjar B and Alsyouf I. Selecting the most efficient maintenance approach using fuzzy multiple criteria decision making. International Journal of Production Economics 2003; 84: 85-100. https://doi.org/10.1016/S0925-5273(02)00380-8

[10] Mirghani MA. A framework for costing planned maintenance. Journal of Quality in Maintenance Engineering 2001; 7: 170-182. https://doi.org/10.1108/13552510110404486

[11] Tan CM and Raghavan N. A framework to practical predictive maintenance modeling for multi-state systems. Reliability Engineering & System Safety 2008; 93: 1138-1150. https://doi.org/10.1016/j.ress.2007.09.003

[12] Ahuja IPS and Khamba JS. Total productive maintenance: literature review and directions. International Journal of Quality & Reliability Management 2008; 25: 709-756. https://doi.org/10.1108/02656710810890890

[13] Alsyouf I. Measuring maintenance performance using a balanced scorecard approach. Journal of Quality in Maintenance Engineering 2006; 12: 133-149. https://doi.org/10.1108/13552510610667165

[14] Chiang J-H and Yuan J. Optimal maintenance policy for a Markovian system under periodic inspection. Reliability Engineering & System Safety 2001; 71: 165-172. https://doi.org/10.1016/S0951-8320(00)00093-4

[15] Savsar M. Effects of maintenance policies on the productivity of flexible manufacturing cells. Omega 2006; 34: 274-282. https://doi.org/10.1016/j.omega.2004.10.010

[16] Vineyard M, Amoako-Gyampah K and Meredith JR. An evaluation of maintenance policies for flexible manufacturing systems: A case study. International Journal of Operations & Production Management 2000; 20: 409-426. https://doi.org/10.1108/01443570010319156

[17] Kans M. An approach for determining the requirements of computerised maintenance management systems. Computers in Industry 2008; 59: 32-40. https://doi.org/10.1016/j.compind.2007.06.003

[18] Komonen K. A cost model of industrial maintenance for profitability analysis and benchmarking. International Journal of Production Economics 2002; 79: 15-31. https://doi.org/10.1016/S0925-5273(00)00187-0

[19] Kutucuoglu K, Hamali J, Irani Z, et al. A framework for managing maintenance using performance measurement systems. International Journal of Operations & Production Management 2001; 21: 173-195. https://doi.org/10.1108/01443570110358521

[20] Parida A. Study and analysis of maintenance performance indicators (MPIs) for LKAB: A case study. Journal of Quality in Maintenance Engineering 2007; 13: 325-337. https://doi.org/10.1108/13552510710829434

[21] Tarakci H, Tang K and Teyarachakul S. Learning effects on maintenance outsourcing. European Journal of Operational Research 2009; 192: 138-150. https://doi.org/10.1016/j.ejor.2007.09.016

[22] Webb GI. Discovering Significant Patterns. Machine Learning 2007; 68: 1-33. https://doi.org/10.1007/s10994-007-5006-x

[23] Wang F. Application of multidimensional association rule techniques in manufacturing resource planning system. In: 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery 10-12 Aug. 2010 2010, pp.1433-1437. https://doi.org/10.1109/FSKD.2010.5569197

[24] Ebeling CE. An introduction to reliability and maintainability engineering. Tata McGraw-Hill Education, 2004.

[25] Golmakani HR. Condition-based inspection scheme for condition-based maintenance. International Journal of Production Research 2012; 50: 3920-3935. https://doi.org/10.1080/00207543.2011.611540

[26] Parida A and Kumar U. Maintenance performance measurement (MPM): issues and challenges. Journal of Quality in Maintenance Engineering 2006; 12: 239-251. https://doi.org/10.1108/13552510610685084

[27] Grall A, Bérenguer C and Dieulle L. A condition-based maintenance policy for stochastically deteriorating systems. Reliability Engineering & System Safety 2002; 76: 167-180. https://doi.org/10.1016/S0951-8320(01)00148-X

[28] Smadi H and Kamrani A. Product Quality-Based Methodology for Machine Failure Analysis and Prediction. International Journal of Industrial Engineerin 2011; 18 2011-12-29.

[29] Nakagawa T and Mizutani S. A summary of maintenance policies for a finite interval. Reliability Engineering & System Safety 2009; 94: 89-96. https://doi.org/10.1016/j.ress.2007.04.004

[30] Wang F-K and Lee W. Learning curve analysis in total productive maintenance. Omega 2001; 29: 491-499. https://doi.org/10.1016/S0305-0483(01)00039-1

[31] Li R, Wu J-f, Wang H-y, et al. Reliability assessment and failure analysis of lithium iron phosphate batteries. Information Sciences 2014; 259: 359-368. https://doi.org/10.1016/j.ins.2013.06.038

[32] Sherwin D. A review of overall models for maintenance management. Journal of Quality in Maintenance Engineering 2000; 6: 138-164. https://doi.org/10.1108/13552510010341171

[33] Yu J and Zhao H. Maintenance plan based on RCM. In: Transmission and Distribution Conference and Exhibition: Asia and Pacific, 2005 IEEE/PES 2005, pp.1-4. IEEE.

[34] Han T and Yang B-S. Development of an e-maintenance system integrating advanced techniques. Computers in Industry 2006; 57: 569-580. https://doi.org/10.1016/j.compind.2006.02.009

[35] Simeu-Abazi Z and Bouredji Z. Monitoring and predictive maintenance: Modeling and analyse of fault latency. Computers in Industry 2006; 57: 504-515. https://doi.org/10.1016/j.compind.2006.02.017

[36] Swanson L. An information-processing model of maintenance management. International Journal of Production Economics 2003; 83: 45-64. https://doi.org/10.1016/S0925-5273(02)00266-9

[37] Purarjomandlangrudi A, Ghapanchi AH and Esmalifalak M. A data mining approach for fault diagnosis: An application of anomaly detection algorithm. Measurement 2014; 55: 343-352. https://doi.org/10.1016/j.measurement.2014.05.029

[38] Nagar A, Hahsler M and Al-Mubaid H. Association rule mining of gene ontology annotation terms for SGD. In: 2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB) 12-15 Aug. 2015 2015, pp.1-7. https://doi.org/10.1109/CIBCB.2015.7300289

[39] Tokola H, Niemi E and Väistö V. Lean manufacturing methods in simulation literature: Review and association analysis. In: 2015 Winter Simulation Conference (WSC) 6-9 Dec. 2015 2015, pp.2239-2248. https://doi.org/10.1109/WSC.2015.7408336

[40] Tan P-N. Introduction to data mining. Pearson Education India, 2006.

[41] Piatetsky-Shapiro G. Discovery, analysis, and presentation of strong rules. Knowledge discovery in databases 1991: 229-238.

[42] Chen W-C, Tseng S-S and Wang C-Y. A novel manufacturing defect detection method using association rule mining techniques. Expert Systems with Applications 2005; 29: 807-815. https://doi.org/10.1016/j.eswa.2005.06.004

[43] Djatna T and Alitu IM. An Application of Association Rule Mining in Total Productive Maintenance Strategy: An Analysis and Modelling in Wooden Door Manufacturing Industry. Procedia Manufacturing 2015; 4: 336-343. https://doi.org/10.1016/j.promfg.2015.11.049

[44] Zhao Y, He Y, Zhou D, et al. Functional risk-oriented integrated preventive maintenance considering product quality loss for multistate manufacturing systems. International Journal of Production Research 2020: 1-18. https://doi.org/10.1080/00207543.2020.1713416

[45] He Y, Gu C, Chen Z, et al. Integrated predictive maintenance strategy for manufacturing systems by combining quality control and mission reliability analysis. International Journal of Production Research 2017; 55: 5841-5862. https://doi.org/10.1080/00207543.2017.1346843

[46] Chen Z, He Y, Zhao Y, et al. Mission Reliability-Oriented Selective Maintenance Optimization for Intelligent Multistate Manufacturing Systems With Uncertain Maintenance Quality. IEEE Access 2019; 7: 109804-109816. https://doi.org/10.1109/ACCESS.2019.2933580

[47] He Z, He Y, Liu F, et al. Big Data-Oriented Product Infant Failure Intelligent Root Cause Identification Using Associated Tree and Fuzzy DEA. IEEE Access 2019; 7: 34687-34698. https://doi.org/10.1109/ACCESS.2019.2904759

[48] Huang J, Chang Q, Zou J, et al. A Real-Time Maintenance Policy for Multi-Stage Manufacturing Systems Considering Imperfect Maintenance Effects. IEEE Access 2018; 6: 62174-62183. https://doi.org/10.1109/ACCESS.2018.2876024

[49] Çiflikli C and Kahya-Özyirmidokuz E. Implementing a data mining solution for enhancing carpet manufacturing productivity. Knowledge-Based Systems 2010; 23: 783-788. https://doi.org/10.1016/j.knosys.2010.05.001

[50] Agrawal R, Imieliński T and Swami A. Mining association rules between sets of items in large databases. In: Acm sigmod record 1993, pp.207-216. ACM. https://doi.org/10.1145/170036.170072

[51] Kamsu-Foguem B, Rigal F and Mauget F. Mining association rules for the quality improvement of the production process. Expert Systems with Applications 2013; 40: 1034-1045. https://doi.org/10.1016/j.eswa.2012.08.039

[52] Alzghoul A, Löfstrand M and Backe B. Data stream forecasting for system fault prediction. Computers & Industrial Engineering 2012; 62: 972-978. https://doi.org/10.1016/j.cie.2011.12.023

[53] Fayyad U, Piatetsky-Shapiro G and Smyth P. From data mining to knowledge discovery in databases. AI magazine 1996; 17: 37.

[54] Adriaans P and Zantinge D. Data Mining Addison Wesley Longman Limited. Edinbourgh Gate, Harlow, CM20 2JE, England 1996.

[55] Fayyad U and Stolorz P. Data mining and KDD: Promise and challenges. Future Generation Computer Systems 1997; 13: 99-115. https://doi.org/10.1016/S0167-739X(97)00015-0

[56] Yevich R. Data mining, Data warehouse: practical advice from the experts. Prentice Hall, 1997.

[57] Vaz CA, de Q.F. Araújo O and de Medeiros JL. Failure Diagnostics Using Data Mining Tools. In: de Brito Alves RM, do Nascimento CAO and Biscaia EC (eds) Computer Aided Chemical Engineering. Elsevier, 2009, pp.1539-1544. https://doi.org/10.1016/S1570-7946(09)70647-9

[58] Ullrich M, ten Hagen K and Lassig J. A data mining approach to reduce the number of maintenance visits in the medical domain. In: Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), 2013 IEEE 7th International Conference on 2013, pp.255-258. IEEE. https://doi.org/10.1109/IDAACS.2013.6662684

[59] Ben-Daya M and Duffuaa S. Maintenance and quality: the missing link. Journal of quality in maintenance engineering 1995; 1: 20-26. https://doi.org/10.1108/13552519510083110

[60] Cakir GS. Development of a condition based maintenance decision model by data mining. School of Industrial Engineering, TUE 2011.

[61] Uhlmann E, Geisert C, Hohwieler E, et al. Data Mining and Visualization of Diagnostic Messages for Condition Monitoring. Procedia CIRP 2013; 11: 225-228. https://doi.org/10.1016/j.procir.2013.07.045

[62] Wang X, Chen B and McGreavy C. Data mining for failure diagnosis of process units by learning probabilistic networks. Process safety and environmental protection 1997; 75: 210-216. https://doi.org/10.1205/095758297529084

[63] Sun J, Wang R, Wang X, et al. Spatial cluster analysis of bursting pipes in water supply networks. Procedia Engineering 2014; 70: 1610-1618. https://doi.org/10.1016/j.proeng.2014.02.178

[64] Iwata H, Tsumoto S, and Hirano S. Maintenance of nursing care plan using similarity-based data mining methods. In: Complex Medical Engineering (CME), 2013 ICME International Conference on 2013, pp.97-102. IEEE. https://doi.org/10.1109/ICCME.2013.6548219

[65] Zhou J, Liu Q, Jin G, et al. Reliability modeling for momentum wheel based on data mining of failure-physics. In: Knowledge Discovery and Data Mining, 2010 WKDD'10 Third International Conference on 2010, pp.115-118. IEEE.

[66] Bastos P, Lopes I and Pires L. A maintenance prediction system using data mining techniques. In: World Congress on Engineering 2012 2012, pp.1448-1453. International Association of Engineers.

[67] Hahsler M, Grün B and Hornik K. A computational environment for mining association rules and frequent item sets. 2005. https://doi.org/10.18637/jss.v014.i15

[68] Liu Y, McMahon C, Ramani K, et al. Editorial for the special issue of information mining and retrieval in design. Elsevier, 2011. https://doi.org/10.1016/j.aei.2011.03.003

[69] Y. He, X. Han, C. Gu, Z. Chen. Cost-oriented predictive maintenance based on mission reliability state for cyber manufacturing systems. Advances in Mechanical Engineering, vol. 10, no. 1, 2018. https://doi.org/10.1177/1687814017751467

[70] E.T. Bekar, P. Nyqvist, A. Skoogh. An intelligent approach for data pre-processing and analysis in predictive maintenance with an industrial case study. Advances in Mechanical Engineering, Vol. 12 (5), pp. 1–14, 2020. https://doi.org/10.1177/1687814020919207

[71] Díaz Vera, Julio César, Negrín Ortiz, Guillermo Manuel, Molina, Carlos, Amparo Vila, Maria. Knowledge Redundancy Approach to Reduce Size in Association Rules. Informatica. Vol. 44, Issue 2, pp. 167–181, 2020. https://doi.org/10.31449/inf.v44i2.2839

[72] Datta, Subrata, Mali, Kalyani. Significant Association Rule Mining with High Associability. 5th International Conference on Intelligent Computing and Control Systems (ICICCS) Intelligent Computing and Control Systems (ICICCS), pp. 1159-1164, May 2021. https://doi.org/10.1109/ICICCS51141.2021.9432237

[73] J. Han, M. Kamber, J. Pei. Data mining: concepts and techniques. 3rd Ed. Morgan Kaufmann, Waltham, Mass. 9, 2011.

[74] R. Glawar, Z. Kemeny, T. Nemeth, K. Matyas, L. Monostori, and W. Sihn. A Holistic Approach for Quality Oriented Maintenance Planning Supported by Data Mining Methods. Procedia CIRP, vol. 57, pp. 259-264, 2016. https://doi.org/10.1016/j.procir.2016.11.045

[75] Lau, H. C. W., Ho, G. T. S., Chu, K. F., Ho, W., & Lee, C. K. M. Development of an intelligent quality management system using fuzzy association rules. Expert Systems with Applications, 36(2), 1801–1815, 2009. https://doi.org/10.1016/j.eswa.2007.12.066

[76] C.K.H. Lee, K.L. Choy, G.T.S. Ho, K.S. Chin, K.M.Y. Law, Y.K. Tse. A hybrid OLAP-association rule mining based quality management system for extracting defect patterns in the garment industry. Expert Systems with Applications, 40 (7), pp. 2435-2446, 2013. https://doi.org/10.1016/j.eswa.2012.10.057

[77] B. Kamsu-Foguem, F. Rigal, and F. Mauget. Mining association rules for the quality improvement of the production process Expert Systems with Applications. Vol. 40, Issue 4, pp.1034-1045, 2013. https://doi.org/10.1016/j.eswa.2012.08.039

[78] D.M. Barrett, J.C. Beaulieu, R. Shewfelt. Color, flavor, texture, and nutritional quality of fresh-cut fruits and vegetables: desirable levels, instrumental and sensory measurement, and the effects of processing. Critical Reviews in Food Science and Nutrition, 50:369–389 (2010). https://doi.org/10.1080/10408391003626322

Submitted: 15.12.2020

Accepted: 14.9.2021

Adel Al-Shayea
Abdulrahman Al-Ahmari
Husam Kaid
Industrial Engineering Department, College of Engineering, King Saud University, P. O. Box 800, Riyadh 11421, Saudi Arabia

Emad Abouel Nasr*
Haitham A. Mahmoud
Industrial Engineering Department, College of Engineering, King Saud University, P. O. Box 800, Riyadh 11421, Saudi Arabia
Faculty of Engineering, Mechanical Engineering Department, Helwan University, Cairo 11732, Egypt

Hisham Al-Mubaid
Computer Science Department, University of Houston – Clear Lake, Houston, TX 77058, USA

Ali K. Kamrani
Industrial Engineering Department, College of Engineering, University of Houston, 213 Engineering Building 2, Houston, TX 77204 4008, USA

*Corresponding author:
eabdelghany@ksu.edu.sa