International Journal of Control and Automation
Vol.8, No.12 (2015), pp.47-64
http://dx.doi.org/10.14257/ijca.2015.8.12.05

Mining Interesting Least Association Rules in Manufacturing Industry: A Case Study in MODENAS

Yunus Indra Purnama[1], Zailani Abdullah[2], Rokhmat Rokhmat[1] and Tutut Herawan[3]

[1]Faculty of Information Technology and Business
Universitas Teknologi Yogyakarta
Kampus UTY, Jalan Lingkar Utara, Yogyakarta, Indonesia
[2]School of Informatics and Applied Mathematics
Universiti Malaysia Terengganu
Gong Badak, Terangganu, Malaysia
[3]AMCS Research Center, Yogyakarta, Indonesia
indra_uty@yahoo.com, zailania@umt.edu.my, rokhmat_uty@yahoo.com, tutut@amcs.co

Abstract

Least association rules are related to the rarity or uncommonness relationship among itemset in database repository. However, mining these rules are quite tricky and seldom discussed since it usually includes with infrequent items or exceptional cases. In manufacturing industry, detecting these rules is very useful and exciting for further analysis such as for market segmentation, prediction, and product arrangements. In this paper, we introduce an enhanced association rules mining method, called Significant Least Pattern Growth (SLP-Growth) and a new measurement named Critical Relative Support (CRS). The novelty of the proposed method is that unlike existing methods, it is used for capturing interesting least items from line-off database. The data which comprised of production of motorcycle/scooter is taken from Malaysia national motorcycle manufacturer called Motorsikal dan Enjin Nasional Sdn Bhd (MODENAS). The results from this research provide useful information for the management to understand the customer demand trends comprehensibly, and enable them to design marketing strategy accordingly.

Keywords: *Data Mining, Least association rules, Critical relative support, Manufacturing industry*

1. Introduction

Data mining is defined as "the nontrivial extraction of implicit, previously unknown, and potentially useful information from data [1]. It is an integral part of knowledge discovery in databases (KDD), which is the overall process of converting raw data into useful information [2]. Until this recent, data mining is still one of the important and popular research areas in knowledge discovery community [3-7]. As an indicator, data mining has been successfully applied in many domain applications including in manufacturing industries [8-12]. One of the well-known data mining techniques that have been widely employed in multidiscipline applications [13] is Association Rules Mining (ARM). It aims at discovering the interesting correlations, frequent patterns, associations or casual structures among sets of items in the data repositories. The problem of association rules (ARs) mining was first introduced by Agrawal for market-basket analysis [14-16]. Typically, two main stages are involved before generating the ARs. First,

ISSN: 2005-4297 IJCA

find all frequent items from transactional database. Second, generate the common ARs from the frequent items. In ARs, an item is said to be frequent if it appears more than a predefined minimum support threshold. Besides that, confidence is another measure that always used in pair with the minimum support threshold. Typically, a set of items is also known as itemset in the context of ARs. By definition, least items are a set of items that is rarely occurred in the database but it may produce interesting ARs. These types of rules are very meaningful in discovering rarely occurring events but significantly important, such as the least ARs in air pollution, system failure, network intrusions, *etc*. From the past developments, many series of ARs mining algorithms were employed the minimum supports-confidence framework in avoiding the overloaded of ARs. The challenge is, by increasing or decreasing the minimum support or confidence values, the interesting rules might be accidently missing out or untraceable.

Since the complexity of the study, difficulties in the algorithms [17]and it may require excessive computational cost, thus very limited attentions have been paid to discover the least ARs. In term of relationship, frequent and least items in the same ARs may have a different degree of correlation. Highly correlated least ARs are referred to the certain items that their frequencies are not satisfying the minimum support but both of them are highly correlated. Moreover, those ARs must fulfil with certain degree of positive degree of correlation. Until this moment, statistical correlation technique has been successfully applied in the transaction databases [18] in finding the negative and positive correlations among items pairs. However, it is not absolutely true that all frequent items have a positive correlation as compared to the least items. In our previous works, we address the problem of mining least ARs with the objectives of discovering significant least ARs but surprisingly they are highly correlated [19-22]. A new algorithm named Significant Least Pattern Growth (SLP-Growth) to extract these ARs is also proposed [20]. The algorithm imposes interval support to extract all frequent and least items family first before continuing to construct a significant least pattern tree (SLP-Tree). The correlation technique to find the degree of correlation in ARs is also embedded in the algorithm [21].

Association rules mining have been applied to capture interesting patterns in many fields [3, 17, 22, 23]. In manufacturing industry, detecting these rules is very useful and exciting for further analysis such as for market segmentation, prediction, and product arrangements. However, it lacks of studies in discussing the association rules mining especially in vehicle industrial applications.

In this paper, we applied the SLP-Growth algorithm to capture the interesting Least ARs from two years of line-off productions of motorcycle/scooter dataset. The dataset was taken from national manufacturer company known as Motorsikal dan Enjin Nasional Sdn Bhd (MODENAS) located in Gurun Industrial Area, Kedah, Malaysia. The results of this study will provide useful information for management in MODENAS to discover the significant association between motorcycles models demand comprehensibly and to design marketing strategies accordingly. It also can be helpful for procurement or operational personnel to do the planning and forecasting more accurately.

The reminder of this paper is organized as follows. Section 2 describes the related work. Section 3 describes the essential rudiments of ARs. Section 4 describes the proposed method, SLP-Growth algorithm and Critical Relative Support (CRS) measurement. This is followed by performance analysis through motorcycle/scooter line-off production dataset in Section 5 and the results are presented in Section 6. Finally, conclusions of this work are reported in Section 7.

2. Related Works

The research interests in applying the data mining techniques in manufacturing industry are kept evolving. Until this recent, several works have been conducted with the various achievements. [8] introduced a framework to organize and apply the knowledge into decision-making process in manufacturing and service applications. In a year later, Kusiak.[9] then discussed the prospects and challenges of data mining applications in the area of product and manufacturing system design. Kruse. [10] suggested new methods for pattern discovery in automotive industry. Three stages are recommended known as development stage, the manufacturing and planning stage, and maintenance and aftercare stage. Bayesian network and Markov network are employed in these methods. Buddhakulsomsiri. [11] proposed a sequential pattern mining algorithm for automotive warranty database. The algorithm used an elementary set concept and database manipulation techniques to extract the hidden patterns from warranty claims. Mavridou. [12] introduced an agent-based framework that uses the data mining techniques to facilitate the mass customization of vehicle in automotive industry. The framework aims at supporting the provision of more personalized products and preserving the advantages of mass production. However, none of them are discussing the data mining application specifically in term of the association among vehicles' model that have been produced.

For the past years, several efforts have been made to propose the scalable and efficient methods for mining frequent ARs. However, mining least ARs is still left behind as compared to this area. As a result and based on frequent pattern mining algorithms, ARs that are rarely found in the database are difficult to discover and always prune out by the minimum support-confidence threshold. In certain domain applications, the least ARs can reveal useful information especially in detecting the exceptional situations. In term of least ARs, few works on the algorithms or methods development have been conducted. Zhou. [17] suggested a method to mine the interesting ARs by considering only infrequent items. The drawback is, Matrix-based Scheme (MBS) and Hash-based scheme (HBS) algorithms are very costly especially in term of hash collision. Ding[24] proposed Transactional Co-occurrence Matrix (TCOM for mining ARs among rare itemsets. However, the implementation wise is quite complex and very expensive. Yun.[25] introduced Relative Support Apriori Algorithm (RSAA) to generate rare itemsets. The challenge is, it takes similar amount of time taken as Apriori when the predefined minimum support threshold is set to be very low. Koh and Rountree[26] suggested Apriori-Inverse algorithm to mine infrequent itemsets without generating any frequent rules. However, it suffers from candidate itemset generations and costly in generating the rare ARs. Liu. [27] proposed Multiple Support Apriori (MSApriori) algorithm to extract the rare ARs. In actual implementation, this algorithm is unavoidable from facing the "rare item problem". From the proposed approaches [17, 24-27], many of them are using the threshold values mechanism to improve the performance of the algorithm. In term of measurements, Brin.[28] introduced objective measure called lift and chi-square as correlation measure for ARs. Lift compares the frequency of patterns against a baseline frequency as computed under statistical independence assumption. Two interesting measures based on downward closure property called all confidence and bond. Lee. [29] suggested two algorithms for mining all confidence and bond correlation patterns by extending the pattern-growth methodology. In term of mining algorithms, Agrawal. [29] proposed the first ARs mining algorithm called Apriori. The main bottleneck of Apriori is, it requires multiple scanning of transaction database and also generates a huge number of candidate itemsets. Han. [30] suggested FP-Growth algorithm which amazingly can break the two limitations as faced by Apriori like algorithms. Currently, FP-Growth is one of the fastest approach and the benchmarked algorithms for mining frequent patterns. The frequent items from transactional database are transformed into prefix tree which is known as frequent pattern tree (FP-Tree). Then, frequent patterns will be extracted by FP-Growth algorithm from this tree structure.

3. Essential Rudiments

3.1. Association Rules (ARS)

ARs were first proposed for market-basket analysis in an attempt to study customer purchasing patterns in retail stores [2]. Recently, ARs have been used in many applications or disciplines such as customer relationship management [11], image processing [31], mining air pollution data [23], educational data mining[19, 32, 33], text mining [33-35], information visualization [22, 33] and manufacturing industry [8-11], Typically, ARM is the process of discovering associations or correlation among itemsets in transaction databases, relational databases and data warehouses. There are two subtasks involved in ARs mining: generate frequent itemsets that satisfy the minimum support threshold and generate strong rules from the frequent itemsets.

Throughout this section the set $I = \{i_1, i_2, \cdots, i_{|A|}\}$, for $|A| > 0$ refers to the set of literals called set of items and the set $D = \{t_1, t_2, \cdots, t_{|U|}\}$, for $|U| > 0$ refers to the data set of transactions, where each transaction $t \in D$ is a list of distinct items $t = \{i_1, i_2, \cdots, i_{|M|}\}$, $1 \le |M| \le |A|$ and each transaction can be identified by a distinct identifier TID.

Definition 1. *A set $X \subseteq I$ is called an itemset. An itemset with k-items is called a k-itemset.*

Definition 2. *The support of an itemset $X \subseteq I$, denoted $\mathrm{supp}(X)$ is defined as a number of transactions contain X.*

Definition 3. *Let $X, Y \subseteq I$ be itemset. An association rule between sets X and Y is an implication of the form $X \Rightarrow Y$, where $X \cap Y = \phi$. The sets X and Y are called antecedent and consequent, respectively.*

Definition 4. *The support for an association rule* $X \Rightarrow Y$, *denoted* $\operatorname{supp}(X \Rightarrow Y)$, *is defined as a number of transactions in D contain* $X \cup Y$.

$$\operatorname{supp}(X \Rightarrow Y) = \mathrm{P}(X \cup Y) \tag{1}$$

Definition 5. *The confidence for an association rule* $X \Rightarrow Y$, *denoted* $\operatorname{conf}(X \Rightarrow Y)$ *is defined as a ratio of the numbers of transactions in D contain* $X \cup Y$ *to the number of transactions in D contain X. Thus*

$$\operatorname{conf}(X \Rightarrow Y) = \frac{\sup \mathrm{p}(X \Rightarrow Y)}{\sup \mathrm{p}(X)}. \tag{2}$$

An item set is a set of items. A *k*-itemset is an itemset that contains *k* items. An itemset is said to be frequent if its support value is satisfied the minimum support threshold (minsupp). A set of frequent itemsets is denoted as L_k. The support of the ARs is the ratio of transaction in *D* that contain both *X* and *Y* (or $X \cup Y$). The support can also be considered as probability $P(X \cup Y)$. The confidence of the ARs is the ratio of transactions in *D* contains *X* that also contains *Y*. The confidence can also be considered as conditional probability $P(Y|X)$. ARs that satisfy the minimum support and confidence thresholds are said to be strong.

3.2. Critical Relative Support

Critical Relative Support (CRS) proposed by [20] is a measure to capture least ARs. In this measure, the support count of antecedent, consequence and association rules are taken into account in deriving CRS value.

Definition 6. (Critical Relative Support). *A Critical Relative Support (CRS) is a formulation of maximizing relative frequency between itemset and their Jaccard similarity coefficient.*

The value of Critical Relative Support denoted as CRS and

$$\operatorname{CRS}(I) = \max\left(\left(\frac{\operatorname{supp}(\mathrm{X})}{\operatorname{suppY}}\right), \left(\frac{\operatorname{supp}(\mathrm{Y})}{\operatorname{supp}(\mathrm{X})}\right)\right) \times \left(\frac{\operatorname{supp}(X \Rightarrow Y)}{\operatorname{supp}(\mathrm{X}) + \operatorname{supp}(\mathrm{Y}) - \operatorname{supp}(X \Rightarrow Y)}\right) \tag{3}$$

CRS value is between 0 and 1, and is determined by multiplying the highest value either supports of antecedent divide by consequence or in another way around with their Jaccard similarity coefficient. It is a measurement to show the level of CRS between combination of the both Least Items and Frequent Items either as antecedent or consequence, respectively.

3.3. Correlation Analysis

After the introduction of ARs, many researches had realized the limitation of the confidence-support framework. By utilizing this framework alone, it is quite impossible to discover the interesting ARs. Therefore, Brin. [28] proposed the correlation rule and depicted together with the existing measures as

$$X \Rightarrow Y \quad (\text{supp}, \ \text{conf}, \ \text{corr}) \tag{4}$$

The correlation rule is a measure based on the minimum support, minimum confidence and correlation between itemsets *X* and Y. There are many correlation measures applicable for ARs. One of the simplest correlation measures is Lift. The occurrence of itemset *X* independent of the occurrence of itemset *Y* if $P(X \cup Y) = P(X)P(Y)$, otherwise itemset *X* and *Y* are dependence and correlated. The lift value between occurrence of itemset *X* and *Y* can be defined as:

$$\text{lift}\ (A, B) = \frac{P(X \cap Y)}{P(X)P(Y)} \tag{5}$$

The equation of (5) can be derived to produce the following definition:

$$\text{lift}\ (X, Y) = \frac{P(Y \mid X)}{P(Y)} \tag{6}$$

or

$$\text{lift}\ (X, Y) = \frac{\text{conf}\ (X \Rightarrow Y)}{\text{supp}\ (Y)} \tag{7}$$

The strength of correlation is measured based on the obtained lift value. If $\text{lift}\ (X, Y) = 1$ or $P(Y \mid X) = P(Y)$ (or $P(X \mid Y) = P(Y)$) then *Y* and *X* are independent and there is no correlation between them. If $\text{lift}\ (X, Y) > 1$ or $P(Y \mid X) > P(Y)$ (or $P(X \mid Y) > P(Y)$), then *X* and *Y* are positively correlated, meaning the occurrence of one implies the occurrence of the other. If $\text{lift}\ (X, Y) < 1$ or $P(Y \mid X) < P(Y)$ (or $P(X \mid Y) < P(Y)$), then *X* and *Y* are negatively correlated, meaning that the occurrence of one discourage the occurrence of the other. Since lift measure is not down-ward closed, it will not be suffered from the least item problem. Thus, least itemsets with low support counts which per chance occur a few times (or only once) together can produce enormous of lift values.

3.4. FP - Growth

Candidate set generation and tests are two major drawbacks in Apriori-like algorithms. Therefore, to deal with this problem, a new data structure called frequent pattern tree (FP-Tree) was introduced. FP-Growth was then developed based on this data structure and currently is a benchmarked and fastest algorithm in mining frequent itemset [36]. The advantages of FP-Growth are, it requires two times of scanning the transaction database. Firstly, it scans the database to compute a list of frequent items sorted by descending order and eliminates rare items. Secondly, it scans to compress the database into a FP-Tree structure and mines the FP-Tree recursively to build its conditional FP-Tree.

A simulation data [12] is shown in Table 1. Firstly, the algorithm sorts the items in transaction database with infrequent items are removed. Let say a minimum support is set to 3, therefore alphabets f, c, a, b, m, p are only kept. The algorithm scans the entire transactions start from T1 until T5. In T1, it prunes from {f, a, c, d, g, i, m, p} to {f, c, a, m, p}. Then, the algorithm compresses this transaction into prefix tree which *f* becomes the root. Each path on the tree represents a set of transaction with the same prefix. This process will execute recursively until the end of transaction. Once the complete tree has been built, then the next pattern mining can be easily performed (For details, see Tables 2 and 3 and Figure 1).

Table 1. A Sample Data

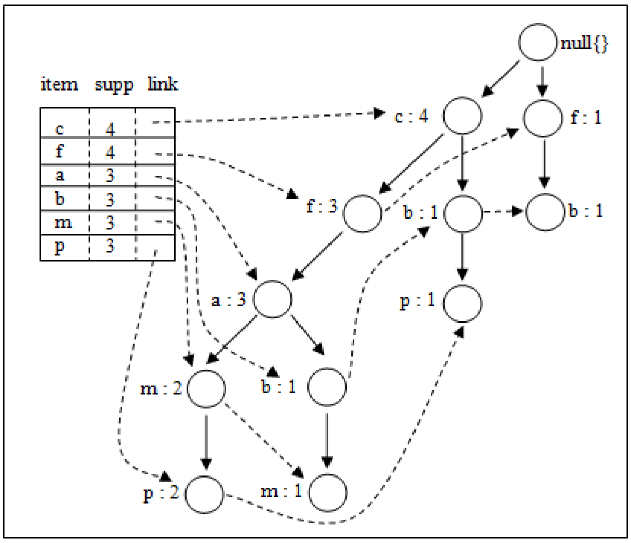
TID	Items
T1	a c m f p
T2	a b c f l m o
T3	b f h j o
T4	b c k s p
T5	a f c e l p m n

Table 2. FP-Growth – Constructing Conditional Items

TID	Items	Sorted Items	Conditional Item
T1	a c m f p	f a c m p	p: f c a m m: f c a a: f c c: f
T2	a b c f l m o	f c a b m	m: f c a b b: f c a a: f c c: f
T3	b f h j o	f b	b: f
T4	b c k s p	c b p	p: c b b: c
T5	a f c e l p m n	f c a m p	p: f c a m m: f c a a: f c c: f

Table 3. FP-Growth - Constructing Conditional FP-Tree

Item	Conditional FP-Tree
c	f : 3
a	f c : 3
b	f c a : 1 f : 1 c : 1
m	f c a : 2 f c a b : 1
p	f c a m : 2 c b : 1



item	supp	link
c	4	
f	4	
a	3	
b	3	
m	3	
p	3	

Figure 1. Constructing the FP-Tree

4. Proposed Method

4.1. Algorithm Development

4.1.1. Determine Interval Support for Least Itemset

Let *I* is a non-empty set such that $I = \{i_1, i_2, \cdots, i_n\}$, and D is a database of transactions where each T is a set of items such that $T \subset I$. An item is a set of items. A k-itemset is an itemset that contains k items. An itemset is said to be least if the support count satisfies in a range of threshold values called Interval Support (ISupp). The Interval Support is a form of ISupp (ISMin, ISMax) where ISMin is a minimum and ISMax is a maximum values respectively, such that $\text{ISMin} \geq \phi$, $\text{ISMax} > \phi$ and $\text{ISMin} \leq \text{ISMax}$. The set is denoted as L_k. Itemsets are said to be significant least if they satisfy two conditions. First, support counts for all items in the itemset must greater ISMin. Second, those itemset must consist at least one of the least items. In brevity, the significant least itemset is a union between least items and frequent items, and the existence of intersection between them.

4.1.2. Construct Significant Least Pattern Tree

A Significant Least Pattern Tree (SLP-Tree) is a compressed representation of significant least itemsets. This trie data structure is constructed by scanning the dataset of single transaction at a time and then mapping onto path in the SLP-Tree. In the SLP-Tree construction, the algorithm constructs a SLP-Tree from the database. The SLP-Tree is built only with the items that satisfy the ISupp. In the first step, the algorithm scans all transactions to determine a list of least items (LItems) frequent items (FItems) and least frequent item (LFItems). LFItems is a combination of LItems and FItems. In the second step, all transactions are sorted in descending order and mapping against the LFItems. It is a must in the transactions to consist at least one of the least items. Otherwise, the transactions are disregard. In the final step, a transaction is transformed into a new path or mapped into the existing path. This final step is continuing until end of the transactions. The problem of existing FP-Tree are it may not fit into the memory and expensive to build. FP-Tree must be built completely from the entire transactions before calculating the support of each item. Therefore, SLP-Tree is an alternative and more practical to overcome these limitations.

4.1.3. Generate Least Pattern Growth (Lp-Growth)

SLP-Growth is an algorithm that generates significant least itemsets from the SLP-Tree by exploring the tree based on a bottom-up strategy. 'Divide and conquer' method is used to decompose task into a smaller unit for mining desired patterns in conditional databases, which can optimize the searching space. The algorithm will extract the prefix path sub-trees ending with any least item. In each of prefix path sub-tree, the algorithm will recursively execute to extract all frequent itemsets and finally built a conditional SLP-Tree. A list of least itemsets is then produced based on the suffix sequence and also sequence in which they are found. The pruning processes in SLP-Growth are faster than FP-Growth since most of the unwanted patterns are already cutting-off during constructing the SLP-Tree data structure. The complete SLP-Growth algorithm is shown in Figure 2.

4.2. Weight Assignment

4.2.1. Apply Measures

The weighted ARs (ARs values) are derived from the formula as described in (1-3, 5). The processes of generating weighted ARs which consist of support, confidence, correlation and CRS are taken place once all patterns are completely produced.

4.2.2. Discover Highly Correlated Least ARS

From the list of weighted ARs, the algorithm will begin to scan all of them. However, only those weighted ARs that have the correlation value more than the predefined correlation threshold are captured and considered as highly correlated. The others ARs will be pruned out and classified as low correlation.

```
Input: Dataset D, ISupp (ISMin, ISMax)
Output: ARs
for items, I in transaction, T do
     Determine support count, ItemSupp
end for loop
Sort ItemSupp in descending order, ItemSuppDesc
for ItemSuppDesc do
     Generate List of frequent items, FItems > ISMax
end for loop
for ItemSuppDesc do
     Generate List of least items,
       ISMin <= LItems < ISMax
end for loop
Construct Frequent and Least Items,
       FLItems = FItems U LItems
for all transactions,T do
  if (LItems ∩ I in T > 0) then
     if (Items in T = FLItems) then
        Construct items in transaction in Descending order, TItemsDesc
     end if
   end if
end for loop
for TItemsDesc do
     Construct SLP-Tree
end for loop
for all prefix SLP-Tree do
     Construct Conditional Items, CondItems
end for loop
for all CondItems do
     Construct Conditional SLP-Tree
end for loop
for all Conditional SLP-Tree do
     Construct Association Rules, AR
end for loop
for all AR do
     Calculate Support and Confidence
     Apply Correlation
end for loop
```

Figure 2. SLP-Growth Algorithm

5. Scenario of Capturing Rules

5.1. Dataset

The experiment was conducted on motorcycle/scooters line-off production dataset. The dataset was obtained from Motosikal dan Enjin Nasional Sdn Bhd (MODENAS) based on the two years of motorcycle/scooters production (2006 to 2007). For monthly motorcycle/scooter scheduling, it uses the information taken from the authorized dealers throughout Malaysia and overseas. The dealers will make sales orders manually to MODENAS subsidiary company known as Edaran Modenas Sdn Bhd (EMOS) and these data will be input into Vehicle Management System (VMS). The main data sources required for VMS in the “motorcycle request and fill-in” module are a dealer particular, motorcycles/scooters models, colors and quantity. In every week, a report of motorcycles/scooters orders will be submitted to several departments in MODENAS. After several internal processes in MODENAS, the data will be input into Material Resource Planning System (MRP) which is running on AS400 Machine. In production term, the motorcycles/scooters are only considered as line-off (completely assemble with a good condition) after they satisfied the quality control checking. The dataset for this experiment is generated from the MRP system and comprises of 31 attributes (models) and 24 records (after compiling from 630 original records without actual date). The rationale behind to compile these numbers into only 24 records are due to the uncontrollable factors in production line such as the availability of colors, motorcycle/scooter parts (components), *etc*. To this, we have a dataset comprises the number of transactions (monthly) is 24 and the number of items (attributes) is 31 (refers to Table 4).

Table 4. Line-Off Production of Motorcycle/Scooter Dataset for 2006-2007

Dataset	Size	# Transactions	# Items
Line-Off Production	4,096 bytes	24	31

5.2. Design

The design for capturing interesting rules from motorcycle/scooters line-off production is illustrated in the following Figure 3.

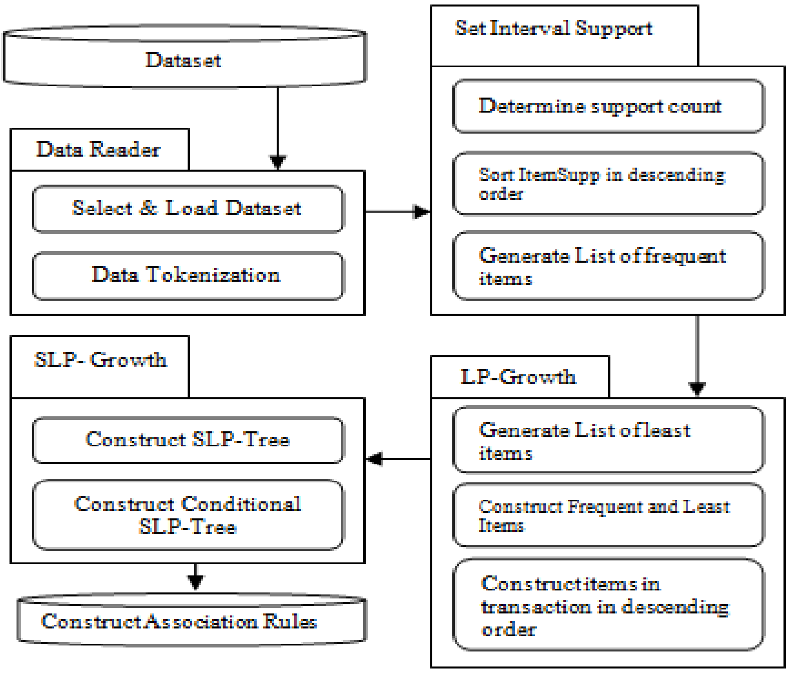

Dataset
Data Reader
Select & Load Dataset
Data Tokenization
Set Interval Support
Determine support count
Sort ItemSupp in descending order
Generate List of frequent items
SLP- Growth
Construct SLP-Tree
Construct Conditional SLP-Tree
LP-Growth
Generate List of least items
Construct Frequent and Least Items
Construct items in transaction in descending order
Construct Association Rules

Figure 3. The Procedure of Mining Critical Least Association Rules

In order to capture the interesting rules and make a decision, the experiment using SLP-Growth method will be conducted on Intel® Core™ 2 Quad CPU at 2.33GHz speed with 4GB main memory, running on Microsoft Windows Vista.

The algorithm has been developed using C# as a programming language. The motorcycle/scooters line-off production used in this model are in a format of flat file.

6. Results

The original data was given in the horizontal format which is only suitable for reporting purposes. It consists of 3 attributes: date, color and total. Due to the confidentiality matters, the actual total is represented in a form of range. Table 5 shows the mapping between the model and a new suggested code. The mapping between the range of quantity and a new recommended code is presented in Table 6. A new set of data was generated based on the combinations of a new model code and a new range of quantity code. For example, if the model is AN110E-D310 and the total production is 162, therefore the item 102 will be appeared in the new dataset. The first 2 number is corresponding to a new model code and the last number is for the range of quantity code. ARs were generated in a form of many-to-one cardinality relationship and the maximum number of antecedents was set to six. The summary of selected ARs for analysis is shown in Table 7.

By embedding FP-Growth algorithm, 2,785 ARs are produced. ARs are formed by applying the relationship of an item or many items to an item (cardinality: many-to-one). Figure 4 depicts the correlation's classification of interesting ARs. For this dataset, the rule is categorized as significant and interesting if it has minimum support more than 5%, must be positive correlation (more than 1.00) and CRS value should be at least 0.8.

Table 7 shows top 20 of interesting ARs with numerous types of measurements. The highest correlation value from the selected ARs is 12.00 (No. 1 to 20). From these ARs, there are three dominant of consequence items, item 341, 151 and 231. In fact, item 341 only appears 8.33% from the entire dataset. Item 341 is for the motorcycle model of "AN120H-A1MY" and its range of quantity is less than 100. For item 151, it is stand for the motorcycle model of "MN120H-A1MY" and also its range of quantity is not more than 100. For the last item, it represents the motorcycle model of "MA120F-A1MY" and it has the range of quantity as similar to the previous two motorcycle models. For both "MN120H-A1MY" and "MA120F-A1MY" models, their occurrences are also 8.33% from the rest of dataset. Table 6 also indicates that all interesting ARs have a value of CRS is equal to 1. Figure 5 illustrates the summarization of correlation analysis with different Interval Support.

From Table 7, the average ratio between antecedent and consequences is 2:1. Its means that, Indeed, extensive research can be carried out to increase the sales of least motorcycle/scooter models by identifying the unique characteristics in the famous models. In others perspective, influenced factors that contributed into the lower sales of this model can be used as a guideline for others model in the future. It can reduce unnecessary cost at the beginning stage before producing unpopular models in the market.

Table 5. The Mapping of a Model and a New Code

No	Model	Code
1	AN110E-D310	10
2	AN110E-E310	11
3	AN110F-D310	12
:	:	:
31	SN150J-A1MY	41

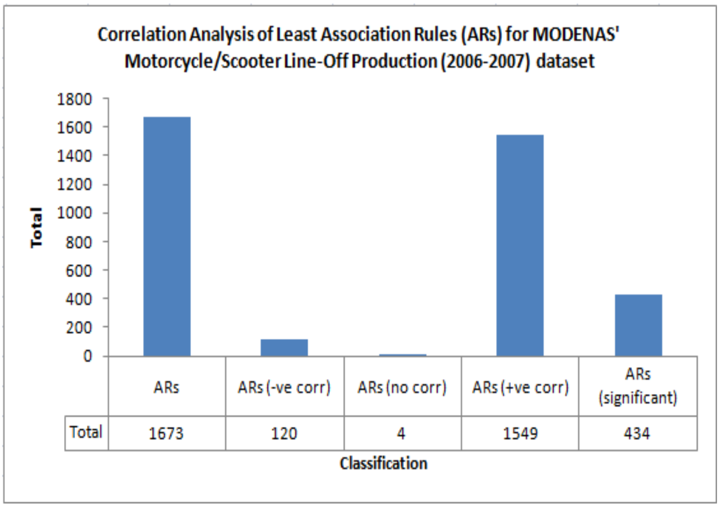
Table 6. The Mapping of the Range of Quantity and a New Code

Quantity Range	Code
< 100	1
100 – 400	2
> 400	3

Table 7. Top 20 of Highest Correlation of Interesting Association Rules Sorted In Descending Order of Correlation

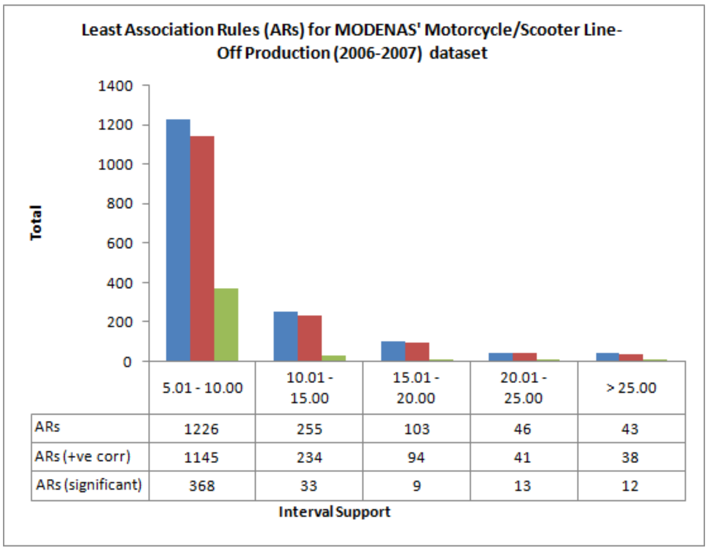
No	Association Rules	Supp	Conf	Corr	CRS
1	273 192 402 392 → 341	8.33	100.00	12.00	1.00
2	412 273 192 402 392 → 341	8.33	100.00	12.00	1.00
3	273 402 392 → 341	8.33	100.00	12.00	1.00
4	412 273 402 392 → 341	8.33	100.00	12.00	1.00
5	412 273 142 402 392 → 341	8.33	100.00	12.00	1.00
6	273 192 142 402 392 → 341	8.33	100.00	12.00	1.00
7	273 142 402 392 → 341	8.33	100.00	12.00	1.00
8	412 353 211 383 → 151	8.33	100.00	12.00	1.00
9	412 211 383 → 151	8.33	100.00	12.00	1.00
10	211 372 383 → 151	8.33	100.00	12.00	1.00
11	353 211 372 383 → 151	8.33	100.00	12.00	1.00
12	353 211 383 → 151	8.33	100.00	12.00	1.00
13	211 383 → 151	8.33	100.00	12.00	1.00

14	412 211 372 383 → 151	8.33	100.00	12.00	1.00
15	412 353 211 372 383 → 151	8.33	100.00	12.00	1.00
16	273 211 382 191 → 231	8.33	100.00	12.00	1.00
17	412 273 211 191 → 231	8.33	100.00	12.00	1.00
18	273 211 191 → 231	8.33	100.00	12.00	1.00
19	412 273 211 382 191 → 231	8.33	100.00	12.00	1.00
20	412 273 382 191 → 231	8.33	100.00	12.00	1.00



	ARs	ARs (-ve corr)	ARs (no corr)	ARs (+ve corr)	ARs (significant)
Total	1673	120	4	1549	434

Figure 4. Classification of ARS using Correlation Analysis



	5.01 - 10.00	10.01 - 15.00	15.01 - 20.00	20.01 - 25.00	> 25.00
ARs	1226	255	103	46	43
ARs (+ve corr)	1145	234	94	41	38
ARs (significant)	368	33	9	13	12

Figure 5. Correlation Analysis of Interesting ARS Using Variety Interval Supports

7. Conclusion

Finding the least association rules from the various models of motorcycle/scooter is very difficult. At the moment, there is nearly no such study has been performed to understand the relationship among different types of models in automotive or manufacturing industry. This information is vital, but it is not easy to capture by typical statistic or data mining approaches. In this paper, we had successfully applied an enhanced association rules mining method, so called SLP-Growth (Significant Least Pattern Growth) and a new measurement named Critical Relative Support (CRS) proposed by [20] for capturing interesting rules line-off production of motorcycle/scooter dataset. The data was taken from our national motorcycle company called Motorsikal dan Enjin Nasional Sdn Bhd (MODENAS) located in Kedah, Malaysia. It is found that SLP-Growth method is suitable to mine the interesting rules which provide faster and accurate results. The results from this research will provide useful information for the management to understand the customer demand trends comprehensibly, and to design marketing strategy accordingly.

Acknowledgment

This research is supported by Universitas Teknologi Yogyakarta Research Grant no vote 556/UW-R/0/X/2013. The authors would like to thanks Mr. Jefri Jasin from department of Information System Technology, MODENAS for providing the data in the journey of completing this work.

References

[1] W.J. Frawley, G. Piatetsky-Shapiro and C. J. Matheus, “Knowledge discovery in databases: An overview”, AI magazine, vol. 13, **(1992)**, pp. 57.

[2] P.-N. Tan, M. Steinbach and V. Kumar, “Introduction to data mining vol. 1: Pearson Addison Wesley Boston, **(2006)**.

[3] T.T.S. Nguyen, H.Y. Lu, T.P. Tran and J. Lu, “Investigation of sequential pattern mining techniques for web recommendation”, International Journal of Information and Decision Sciences, vol. 4, **(2012)**, pp. 293-312.

[4] L. Lin, M.-L. Shyu and S.-C. Chen, “Association rule mining with a correlation-based interestingness measure for video semantic concept detection”, International Journal of Information and Decision Sciences, vol. 4, **(2012)**, pp. 199-216.

[5] P. Pahwa, R. Arora and G. Thakur, “An efficient algorithm for data cleaning”, Intelligence Methods and Systems Advancements for Knowledge-Based Business, **(2012)**, pp. 305.

[6] A. Murthy and V. Nagadevara, “Predictive Models in Cybercrime Investigation: An Application of Data Mining Techniques”, Advancing the Service Sector with Evolving Technologies: Techniques and Principles: Techniques and Principles, **(2012)**, p. 166.

[7] Z. Wang, R. Yan, Q. Chen and R. Xing, “Data mining in nonprofit organizations, government agencies, and other institutions”, Advancing the Service Sector with Evolving Technologies: Techniques and Principles: Techniques and Principles, **(2012)**, pp. 208.

[8] A. Kusiak, “Data mining: manufacturing and service applications”, International Journal of Production Research, vol. 44, **(2006)**, pp. 4175-4191.

[9] A. Kusiak and M. Smith, “Data mining in design of products and production systems”, Annual Reviews in Control, vol. 31, **(2007)**, pp. 147-156.

[10] R. Kruse, M. Steinbrecher and C. Moewes, “Data Mining Applications in the Automotive Industry”, in 4th International Workshop on Reliable Engineering Computing (REC 2010) Edited by Michael Beer, Rafi L. Muhanna and Robert L. Mullen Copyright, **(2010)**, pp. 23-25.

[11] J. Buddhakulsomsiri and A. Zakarian, “Sequential pattern mining algorithm for automotive warranty data”, Computers & Industrial Engineering, vol. 57, **(2009)**, pp. 137-147.

[12] E. Mavridou, D. D. Kehagias, K. Kalogirou and D. Tzovaras, “An agent-oriented data mining framework for mass customization in the automotive industry”, in Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-vol. 03, **(2008)**, pp. 575-578.

[13] A. Ceglar and J.F. Roddick, “Association mining”, ACM Computing Surveys (CSUR), vol. 38, **(2006)**, p. 5.

[14] R. Agrawal, T. Imielinski and A. Swami, “Database mining: A performance perspective”, Knowledge and Data Engineering, IEEE Transactions on, vol. 5, **(1993)**, pp. 914-925.

[15] R. Agrawal, T. Imieliński, and A. Swami, “Mining association rules between sets of items in large databases”, in ACM SIGMOD Record, **(1993)**, pp. 207-216.

[16] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules”, in Proc. 20th int. conf. very large data bases, VLDB, **(1994)**, pp. 487-499.

[17] L. Zhou and S. Yau, “Association rule and quantitative association rule mining among infrequent items”, in Proceedings of the 8th international workshop on Multimedia data mining:(associated with the ACM SIGKDD 2007), **(2007)**, pp. 9.

[18] H. Xiong, S. Shekhar, P.-N. Tan and V. Kumar, “Exploiting a support-based upper bound of Pearson's correlation coefficient for efficiently identifying strongly correlated pairs”, in Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, **(2004)**, pp. 334-343.

[19] Z. Abdullah, T. Herawan, N. Ahmad and M. M. Deris, “Extracting highly positive association rules from students’ enrollment data”, Procedia-Social and Behavioral Sciences, vol. 28, **(2011)**, pp. 107-111.

[20] Z. Abdullah, T. Herawan and M. M. Deris, “Mining significant least association rules using fast SLP-growth algorithm”, in Advances in Computer Science and Information Technology, ed: Springer, **(2010)**, pp. 324-336.

[21] Z. Abdullah, T. Herawan and M.M. Deris, “Scalable model for mining critical least association rules”, in Information Computing and Applications, ed: Springer, **(2010)**, pp. 509-516.

[22] Z. Abdullah, T. Herawan and M. M. Deris, “Visualizing the construction of incremental disorder Trie Itemset data structure (DOSTrieIT) for frequent pattern tree (FP-tree)”, in Visual Informatics: Sustaining Research and Innovations, ed: Springer, **(2011)**, pp. 183-195.

[23] M. Mustafa, N. Nabila, D. Evans, M. Saman and A. Mamat, “Association rules on significant rare data using second support”, International Journal of Computer Mathematics, vol. 83, **(2006)**, pp. 69-80.

[24] J. Ding, “Efficient association rule mining among infrequent items: University of Illinois at Chicago”, 2005.

[25] H. Yun, D. Ha, B. Hwang, and K. H. Ryu, “Mining association rules on significant rare data using relative support”, Journal of Systems and Software, vol. 67, **(2003)**, pp. 181-191.

[26] [26]Y. S. Koh and N. Rountree, “Finding sporadic rules using apriori-inverse”, in Advances in Knowledge Discovery and Data Mining, ed: Springer, **(2005)**, pp. 97-106.

[27] B. Liu, W. Hsu and Y. Ma, “Mining association rules with multiple minimum supports”, in Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, **(1999)**, pp. 337-341.

[28] S. Brin, R. Motwani and C. Silverstein, “Beyond market baskets: Generalizing association rules to correlations”, in ACM SIGMOD Record, **(1997)**, pp. 265-276.

[29] Y.-K. Lee, W.-Y. Kim, Y. D. Cai and J. Han, “CoMine: Efficient Mining of Correlated Patterns”, in ICDM, **(2003)**, pp. 581-584.

[30] J. Han, J. Pei, Y. Yin and R. Mao, “Mining frequent patterns without candidate generation: A frequent-pattern tree approach”, Data mining and knowledge discovery, vol. 8, **(2004)**, pp. 53-87.

[31] C.C. Aggarwal and P.S. Yu, “A new framework for itemset generation”, in Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems, **(1998)**, pp. 18-24.

[32] Z. Abdullah, T. Herawan, N. Ahmad and M. M. Deris, “Mining significant association rules from educational data using critical relative support approach”, Procedia-Social and Behavioral Sciences, vol. 28, **(2011)**, pp. 97-101.

[33] T. Herawan, I.T.R. Yanto and M. M. Deris, “Soft set approach for maximal association rules mining”, in Database Theory and Application, ed: Springer, **(2009)**, pp. 163-170.

[34] T. Herawan and M.M. Deris, “A soft set approach for association rules mining”, Knowledge-Based Systems, vol. 24, **(2011)**, pp. 186-195.

[35] T. Herawan, P. Vitasari and Z. Abdullah, “Mining interesting association rules of student suffering mathematics anxiety”, in Software Engineering and Computer Systems, ed: Springer, **(2011)**, pp. 495-508.

[36] J. Han, J. Pei and Y. Yin, “Mining frequent patterns without candidate generation”, in ACM SIGMOD Record, **(2000)**, pp. 1-12.