Article

Using Data Mining Methods for Predicting Sequential Maintenance Activities

Sadok Rezig *, Zied Achour and Nidhal Rezg

Laboratoire de Génie Informatique, de Production et de Maintenance, UFR MIM, Université de Lorraine,
57000 Metz, France; zied.achour@univ-lorraine.fr (Z.A.); nidhal.rezg@univ-lorraine.fr (N.R.)
* Correspondence: sadok.rezig@univ-lorraine.fr; Tel.: +33-372-748-045

check for
updates

Abstract: A data mining approach is integrated in this work for predictive sequential maintenance along with information on spare parts based on the history of the maintenance data. For most practical problems, the simple failure of one part of a given piece of equipment induces the subsequent failure of the other parts of said equipment. For example, it is frequently observed in mining industries that, like many other industries, the maintenance of conventional equipment is carried out in sequence. Besides, depending on the state of parts of the equipment, many parts can be consumed and replaced. Consequently, with a group of spare parts consumed sequentially in various maintenance activities, it is possible to discover sequential maintenance activities. From maintenance data with predefined support or threshold values and spare parts information, this work determines the sequential patterns of maintenance activities. The proposed method predicts the occurrence of the next maintenance activity with information on the consumed spare parts. An industrial real case study is presented in this paper and it is well-noticed that our experimental results shed new light on the maintenance prediction using data mining.

Keywords: data mining; predictive maintenance; sequential pattern; spare parts

1. Introduction

Data mining methods are tools that combine the techniques of artificial intelligence, statistical analysis, and computer science, namely, databases and graphic visualizations, in order to extract and obtain information that is not explicitly represented in the original data that can be more profitable and interesting. Indeed, the purpose of data mining is to extract the relevant information from a large amount of data and, thus, build models of information and knowledge based on fixed criteria. Additionally, to be able to detect standard profiles, recurring behaviors, rules, links, unknown trends (not fixed a priori), and particular structures concisely give the essence of the useful information to assist the decision [1]. It is, therefore, a filtering process that extracts the relevant information from a large amount of information. According to MIT (Massachusetts Institute of Technology), this is one of 10 emerging technologies that will "change the world" in the 21st century.

Data Mining is based on two types of methods; Descriptive methods and Predictive methods. The descriptive approach (Patterns) [2] highlights the information present but hidden by the volume of data. This type of method makes it possible to reduce and synthesize the data and does not require a variable to explain it. However, the predictive method (Modeling) aims to extrapolate new information from the present information (this is, the case of scoring) and, unlike descriptive methods, predictive methods require the explanation of the variables [3].

In the presented industrial case, data mining methods are used as a tool that allows the production of added value by improving productivity, maximizing revenue, minimizing costs, and also ensuring the availability of the equipment based on algorithms and statistical and historical data. This enables

real-time results to facilitate the interventions and control over time. Data Mining can also optimize quality and reduce the scrap by up to 30% [4]. It ensures the sequencing of the maintenance activities with the support of information on the spare parts at the base of the history of the maintenance actions. In addition, the data mining also ensures the determination of the optimal supply of spare parts; in maintenance activities, the availability and good management of spare parts remain an indisputable need that varies according to the severity of the system maintained. In a maintenance intervention, it is often used to act on a set of spare parts, hence, the concept of dependence [5]. Moreover, it is often found that in the case of lacking spare parts, the cost of storage of one or more other spare parts increases.

In the work of Moharana and Sarmah [6], the authors proposed an approach to incorporate the dependence of the elements in the periodic review policy in order to determine the optimal stock of the dependent spare parts by considering the common cycle time and the filling rate of each spare part. At first, a dependency calculation was made for the associated spare parts from the consumption history. Next, the stock management policy was applied in a standard way.

In the last few years, companies have expressed concern on improving the dynamic globalized market where we have seen remarkable improvements in performance in the industrial sector. These changes are transversal to all of the company's processes, which also affect its maintenance functions.

To reduce defects and keep the systems and equipment running, companies have incorporated tools into their Information and Communication Technology (ICT) systems. The benefits are evident in terms of the quality and cost savings, especially those related to the data processing time and the accuracy of the knowledge obtained. In their daily work, companies produce and store huge amounts of data of different natures, increasing the difficulty of the use and processing of the data in real time. In this context, given the relevance of the data collected in industrial facilities, we seek to propose a forecast model of predictive maintenance activities using data mining techniques by means of this topic. This data mining will identify the dependencies between the spare parts and then predict the occurrence of each maintenance sequence.

Manufacturers are still confronted with an immense downtime and have to bear additional costs for transporting spare parts [7]. Consequently, there is a need to learn the maintenance activity sequences performed on the equipment and the way to approximately use the spare parts. This process will reduce the downtime activities by determining the occurrence of the next maintenance activity and the real need for related spare parts. A Sequential pattern mining technique (i.e., one of the data mining techniques) will be a better approach for deducing the sequences and predicting the number of spare parts to be used. The sequential pattern mining approach tracks the relationships or patterns among the data objects over the time periods that are reported in the database. The records of the daily maintenance activities are captured in the maintenance database of the company with timestamp information. The usual maintenance transaction has important attributes such as the maintenance number, the date of maintenance, the equipment number, the equipment sub-section number, the maintenance type, the spare parts used, the quantity of the spare parts, the cost of the spare parts, the labor cost, etc. Using this information, one can find out the relationship between the activities conducted relating to a particular day's maintenance and the actions required to be performed for the next day's maintenance period. Sometimes this process is also useful for unearthing the lowest time gap between the same maintenance activities due to faulty maintenance work done initially.

In this work, the sequential pattern mining algorithm [8] is used on a maintenance dataset which is collected from a large maintenance database of a mining company. Initially, the frequent sequential patterns of the time interval are extracted by using the support or threshold values. Next, the sequential rules are generated from the frequent patterns and a proposed rule based on the classification approach is applied to predict the occurrence of the next maintenance activities. Finally, the spare part codes are mapped on to the discovered sequential maintenance activities.

The rest of the paper is organized as follows. In Section 2, we summarize the research methods presented in the literature and the procedure for mining sequential patterns and discovering sequential rules. The proposed framework for mining sequential maintenance activities and spare parts mapping are described in Section 3. In Section 4, an industrial case study is presented to illustrate the performance of the proposed method. The results and analysis of the case study are also made in the same section. Finally, the conclusions and perspectives are provided in Section 5.

2. Background

The frequent patterns present a considerable advantage for performing other data mining activity processes, e.g., clustering, association-rule mining, classification, etc. In the work by Agrawal et al. [7,8], one can find frequent items from a large database. Discovering itemset/frequent patterns is an important process for performing other data mining activity processes such as association rule mining, classification, prediction, clustering [7,8]. These approaches are called Apriori-like algorithms and pattern growth methods; the apriori-like algorithms follow a basic property like downward closing or anti-monotone, which means that any subset of the frequent itemsets is also frequent. This property is used to eliminate the non-frequent itemset with the given support or threshold support values. The presented approach assumes that all items are binary variables and considers whether they are consumed or not. Not many researchers have proposed an extension of the itemset mining, called itemset mining, with quantities that consider the consumption of the item along with the quantity [9]. Other researchers have investigated the weighted itemset mining considering the occurrence of the item along with its weight or importance [10].

In the work of Huang et al. [11], the authors have introduced different algorithms that describe sequential pattern mining problems for determination of the frequent sequences of the items which should qualify a user-specified minimum support. Then, Bayardo and Agrawal [12] proposed a novel methodology for ranking the association rules and introduced an algorithm for extracting the best rules using rule support and confidence from the large datasets. In Reference [13], the authors developed a new algorithm called PrefixSpan where the global database is projected into a set of smaller (local) databases and sequential patterns that are constructed by exploring frequently occurring datasets of the local databases.

In the recent scientific research, the sequential pattern mining technique is extensively used by researchers in various directions, such as incremental sequence mining, biological sequence mining, multi-dimensional sequence pattern mining, and approximate sequence mining in noisy environments [14]. For example, when a customer buys a computer, he/she is likely to buy a printer, and then a storage disk at a later point in time. For this retail information, managers can do suitable shelf replacements and promotional activities. Several works proposed different algorithms for mining sequential patterns from a large database. The problem of performance is the main issue for sequential pattern mining. Thus, over time, researchers have improved performance by adding some constraints to the mining process. There are many works which contribute to constrained sequential mining. In the work by Masseglia et al. [15], the author used weight constraints to reduce the number of unimportant patterns. In addition, in Reference [16], the authors proposed a mining model by incorporating user-defined constraints for discovering the knowledge that would satisfy the user's needs in a better way. In Reference [17], a Generalized Sequential Models (GSP) algorithm was used for discovering sequential patterns with time constraints. In addition, a new algorithm called the Graph Time Constraints (GTC) algorithm [17] was proposed for mining patterns in large databases. In Reference [18], a new methodology for extracting weighted sequential patterns by considering the time interval weight was developed.

Many research works have applied data mining techniques to achieve various objectives such as detecting faults, predicting failure probabilities, predicting maintenance intervals, prioritizing equipment, determining fault trends, identifying the cause of failures, etc. In Reference [19], the author presented a text mining method by using the abstract and keywords of 150 research papers and revealed

that only 8% of data mining studies cover the area of maintenance. In addition, Betanov et al. [20] proposed a rule-based system for the maintenance management of diagnosing maintenance model selection. The model studied historical failure data and recommended an appropriate policy with optimal preventive maintenance intervals. The aircraft maintenance data are analyzed in Reference [21], which discovered that the parameters link failures, diagnoses, and repair actions in order to enhance the maintenance practices. In Reference [22], the authors suggested a neural-network-based prediction model for assessing the risk priority of medical equipment. Their model was capable of predicting the risk factor assessment for the service departments in large hospitals. From the above literature review, it is remarkable that the sequential pattern mining technique has not been applied by any researchers in the context of maintenance activities for the subsequent management of spare parts up to date. In this work, we have made an attempt to bridge this gap in the literature.

A sequence database consists of sequences of ordered elements or events recorded with or without time-stamped information. Let $I = \{I_1, I_2 \ldots I_p\}$ be the set of all items and s be the sequence which can be written as $\langle e_1, e_2 \ldots e_n \rangle$ and represents the set of all events. For example, in this sequence, e_1 is performed before e_2, and e_3 after e_2. An event can be linked to one or more articles. In Table 1 below, if $x = \langle a, (a, b, c) \rangle$ and $y = \langle (b, c) \rangle$ then x is called the super-sequence of y [23]. Thus, it is noted that in a data sequence s which is a set of then-tuple $\langle SID, s \rangle$, where SID is the sequence id and s is the sequence, if $\langle SID, s \rangle$ contains a sequence α, then α is a subsequence of s. The support value of the α sequence is given as follows: $Supp_s = \frac{|\{ \langle SID, s \rangle | (\langle SID, s \rangle \in S)(\alpha \subseteq s) \}|}{|\{ \langle SID, s \rangle | (\langle SID, s \rangle \in S) \}|}$.

Table 1. An example of a sequence database.

Sequence ID	Sequence
10	<a, (a,b,c), (a,c), d, (c,f)>
20	<(a,d), c, (b,c), (a,e)>
30	<(e,f), (a,b), (d,f), c, b>
40	<e, g, (a,f), c, b, c>

To better understand this, we take the following example.
The supported value of some sequences is given in Table 2 below.

Table 2. The sequential patterns for a 1-sequence.

Patterns	Sequence	Support (%)	Frequent/Infrequent
Pattern 1	<(a)>	100	Frequent
Pattern 2	<(a,e)>	25	Infrequent
Pattern 3	<(a,d)>	25	Infrequent
Pattern 4	<(a,f)>	25	Infrequent
Pattern 5	<(a,c)>	25	Infrequent
Pattern 6	<(a,b)>	50	Frequent
Pattern 7	<(a,b,c)>	25	Infrequent
Pattern 8	<(b)>	100	Frequent
Pattern 9	<(b,c)>	50	Frequent
Pattern 10	<(c)>	100	Frequent
Pattern 11	<(c,f)>	25	Infrequent
Pattern 12	<(d)>	75	Frequent
Pattern 13	<(d,f)>	25	Infrequent
Pattern 14	<(e)>	75	Frequent
Pattern 15	<(e,f)>	25	Infrequent
Pattern 16	<(f)>	75	Frequent
Pattern 17	<(g)>	25	Infrequent

From Table 2, it is clear that the 8 sequential models (<(a)>, <(a,b)>, <(b)>, <(b,c)>, <(c)>, <(d)>, <(e)> and <(f)>) are generated from 17 possible models.

However, the above example is not taken into account because the timestamp information is important in order to identify the intervals between the events in the sequential crawl. For example, suppose a customer buys a computer today and, just after a month, the same customer visits a store to buy a printer. A model was proposed by considering the time constraints from the time-stretched sequences. An example is given here to illustrate the sequential process with a time interval (i.e., see Table 3). The same sequence data are used, including the time-stamped information. The authors defined the following constraints for extended sequences in the mining time [24]: Minsup: The minimum number of sequences that a sequential model must contain; MinInterval: the minimum allowed interval between two successive events; MaxInterval: the maximum allowed interval between two successive events; MinWholeInterval: the minimum interval allowed between the first event and the last event; MaxWholeInterval: intervalle maximal autorisé entre le premier événement et le dernier événement.

Table 3. The extract of the frequent sequential models with temporal information.

Patterns	Sequence	Support (%)
Pattern 1	$<t = 0, (c)>$	75
Pattern 2	$<t = 0, (b,c)>$	75
Pattern 3	$<t = 0, (b)>$	100
Pattern 4	$<t = 0, (a,b,c)>$	75
Pattern 5	$<t = 0, (a,b)>$	100
Pattern 6	$<t = 0, (a,c)>$	75
Pattern 7	$<t = 0, (a)>$	100
Pattern 8	$<\{t = 0, (b)\}, \{t = 1, (c)\}>$	75
Pattern 9	$<\{t = 0, (b)\}, \{t = 1, (a,b)\}>$	75
Pattern 10	$<\{t = 0, (b)\}, \{t = 1, (a,c)\}>$	75
Pattern 11	$<\{t = 0, (b)\}, \{t = 1, (a)\}>$	100
Pattern 12	$<\{t = 0, (b)\}, \{t = 1, (b)\}>$	75
Pattern 13	$<\{t = 0, (a)\}, \{t = 1, (a,b)\}>$	75
Pattern 14	$<\{t = 0, (a)\}, \{t = 1, (a)\}>$	75
Pattern 15	$<\{t = 0, (a)\}, \{t = 1, (b)\}>$	75
Pattern 16	$<\{t = 0, (a,b)\}, \{t = 1, (a)\}>$	75

This work sheds a new light on in the literature with regards to the application of the sequential data mining technique for the prediction of the maintenance sequence and the management of spare parts and, thereby, this research contributes to the body of knowledge.

3. The Proposed Model

Our proposed model is intended to integrate the sequential maintenance models with their frequent group of spare parts. We describe the framework of the proposed model here, such as the maintenance data collection, the sequential model generation, the frequent generation of spares groups, and the integration of sequential maintenance activities with the associated spares. The model frame describes the sequence of time intervals for maintenance activities and the corresponding spare parts (i.e., Figure 1).
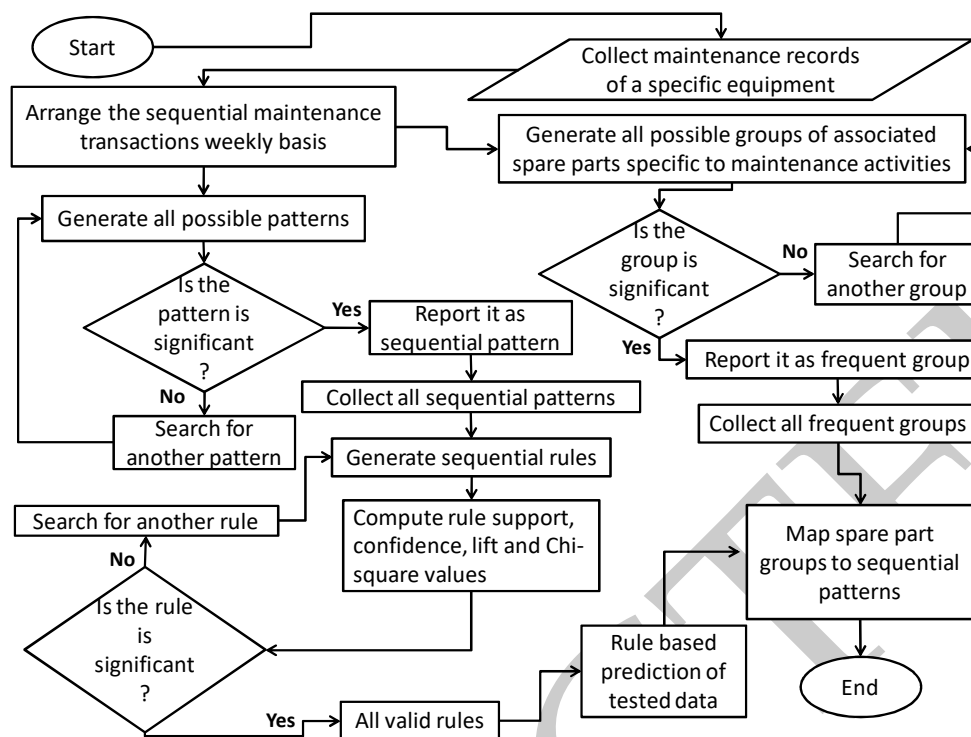
Figure 1. The organization chart of the proposed model.

3.1. Collection of Maintenance Data

Generally, maintenance data includes different attributes such as the location, equipment number, equipment sub-section number, maintenance type, maintenance start date, date of maintenance, end of maintenance, the spare parts used, quantities of the spare parts, description of the repair or replacement, etc. This information is stored in a maintenance database that is also linked to an integrated hardware management or enterprise resource planning (ERP) database. The data cleansing process removes the unwanted records or orphan records that have been entered by users incorrectly or are unnecessary because of a system error. After cleanup, a separate table or database view is created to store these clean data for classification, forecasting, or trend analysis of other mining activities.

Only the maintenance date information for a single device is used to retrieve the sequential patterns. The complexity of generating the sets of items or the pattern combination increases as the number of items increases. In the case of sequential pattern mining, the possible number of models using Generalized Sequential Models (GSP) extraction [17] is $n^2 + n(n-1)/2$, where n is the number of elements present in the database. For example, 51 patterns can be generated using 6 items with a length of 1. In this case, an a priori algorithm was used in a GSP to cut all the candidates. In this study, sequential maintenance activities with time stamps are used because the time information can be used more practically in maintenance compared to the sequences without time information. A pattern is called sequential when the sequence support is greater than or equal to the threshold support values. For the time slot sequences, four other threshold values such as the MinInterval, MaxInterval, MinWholeInterval, and MaxWholeInterval are used with the Minsupp values. In this study, these thresholds were identified, the sequential models were generated and the results were compared with different threshold values. However, in reality, these values must be suggested by decision-makers or experts.

3.2. Generation of Sequential Rules

Once the sequential model is generated, the next step is to generate sequential association rules among the element sets. A similar a priori algorithm is used to generate the association rules.

Usually, a rule consists of two components called the antecedent (or condition) and the consequence (or part of the conclusion). It is written in the form of an IF-THEN expression as IF x THEN y or x. The rule is evaluated using various statistical measures such as a rule support, confidence, and Chi-square values [25–28].

3.3. Classification by Rules of Maintenance Activities

After validating the selection of the sequential rules using a set of training data, these rules can be used to classify the test data. The consequent part of a new record must be predicted using this method. Referring to Reference [23], the accuracy of the prediction can be measured using two important ratios:

- The coverage measure, which shows the proportion of *n*-tuplets that respect the rule R

$$Coverage = \frac{\eta_{covers}}{|D|} \tag{1}$$

- The rule Accuracy, which checksthe accuracy of the rule

$$accuracy = \frac{\eta_{correct}}{\eta_{covers}} \tag{2}$$

3.4. Generation of Frequent Spare Parts Groups

In this step, we need to determine the frequently used spare parts for individual maintenance activities. Few spare parts are used in some maintenance operations and, in some cases, many spare parts are used for the same maintenance activity. Thus, by analyzing the historical consumption of spare parts demand, we can find the best group that is frequently used in the same type of maintenance activity.

Example: We let Minsup be 40%. In Table 4 below, we observe that {S1, S2, S4} and {S1, S3, S4} have the highest spare parts coverage with support values of 42.13% and 41.71%, respectively. Thus, the best group of spare parts is the highest from among the both of them, i.e., {S1, S2, S4}.

Table 4. The spare part groups with support.

Spare Part Groups	Support (%)
S1	87.24
S2	56.28
S3	55.42
S1, S2	50.42
S1, S3	49.28
S4	45.23
S1, S2, S4	42.13
S1, S3, S4	41.71
S2, S4	40.57

4. Concrete Industrial Example

In this example, the data were collected from a mining company in southern India. The company has implemented integrated materials management since 2004 by integrating indentation, supplier sourcing, purchasing, and warehousing (inventory). Thus, all the spare parts consumption information is stored in a central database. The company has three large mining units that use a variety of special mining equipment, such as 28-wheel bucket wheel excavators, 12-wheel belt conveyors, 16-wheel spreaders, and 9-car trippers. The company manages nearly 150,000 items, of which 105,217 are spare parts that are mainly consumed by the mines and power generation equipment.

In this study, the equipment belt conveyor and the proposed model are considered to be applied to seven different belt conveyor related maintenance activities which consume approximately 2643 spare parts at various mining sites. We collected the spare parts from the consumption history. Regarding the maintenance activities of the 1800 conveyors in Mine-I (Table 5 below), the data show the seven maintenance activities and the few related spare parts used during the maintenance activities. We collected the maintenance data and spare part consumption data from April 2006 to February 2012. We used binary transaction data for the spare parts consumption. The number '1' indicated that the spare part is consumed and '0' indicated that the spare part was not consumed during the maintenance activity.

Table 5. The spare parts information of the industrial example.

Maintenance	Spare Part Code	Part Description
POWER SECTION MAINTENANCE (M01)	490040037 (S11)	FANOX ELECTRONIC MOTOR PROTECTION RELAY
	490080064 (S12)	ELECTRICAL SPARES FOR AUTOMATIC BUFFING
	499120322 (S13)	4 POLE, 440 VOLTS, 50 CYCLES, 40 AMPS
	490040016 (S14)	ELECTRONIC LIMIT SWITCH GW 2-000G-ITEM
	490000879 (S15)	GASKET 167.5X178X0.7 PTNO 400V
	490030120 (S16)	DOUBLE SOLENOID VALUE FOR 220V AC MOTOR
LOOSE FIXING AND NOISE (M02)	497030005 (S21)	HEATER PLATE BOLTS AND NUTS, SIZE: 6MM D
	497030054 (S22)	SPECIAL BOLT AND NUT FOR BTR CLAMPS AS P
	499120217 (S23)	BTR CLAMP 2 $\frac{1}{2}$ INCH WITH NECESSARY BOLT
	499140003 (S24)	LINK BOLT WITH SELF LOCKING NUT FOR TI/RI FOR 1800 MM
	499140107 (S25)	LINK BOLT WITH SELF LOCKING NUT AND WASH
	499140110 (S26)	LINK BOLTS OF SIZE 15.9MM DIA X 66.5MM
	499150142 (S27)	8.8 GRADE BOLT AND NUT OF SIZE 24X60MM
	499180010 (S28)	FISH PLATE BOLTS TYPE-II WITH NUTS SIZE
	499180011 (S29)	NOSE PIECE BOLT and NUT AS PER
IDLER MAINTENANCE (M03)	499130207 (S31)	SIDE TROUGHING IDLER SIZE 159MM DIA X 1100MM LONG
	499130409 (S32)	CENTRE TROUGHING IDLER SIZE 159MM DIA X 1100MM LONG
	499130211 (S33)	RETURN TROUGHING IDLER OF 159MM DIA X 1100MM LONG
	499140003 (S34)	LINK BOLT WITH SELF LOCKING NUT FOR ALL IDLERS
	499140108 (S35)	LINK PLATE FOR TROUGHING AND RETURN IDLER
	499140121 (S36)	SHAFT FOR THE CENTRE ROLLER OF TROUGHING
	499140167 (S37)	SHAFT DIA 60/62 X 450 MM LONG FOR IMPACT IDLER
	499140241 (S38)	BARRELS FOR 2000MM CONVEYOR RETURN IDLER
FRAME MAINTENANCE (M04)	492010408 (S41)	L ROD WITH PIN FOR 2400MM CONVEYOR STANDARD FRAME
	499140103 (S42)	STANDARD SHIFTABLE FRAMES IN COMPLETELY
	499140118 (S43)	STRESS FRAMES FOR 1800 MM CONVEYOR SYSTEM
	499170002 (S44)	SHIFTABLE A FRAME PLATE SET
	499170022 (S45)	CABLE CLAMP FOR 1800MM CONVEYOR FRAMES
	499170036 (S46)	MAIN FRAME FOR FOUR ROLLER TRACK SHIFTER
BELT MAINTENANCE (M05)	499120215 (S51)	METAL COVER SHEET FOR 1800 SC BELT
	499120001 (S52)	STEEL CORD CONVEYOR BELT 1800MM ST 2250
	499120221 (S53)	E.R.W. PIPE SHED FOR 1800 MM BELTS AS PE
	499120106 (S54)	BELT CRAMP FOR 1800MM ST 4000 S.C BELT
	499120102 (S55)	BBELT FASTNERS PLATE TYPE (3 INCH)
	497020001 (S56)	BELT PULLING CLAMPS 50MM
	490070006 (S57)	BELT DRESSING M/C CUTTER BLADES OF STEEL
	497010020 (S58)	COTTON SHEET FOR BELT JOINTS WIDTH 1000
	490030132 (S59)	BELT PRESSING AND LIFTING ARM 5/16 INCH ID, DWB HOSE

Table 5. *Cont.*

Maintenance	Spare Part Code	Part Description
HYDROULIC SYSTEM MAINTENANCE (M06)	490000092 (S61)	HYDRAULIC CYLINDERS—ALUMINIUM ALLOY FO
	490001251 (S62)	HYDRAULIC CLAMP TOWER JACKS CYLINDER (IT)
	490010217 (S63)	HYDRAULIC JACK OIL SEAL: 315 X 355 X 33.3 FOR WAGNER PRESS
	490030045 (S64)	HYDRAULIC OIL PRESSURE GAUGE (1) RANGE
	490030045 (S65)	PRESSURE FOOT FOR NILOS HYDRAULIC JACK O
	499050201 (S66)	HYDRAULIC HOSE CONFIRMS TO SAE 100 R2 ID
	499050297 (S67)	HYDRAULIC HOSE OF ID5/16 INCHES AND LENGTH 3000MM
	499120101 (S68)	END BOLT FOR HYDRAULIC TRAVERSEFOR 1800
	499120304 (S69)	NILOS HYDRAULIC CYLINDERS WITH "T" JOINT
BEARING AND LUBRIFICATION (M07)	4990000122 (S71)	BEARING HOUSINGS FOR TROUGHING AND RETURN
	491020009 (S72)	BEARING COVER INNERANDOUTER AND LABYRINTH
	499130007 (S73)	GREASE SEAL RING FOR IMPACT IDLERS OF 1800MM CONVEYOR
	499130215 (S74)	BARREL WITH BEARING HOUSINGS WELDED ON B
	499130401 (S75)	BEARING HOUSING FOR IMPACT IDLER OF 1800
	499130416 (S76)	GREASE SEAL RING FOR TROUGHING IDLER ANDR

Table 6 shows the steps of our solving method.

Table 6. The progress of our solving method.

Step	Description
Collection of maintenance data	GSP (Generalized Sequential Models) Support Values Thresholds values Criteria (MinInterval, coverage, etc.)
Generation of sequential rules	Rule support Rule Confidence Lift Chi-square measure
Classification by rules of maintenance activities	Accuracy and Coverage
Generation of frequent spare parts groups	Historic and Support values

Discussion of Results

We considered the seven maintenance activities of the mobile transport conveyor such as power section maintenance, loose fixings and noise reduction, idler maintenance, conveyor frame maintenance, belt maintenance, hydraulic system maintenance and bearing, and lubrication for our analysis. In Table 6, the maintenance activities related to spare parts codes and descriptions are given.

For computational simplicity, we used 40% support values for demonstrating the rule-based classification approach. Sequences having a single maintenance activity are ignored from our model as they do not indicate the next maintenance activity information. This will only be used while computing the confidence value of the sequential association rules. In this analysis, we assumed that there are no occurrences of parallel maintenance activities performed for the conveyor belt. In our analysis, the symbol '—>' indicates the sequence direction and the symbol '==>' indicates the rule direction.

- Collection of maintenance data:

Using the GSP algorithm, 58 sequential models were generated with a 30% support threshold value and, likewise, 27 models were generated with 40% support threshold. The results are given in Table 7.

Table 7. The determined sequential patterns.

S1. No.	Sequences	Intervals (weeks)	Support (%)
1	M01	-	99.22
2	M03	-	96.39
3	M05	-	83.61
4	M02	-	80.33
5	M04	-	74.76
6	M06	-	74.49
7	M04→M03	0–1	70.20
8	M03→M02	0–1	67.69
9	M01→M03	0–2	59.15
10	M01→M03	0–1	57.87
11	M05→M03	0–3	57.51
12	M06→M03	0–1	57.33
13	M06→M03	0–3	51.67
14	M05→M06	0–2	49.66
15	M05→M06→M03	0–2–3	49.29
16	M05→M02	0–4	48.61
17	M05→M03→M02	0–3–4	48.52
18	M06→M02	0–2	48.47
19	M06→M03→M02	0–1–2	48.38
20	M05→M06→M02	0–2–4	48.29
21	M05→M06→M03→M02	0–2–3–4	48.20
22	M01→M02	0–2	44.04
23	M01→M03→M02	0–1–2	43.86
24	M04→M02	0–2	42.54
25	M04→M03→M02	0–1–2	42.36
26	M01→M04	0–1	40.07
. . .	. . .	. . .	. . . .
55	M06→M01→M04→M03	0–1–2–3	31.99
56	M05, M06→M01→M03	0–1–3	31.99
57	M01→M04, M06→M03	0–1–3	31.95
58	M05, M06→M01→M04→M03	0–1–2–3	31.81

One can remark that the sequences have the maximum support values as follows:

- For a combination of 2: M04→M03→
- For a combination of 3:

	Support	Average interval
M06→M03→M02	48.29	1(0–1–2)
M05→M06→M03	49.29	1.5(0–2–3)

One can note that the second sequence has the maximum support value, however, the chosen sequence is M06→M03→M02 because of its lower average interval.

- Generation of sequential rules

Since the sequences are identified, one can calculate the following 4 measures: rule support, rule confidence, lift, and Chi-square measure. From these 4 measurements, one can deduce if the sequences are significant.

- Classification by rules of maintenance activities

After determining the rules, it was necessary to test them and calculate the coverage and accuracy rates. The results are as follows (i.e., see Table 8).

Table 8. The result of the classification.

S1. No.	Rules	η_{covers}	$\eta_{correct}$	Coverage (%)	Accuracy (%)
1	M05→M06→M03⇒M02	92	88	40.10	95.65
2	M05→M06⇒M03→M02	92	82	40.10	89.13

- Generation of frequent spare parts groups

The last step is to link the set of spare parts groups for each type of maintenance based on the history. Then, for each group, we calculated its support value in order to choose the maximum value (see Figure 1).

After the generation of the sequential rules for classification, we tested these rules with the test data of 225 maintenance transactions for validation. The results of the rule coverage and accuracy are given in Table 8. It is observed that the rules showed accurate predictions with the test data. The next step of the model was to find out the frequent spare parts group for each maintenance activity separately. In Table 9, we have shown the 350 records of spare parts consumption for Belt maintenance (*M03*). The spare parts group is determined by using the procedure mentioned in Section 3 and the result sare given in Table 10 by considering a support value of 50%.

Table 9. The spare parts information for M03.

Maintenance ID	Type of Maintenance	Spare Parts Information								
		S31	S32	S33	S34	S35	S36	S37	S38	S39
1	*M03*	1	1	1	1	0	0	1	0	1
2	*M03*	1	0	1	1	1	1	1	1	1
3	*M03*	1	1	0	0	1	1	1	1	0
4	*M03*	0	1	1	1	1	1	1	1	1
5	*M03*	1	1	1	1	1	1	1	0	0
6	*M03*	1	1	1	1	1	1	1	1	1
. . .	. . .					. . .				
350	*M03*	1	0	1	1	1	0	0	1	1

Table 10. Selecting the spare parts group with a 50% support.

Maintenance Activity	Spare Parts Group	Support (%)
	S33, S37, S39	57.14
	S31, S33, S34	56.28
	S33, S34, S38, S39	55.42 *
	S33, S34, S35, S39	53.42
IDLER MAINTENANCE M03	S31, S33, S34, S39	52.28
	S35, S36, S37, S38	52.00
	S31, S33, S34, S37	49.71
	S31, S33, S34, S36, . . . , S39	46.57

* One can remark that spare parts {S33, S34, S38, S39} are selected as best group for M03. Similarly, {S52, S54, S55, S57, S59}, {S61, S63, S64, S68}, and {S23, S25, S26, S27, S29} are selected for M05, M06, and M02 respectively. Now these groups are mapped to individual maintenance activities and intervals.

5. Conclusions and Perspectives

In this work, an attempt was made to suggest a list of sequential maintenance activities based on the historical records of maintenance data. Our sequential mining approach gave a better way of analyzing the sequences compared to the traditional statistical analyses used, particularly when the volume of the maintenance data is large. Generally, maintenance managers suggest maintenance activities on a rule of thumb basis or as derived from their experience and they physically control the equipment at regular intervals. The suggested method tries to avoid these manual efforts. Thus, one can generate a longer sequence which covers a higher number of maintenance activities

based on the threshold value. Next, the frequent spare parts group is mapped to the maintenance activity of the generated sequence based on the threshold support values set by the decision-makers. If the managers want to cover a higher number of spares, they can reduce the threshold values. This helps the maintenance managers carry these spare parts directly to the maintenance location and the number of trips from the stores to the maintenance location will be reduced, thereby decreasing the transport costs. Finally, the proposed approach for suggesting maintenance activities is dynamic in nature as it is directly connected to the maintenance database. The generated sequential patterns may change periodically depending on the actual changes or personalizations to the maintenance activities.

The developed model analyses the past preventive maintenance records of a given piece of equipment and tries to determine the sequential activities of different maintenance activities, including the time stamp information. In addition, the timestamp information can be used to prioritize a maintenance activity that has been ignored in a particular piece of equipment. In future studies, one can try to weigh the time intervals to determine the sequences. Our model can be extended for analyzing the failure maintenance activities and to perform root cause analyses which can give suggestions that are more valuable to maintenance managers in order for them to take corrective actions prior to the next occurrence of a failure.

References

1. Hirate, Y.; Kato, S.; Yamana, H. Web structure in 2005. In *Algorithms and Models for the Web-Graph, Proceedings of the International Workshop on Algorithms and Models for the Web-Graph, Banff, AB, Canada, 30 November–1 December 2006*; Aiello, W., Broder, A., Janssen, J., Milios, E., Eds.; Springer: Berlin/Heidelberg, Germany, 1842; pp. 36–46.

2. Hand, D.J. Principles of data mining. *Drug Saf.* **2007**, *30*, 621–622. [CrossRef] [PubMed]

3. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA data mining software: An update. *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 10–18. [CrossRef]

4. Pujari, A.K. *Data Mining Techniques, 2001*; Sangam Books Ltd.: Telangana, India, 1947.

5. Joachims, T. Optimizing search engines using clickthrough data. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, AL, Canada, 23–26 July 2002; pp. 133–142.

6. Moharana, P.C.; Sharma, B.M.; Biswas, D.R.; Dwivedi, B.S.; Singh, R.V. Long-term effect of nutrient management on soil fertility and soil organic carbon pools under a 6-year-old pearl millet–wheat cropping system in an Inceptisol of subtropical India. *Field Crop. Res.* **2012**, *136*, 32–41. [CrossRef]

7. Agrawal, R.; Imieliński, T.; Swami, A. Mining association rules between sets of items in large databases. *ACM SIGMOD Rec.* **1993**, *22*, 207–216. [CrossRef]

8. Agarwal, B. Participatory exclusions, community forestry, and gender: An analysis for South Asia and a conceptual framework. *World Dev.* **2001**, *29*, 1623–1648. [CrossRef]

9. Srikant, R.; Agrawal, R. Mining sequential patterns: Generalizations and performance improvements. In *Advances in Database Technology—EDBT '96, Proceedings of the International Conference on Extending Database Technology, Avignon, France, 25–29 March 1996*; Apers, P., Bouzeghoub, M., Gardarin, G., Eds.; Springer: Berlin/Heidelberg, Germany, 1842; pp. 1–17.

10. Yun, C.H.; Boggon, T.J.; Li, Y.; Woo, M.S.; Greulich, H.; Meyerson, M.; Eck, M.J. Structures of lung cancer-derived EGFR mutants and inhibitor complexes: Mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cells* **2007**, *11*, 217–227. [CrossRef] [PubMed]

11. Huang, S.; Li, R.; Zhang, Z.; Li, L.; Gu, X.; Fan, W.; Lucas, W.J.; Wang, X.; Xie, B.; Ni, P.; et al. The genome of the cucumber, *Cucumis sativus* L. *Nat. Genet.* **2009**, *41*, 1275–1281. [CrossRef] [PubMed]

12. Bayardo, R.J., Jr.; Agrawal, R. Mining the most interesting rules. In Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 15–18 August 1999; pp. 145–154.

13. Pei, J.; Han, J.; Mortazavi-Asl, B.; Wang, J.; Pinto, H.; Chen, Q.; Hsu, M.C. Mining sequential patterns by pattern-growth: The prefixspan approach. *IEEE Trans. Knowl. Data Eng.* **2004**, *11*, 1424–1440.

14. Chi, Y.; Wang, H.; Philip, S.Y.; Muntz, R.R. Catch the moment: maintaining closed frequent itemsets over a data stream sliding window. *Knowl. Inf. Syst.* **2006**, *10*, 265–294. [CrossRef]

15. Masseglia, F.; Poncelet, P.; Teisseire, M.; Marascu, A. Web usage mining: Extracting unexpected periods from web logs. *Data Min. Knowl. Discov.* **2008**, *16*, 39–65. [CrossRef]

16. Chang, J.H. Mining weighted sequential patterns in a sequence database with a time-interval weight. *Knowl.-Based. Syst.* **2011**, *24*, 1–9. [CrossRef]

17. Washio, T.; Motoda, H. State of the art of graph-based data mining. *ACM SIGKDD Explor. Newsl.* **2003**, *5*, 59–68. [CrossRef]

18. Yun, U. A new framework for detecting weighted sequential patterns in large sequence databases. *Knowl.-Based. Syst.* **2008**, *21*, 110–122. [CrossRef]

19. Choudhary, A.K.; Harding, J.A.; Tiwari, M.K. Data mining in manufacturing: A review based on the kind of knowledge. *J. Intell. Manuf.* **2009**, *20*, 501. [CrossRef]

20. BETANOV. Cemil. In *Introduction to X. 400, 1993*; Artech House, Inc.: Norwood, MA, USA, 1974.

21. Young, T.; Fehskens, M.; Pujara, P.; Burger, M.; Edwards, G. Utilizing data mining to influence maintenance actions. In Proceedings of the 2010 IEEE AUTOTESTCON, Orlando, FL, USA, 13–16 September 2010; pp. 1–5.

22. Al-Naima, F.; Al-Timemy, A.H.A. A neural network based algorithm for assessing risk priority of medical equipments. In Proceedings of the 2010 7th International Multi-Conference on Systems, Signals and Devices, Amman, Jordan, 27–30 June 2010; pp. 1–6.

23. Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*, 3rd ed.; Morgan Kaufmann Publishers: Burlington, MA, USA, 1984; pp. 230–240.

24. Hirate, Y.; Yamana, H. Generalized Sequential Pattern Mining with Item Intervals. *J. Comput.* **2006**, *1*, 51–60. [CrossRef]

25. Azevedo, P.J.; Jorge, A.M. Comparing rule measures for predictive association rules. In *Machine Learning: ECML 2007, Proceedings of the 18th European Conference on Machine Learning, Warsaw, Poland, 17–21 September 2007*; Kok, J.N., Koronacki, J., Mantaras, R.L.D., Matwin, S., Mladenič, D., Skowron, A., Eds.; Springer: Berlin/Heidelberg, Germany, 1842; pp. 510–517.

26. Sá, C.R.D.; Azevedo, P.; Soares, C.; Jorge, A.M.; Knobbe, A. Preference rules for label ranking. *Inf. Fusion* **2018**, *40*, 112–125.

27. Thaseen, I.S.; Cherukuri, A.K. Intrusion detection model using fusion of chi-square feature selection and multi class SVM. *Comput. Inform. Sci.* **2017**, *29*, 462–472.

28. Fedrizzi, M.; Ferrari, F. A chi-square-based inconsistency index for pairwise comparison matrices. *J. Oper. Res. Soc.* **2018**, *69*, 1125–1134. [CrossRef]