Information Technology Journal 3 (2): 150-162, 2004
ISSN 1682-6027

Identifying the Common and Critical Parts for MTO Products Using a Data Mining Approach

Kuo-Qin Yan and [1]Shu-Ching Wang
Department of Business Administration, [1]Department of Information Management,
Chaoyang University of Technology
168 Gifeng E. Rd., Wufeng, Taichung County, Taiwan, 413, Republic of China

Abstract: For a manufacturer which bases make-to-order (MTO) as its product positioning strategy, in such competitively growing business in which customers demand a shorter and shorter delivery time, it's important for the manufacturer to avoid unnecessary inventory of semi-finished goods but improve the flexibility of delivery and to minimize the uncertain interferences in the middle of manufacturing in order to better respond to the customers. We propose a four-step method. In the first step, employing RFM indexing method to find the most potential profit-maximizing product from the product line for export. Second step is to do a mining on sequence pattern of the bill of material (BOM) which is picked from the RFM to obtain the large frequent sequence pattern of processing. From these largest sequence patterns, in the third step, create the corresponding sequence rules and present in visualization the common parts BOM and critical parts. Fourth step is to generate a feasible as well as safe optimized quantity of production. Our proposed method can, provide as an assistance to the enterprises that they can better control the commonness of their products and processes and help them find the most appropriate parts or semi-finished goods for make-to-stock during the safe planning horizon so that they can effectively and accurately increase the inventory of these critical parts and minimize the uncertain interferences in the middle of manufacturing in order to better respond to the customers.

Key words: Decision support, data mining, inventory management strategy, information technology management

INTRODUCTION

Manufacturing competing strategy in terms of order of priority, or manufacturing competitive ability index, includes four critical elements: "quality", "cost", "time" and "flexibility"[1]. From customers' perspective when choosing a supplier, "quality" means low defect rate and reliable/durable product manufactured. "Cost" means price for the product. For suppliers it's the manufacturing cost of the product and the sales price minus the manufacturing cost is the profit. "Time" means on-time delivery, fast delivery and development speed of new products. "Flexibility" means the ability of supplier to customize product for customers, a broad product line, rapid volume change according to order placement, rapid product mix change and rapid design change.

For a manufacturer, subjects such as reducing the manufacturing cost, decreasing the defect rate, increasing the product quality and meeting the shorter and shorter delivery time have been actively under way for improvement in operational effectiveness (OE) of many manufacturers. Looking at those enterprises which base on make-to-order strategy, we know they are basically similar in view of their strategies, market segmentation and quality and price of products with one another; as a consequence, their quality and price are determinants of their survival instead of their competitive advantages. As a matter of fact, the critical factors in regard to competition of make-to-order enterprises are the length of time from placement of order to delivery which we call it lead time, the delivery schedule, quantitative steadiness and flexibility[1,2].

Order decoupling point of make-to-order based strategy is at raw material step[3-6]. Once an order is placed, it starts to activate the internal value chain mechanism of the enterprise, purchase raw materials, make arrangement on production schedule etc. A series of time points of production are created right after the customer places the order which occurs at the initial raw material purchase step, thus make-to-order based enterprises will have lower

Corresponding Author: Shu-Ching Wang, Department of Information Management, Chaoyang University of Technology, 168 Gifeng E. Rd., Wufeng, Taichung County, Taiwan, 413, Republic of China Tel: 886-4-2332-3000 ext. 3131 Fax: 886-4-2374-2337 E-mail: scwang@mail.cyut.edu.tw

inventory of final goods, semi-finished goods and products but longer lead time than make-to-stock enterprises. Jan Olhager has mentioned[4] from the perspective of strategic competitive advantage in terms of the impact of order decoupling point on production strategy, the make-to-order enterprises which need to solve the problem of lengthy lead time can strategically select to forward shift the order decoupling point and convert it to assembly-to-order (ATO) production strategy or to keep the current order decoupling point in which it maintains the advantages of lower inventory cost and customized product and attempt to reduce the lead time as well as enhance and improve the efficiency of flexibility.

In a number of current references that have been published in regard to production plan[7-10] most of which employs operation research and multi criteria decision making methods. With lowest inventory cost and shortest lead time as target functions plus limitations such as service standard, shortage cost, set-up cost and so forth, establish a mathematical model used to find the optimized quantity of production which is kind of mainly quantity planning based. Another method is by J. Huiskonen who values the characteristics of product and marketing which contribute to the sales volume and turnover by identifying which products are to be produced under different production strategies[11]. For researches on product modulation, which is mostly assumed that common modules used for the products are given, improving traditional BOM in data structure of information system is therefore the main direction of research[12-14]. Some examples are-generic BOM (GBOM), planning BOM (PBOM), modular BOM (MBOM), super BOM (SBOM), add/delete BOM (A/D BOM), bill of manufacture (BOMFR) etc.

As for finding the common modules from current BOM of the enterprise and using these common modules to determine an optimized quantity of production are very rare. Especially for industries with their products cannot be easily modularized or make-to-order based industries which may not be able to modularize their product lines in a short period of time, we, however, propose a method which can help them to keep the current architecture of information systems on-the-fly without making big change in structure of their product materials, take advantage of the past sales history and combine with data mining so as to instantly and dynamically identify the group of profit-maximizing products, divide vertically the frequent sequence rules of BOM to obtain the common modules among the enterprises, take from these modules and common manufacturing rules as basis of inventory production. And lastly apply these common rules to make a proper optimized production plan so that the enterprises can have higher flexibility of production and swiftly shift their resources of production to better respond to the unexpected.

Buffer strategy: Table 1 lists the characteristics of different production environments. Enterprises, due to different industry environments, will have different properties of products[2]. The first three properties: products, products demand and capability are determined by the industry environment under which the enterprise is. Some limitations such as the enterprise has the standardized products, the products have to be entirely customized according to the customers' requirements, the products demand are forecasted and capability is carefully and thoroughly planned in advance are basically fixed, the enterprise will not change its proper buffer strategy easily. To determine a proper buffer strategy needs careful consideration in terms of two dimensions-quantity demanded and types of products available. For industries with large quantity of products demanded and less changes in specifications, they are better off by going the route of make-to-stock (MTS); for industries with less quantity of products demanded and entire customization of products, they are more inclined to go the route of engineer-to-order (ETO); for industries in between, they need make-to-order (MTO) or assembly-to-order (ATO) based strategies depending if there is pre-defined group of products available[4,6]. Under different manufacturing environments, enterprises respond to customers differently; therefore, a buffer strategy under certain manufacturing environment is a consequence composed of products, competitions and strategies altogether[4,15].

For make-to-order (MTO) based industries, their bill of material (BOM) is usually presented in V shape. The bottom part of V shape is made up of a couple of pure raw materials that form its acute shape. In the middle part of V shape or during the manufacturing, there are plenty of combinations between processing procedures and manufacturing, in which their various relationships will result in more optional final goods in the final goods stage. These products which differ slightly among themselves form the aperture in the V shape of BOM[16]. Make-to-order enterprises, due to a number of varying options on processing by small and medium businesses as they cannot provide the entire processing capability or their not having the relative economic scale, may be involved in more outsourcing of the processing, management contracts and longer production time and higher variation and complexity of controls. A slight alteration in these factors may affect the promised delivery time, that is, the ability of on-time delivery.

Table 1: Characteristics of different production environments[2]

Properties	MTS	ATO	MTO	ETO
Products	Standardization	Pre-defined group of products	Customization	Entire customization
Products demand	Can forecast			Cannot forecast
Capability	Can plan			Cannot plan
Lead time	Unimportant	Important	Important	Most important
Critical competition	Distribution	Last Assembly	Manufacture, Last Assembly	Entire process flow
Process complexity	Distribution	Assembly	Part Manufacture	Development
Process uncertainly	Lowest			Highest
High-level management focus	Marketing/delivery	Creativity	Capability	Customer contract
Mid-level management focus	Inventory management	MPS/Customer order	Monitor and control On-The-Fly/ Customer order	Project management

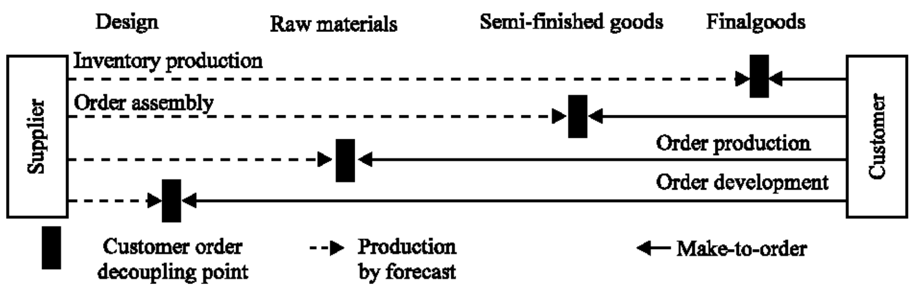

Fig. 1: Order decoupling points under different production environments[2]

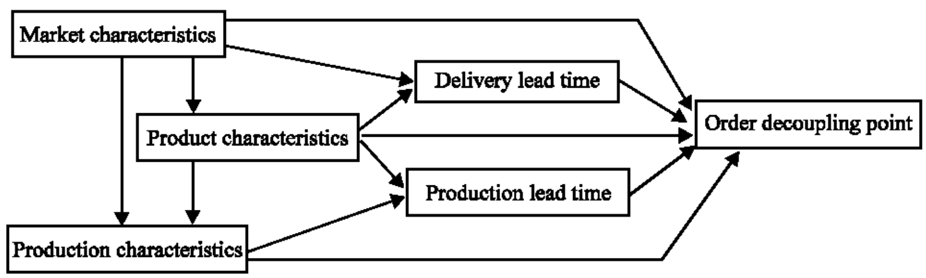

Fig. 2: Factors affecting the order decoupling point[4]

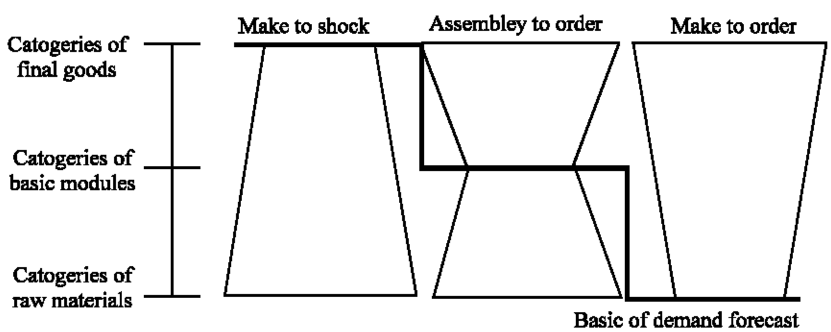

Fig. 3: Basis of demand forecast for make-to-order enterprises[2]

Because of various options of products and uncertain factors related to the customers, make-to-order enterprises usually cannot forecast accurately about the market's demand and scale in advance in order to make sure they have sufficient inventory to buffer the uncertainty and unexpected demand of customers. As shown in Fig. 1, in the make-to-order based enterprises, the main raw materials are the basic materials of all products, it's feasible to produce or purchase through forecast; yet during the processing of components to semi-finished goods to final goods in which there will be some big customizations needed to be made that makes difficult to forecast the demand, the enterprises must go fully along with customer's order on master product scheduling

(MPS). That customer places the order has determined the activation of internal value chain activity in raw material step for those make-to-order enterprises, this draws the line between forecast and single production; the preceding activities are forecast driven to produce or purchase, the following activities are customer driven to produce and process by specifications shown on the order. This point is called order penetration point (OPP)[4] or order decoupling point (ODP)[5,6] .

Of the range from which the enterprise can choose the order decoupling point, also the range within which the enterprise can choose the production strategy, the main determining factors are market characteristics, product characteristics, production characteristics, delivery lead time and production lead time[4]. Fig. 2 presents the interactive relationship among these five elements or factors for the order decoupling point. If reducing the delivery dead time is necessary for obtaining orders, the make-to-order based enterprises, however, must adequately improve its manufacturing efficiency in the way that shifts the order decoupling point forward.

Figure 3 explains, for those make-to-order enterprises wanting to lessen the impact of variations that may occur during the manufacturing and decrease the delivery lead time, they should upgrade the forecast which is based on raw material step to basic module step as much as they possibly can; in other words, they need to enhance the common property and modularization and make the basic module by inventory production through forecasting, in this way they can effectively reduce the delivery lead time and increase the delivery flexibility. The purpose of our research is, through the sequence pattern mining of BOM, to find the largest common part and basic module in V-shaped BOM for the make-to-order enterprises.

RFM model: In application of database direct marketing, RFM Model is usually used for quantification assessment that is carried out on customer and product values[17]. R, stands for Recency, represents the most recent purchase that has been made. F, stands for Frequency, is the frequency of purchases that have been made. M, stands for Monetary, is the amount of money spent for purchase. RFM Model uses these three, visually simple but quite effective quantification indices to assess the customer purchase behavior or attractiveness of the product to the customer, which is still commonly used in applications of database marketing or relationship marketing[18,19]. "Recency" computes the length of time from the last purchase made on the product or by the customer up to present. The shorter time distance of purchase made on the product, the higher probability that this product will be purchased next time. The longer time from last purchase made on the product up to present suggests that this product may be outdated or the customer has made his purchase elsewhere. "Frequency" measures the number of times of purchase made on the product during a certain time period. The length of time could be one quarter or one month; this is generally affected by the characteristics of different industries, characteristics of products and the cycle time of purchase. "Frequency" shows that the interaction between products and customers; the higher the frequency of purchase made, the more this product draws the customer's attention and the higher frequency the customer will buy again. "Monetary" determines the total amount of sales on the products during a certain time period. The higher amount of money spent on certain product means that the customer is more interested on this product; in other words, this product is likely to contribute more to company's profit in the future. When the customer spends more on certain product, it means that the customer has large demand for this product and is very likely to buy again. Getting a hang and understanding the most valuable group of customers as well as products can help designing a more accurate and effective promotion solution and preventing from unnecessary cost of promotion and inventory on invaluable group of customers and products. Our method uses RFM Model to identify, from the product line, the value of inventory production of product and the product of high inventory production value as a forecast that the customer in the future will place order for a certain frequency, quantity and money on the product. Maintaining a proper inventory on this type of products will improve and increase the enterprise's ability to respond to the customers in order to satisfy the customer on changing delivery flexibility.

Sequence rules: Data mining is also known as knowledge discovery, in which its purpose is to mine the hidden, past unknown but may be valuable information in future application such as knowledge rule, constraint and regularity[20]. The earliest proposed and most widely-used method in data mining is called association rule. Association rule analyzes and matches the transactional data in relational database to mine the association in transactions which customer purchases the product[21]. Association rule uses the following to present the reliability of association and rule between products: $A \cap B \rightarrow C$ [Support:S,Confidence:C]. Support S represents the percentage of the simultaneous occurrences of both A, B and C over the total count of the transactional records. Confidence C is the percentage in the transactional records which have the occurrence of both

A and B and contains C. The definition tells that the association rule is not picking on the sequence of product items in a single transactional record. For example, Transaction 1 contains three items-A, B and C. Transaction 2 contains items-B, C and A. Although the three items occur in different sequences on both transactions. Transactions 1 and 2 are both regarded as the same pattern. On some circumstances, in addition to taking account of the product that is purchased in the transaction, the time when the transaction occurs is another important factor to be considered also. For this reason, Agrawal and Srikant[22] first proposes a sequence rule mining method based on Apriori, in which its aims at solving the problem when there is difference on the order of priority in the transactional items. Sequence rule is defined as follows: If one sequence $(a_1,a_2,\ldots,a_n)$ contains another sequence $(b_1,b_2,\ldots,b_n)$ and exists in $i_1<i_2<\ldots<i_n$, then $b_1\subseteq a_1$, $b_2\subseteq a_2$, …, $b_n\subseteq a_n$. Hence, the sequence rule is generated as a result of finding the frequent sequence that is bigger than the user setting but of smallest support from the customer transactional database D. In each transaction from D, it contains the customer number, transactional time, purchased product item in which the sequence of each transaction is determined from the time when the transaction occurs.

From sequence data aggregate, type of sequence pattern can be found. According to the generalization from the research of sequence rule done by Chen *et al.*[23], the current proposed sequence rule mining methods capable of sequence pattern mining can be classified into three categories. (1) Similar Pattern: Find similar patterns as to the Euclidean distance and associated coefficients among patterns. (2) Period Patterns: Find patterns having cyclic and periodical characteristics from data aggregate containing time signature. (3) Frequent Pattern: List patterns which occur more than all patterns of smallest support in data aggregate so as to find this type of patterns. Of the above three types of sequence patterns mining, the frequent pattern mining is the main body of current researches[7,10,11,24,25]. Of the frequent sequence patterns which can mine different rules are classified as (1) Continuous Pattern, (2) Discontinuous Pattern and (3) Hybrid Sequence Pattern, proposed by Chen *et al.*[23], which can mine the sequence patterns and non-sequence patterns at the same time. The difference among three different sequence patterns is explained in the following: There are three sequences. Sequence 1: A, B, Y, K, F, Sequence 2: A, B, K, F, P and Sequence 3: C, A, B, C, R, K, F. Let the minimum support be 3, then the algorithm which mines continuous patterns can find <AB> and <KF> sequence patterns but is not able to identify the sequence patterns such as <AB*KF>. * represents a string of item set whose length is bigger than 0 and is not fixed. As for the algorithm which mines discontinuous patterns, it can effectively find those discontinuous sequence patterns such as <A*B*K*F>, <K>, <B*K*F> and <B, F>, yet it cannot effectively mine the continuous patterns such <AB> and <KF>. Therefore Chen *et al.*[23] proposes hybrid sequence pattern which is not only able to mine both continuous and discontinuous patterns at the same time but is also able to mine the hybrid patterns as middle items that have no fixed length such as <ABK*F>, <AB*KF> and <AB*K*F>. Sequence pattern mining is quite widely used in practical applications. An example can be seen from its applications mentioned by Chen *et al.*[23] these applications include mining non-specific user browsing behaviour model[26] the order of priority in customer's purchase and his shopping behaviour at shopping mall, the order of places to visit in a trip, the forecast of possible plan failures, the gene sequence in gene database and what not[23]. We propose a BOM sequence pattern mining method and because each part number in the BOM must be the only characteristic thus we must employ continuous sequence pattern mining. In essence of mining events, the BOM lists the sequence of processing, which is not the aggregate of events that have already taken place, but rather a description of an aggregate of events about production plan that will be taken in the future. Therefore our method wants to solve the data mining problem that describes the meta data of processing.

MATERIALS AND METHODS

Figure 5 shows the process flow and procedures of our proposed method in four steps, from the selection phase on products of high delivery frequency, the mining phase on BOM sequence pattern, the generation phase of sequence rules and the phase of optimized quantity of production. First phase is to conduct a RFM Model analysis on all sales data based on three dimensions: time, frequency and money involved on the purchase in order to find the most valuable product which is also the most potential product that customer would like to buy again in the future. Second phase is to do a data mining on the sequence pattern of BOM then find the child part number which occurs more than large 1 item set of smallest support on the RFM Model selected BOM. Next is one more time of scanning on the BOM, according to the processing sequence, keep the large 1 item set of largest support and eliminate the processes whose lengths are smaller than the minimum sequence support also to obtain the sequence pattern of BOM. Third phase is to generate rules of common semi-finished products by the mined

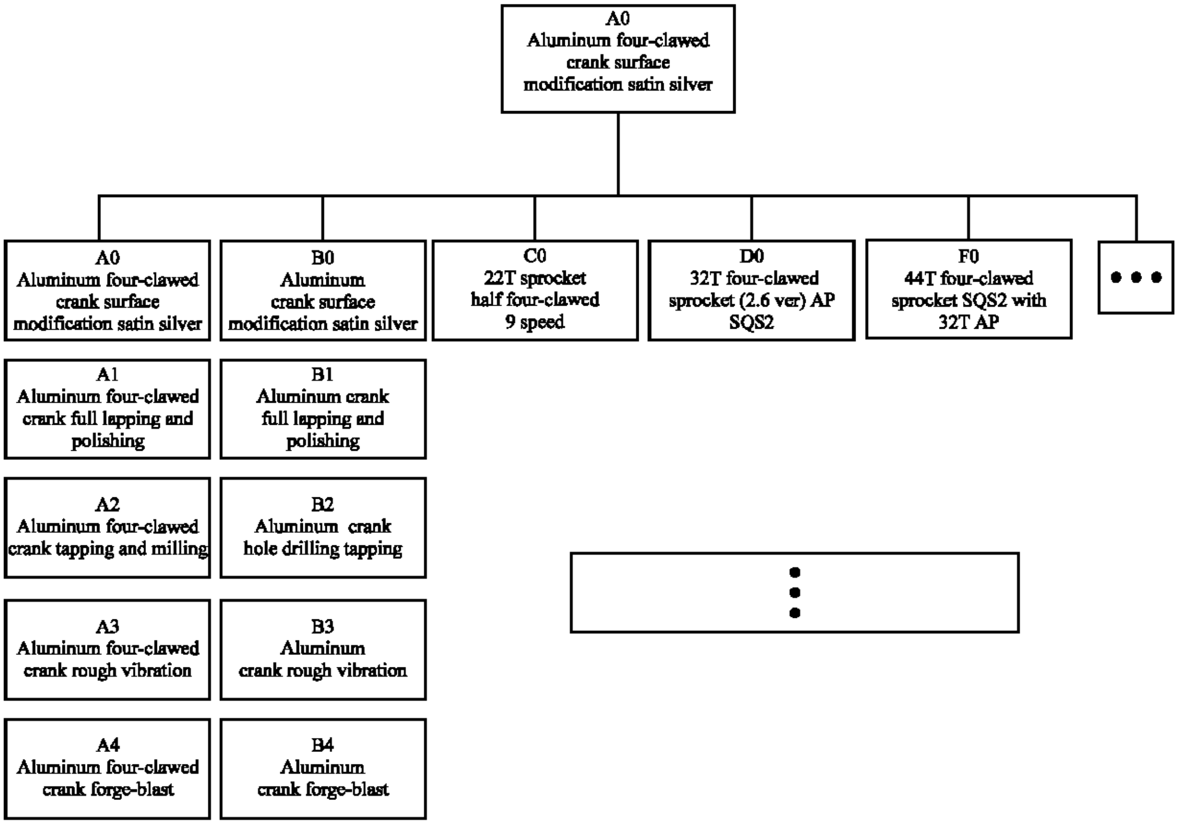

Fig. 4: Example of BOM

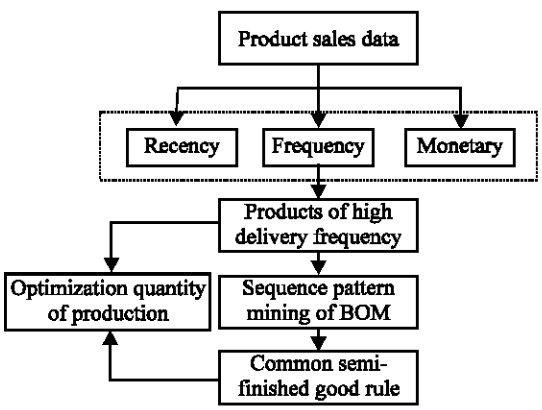

Fig. 5: The process flow and procedures of our proposed method

sequence patterns and present these rules in visualization that demonstrates the dependent relationship among raw materials, common processing, semi-finished goods and final goods and provide as a source of reference for materials purchase, outsourcing, production schedule and when R and D related personnel make materials requirement planning (MRP) in enterprise resource planning (ERP) and when they develop new products. Fourth phase is to use the largest common part rules by the sales cycle and quantity of the common final goods and forecast the probability value of future sales on the final goods to meet the expectation and make a safe and feasible optimized quantity of largest common semi-finished goods production.

Frequent product screen phase: Enterprises during their growth always put their main capability on products which bring them the highest profit and change plans accordingly with time. Despite these changes are due to new strategy, spontaneous declination of the life cycle on the product itself, the new product development because of marketing demand, the improvement over the older product and processing, seasonal effects, or popularity factors, all of these play part in volume of sales. Sooner or later, these changes will reflect on the raw materials for production, the processing and possible semi-finished goods that may utilize the backup material inventory. Therefore our method, during the selection phase on products of high delivery frequency, employs the commonly-used RFM Model in database marketing that dynamically adapts itself to external factors to find the most likely product that customer may continuously

Table 2: Search of recently purchased products aggregate from order data table

```
SELECT DISTINCT B.product_no, B.order_no, B.ord_qty, B.total_price
FROM
Order_Master A, order_detail B
WHERE A.ord_date BETWEEN :start_date AND end_date
ORDER BY A.ord_date
```

Table 3: Search of sales frequency on products aggregate from order data table

```
SELECT DISTINCT B.product_no, SUM(B.order_qty),
COUNT(B.product_no) FROM Order_Master A,Order_Detail B
WHERE A.ord_date BETWEEN :start_date AND end_date
AND A.Order_No = B.Order_No
GROUP BY B.product_no
ORDER BY COUNT(B.product_no) DESC
```

Table 4: Generating a sequence data

Product-id	Item set
001	A4, A3, A2, A1, A0
001	B4, B3, B2, B1, B0
001	C6, C5, C4, C3, C2, C1, C0
001	D7, D6, D5, D4, D3, D2, D1, D1
001	F6, F5, F4, F3, F2, F1, F0

Table 5: Conversion of BOM to sequence data

Product-id	Item set
001	A4, A3, A2, A1, A0
001	B4, B3, B2, B1, B0
001	C6, C5, C4, C3, C2, C1, C0
001	D7, D6, D5, D4, D3, D2, D1, D1
001	F6, F5, F4, F3, F2, F1, F0

Table 6: Calculation of supports on sub part numbers of products after RFM selection

```
SELECT Child, COUNT(Child) FROM BOM
WHERE Parent IN
(SELECT Product_No FROM RFM)
GROUP BY Child
HAVING COUNT(Child) > Min_Sup
```

Table 7: Generation of frequent processing sequence on highly potential of products being purchased

```
Algorithm Generate_Frequent_Sequences

//RFM: The Data Sheet of RFM
//BOM: Bill Of Material
//Min_Sequence : Minimum Sequence
//Temp _Sequence: Temporary Sequence
//Frequence Sequence : Frequence Sequence
//Min_Support:Minimum Support

for each RFM do
begin
   for each BOM level do
   begin
      Search every level child from BOM ;
      If Child < Min_Support then
         Next Sequence
      else Add item_no and support To Temp _Sequence ;
   End ;
   If Temp_ Sequence < Min_Sequence then
      Delete Temp_ Sequence
   Else Add Temp_ Sequence To Frequence_Sequence ;
end ;
```

purchase in the future. In database marketing, the RFM Model exploits recency, frequency and monetary attributes to identify the customers and prevailing products values in the database.

In evaluation of the recency attribute, we apply the SQL scripts in Table 2 on the Master-Detail order data table (or sales data table) in the transactional database, during the time period from start_date to end_date, to find the data containing customers' purchased products. These sales data are arranged in sequence starting off the most recent transaction date and in like manner until the earliest transaction date. Of these products aggregate, the method for determining the R value in each individual product can be easily achieved by the 20/60/20 distributive principle[17] which assigns each product a different R value; in this case, the most preceding products in the sequence take up 20% because of the more recent transaction dates, hence the R value is 3; the middle 60% products have a R value of 2; the rest of 20% products which may be unpopular or become less competitive that lack of people's interests thus have a R value of 1.

In the frequency attribute, we apply the SQL scripts in Table 3 to generate a score of frequency for each purchased product. The SQL scripts in Table 3 generate a sequence containing the frequency of each product that has been purchased during the time period from start_date to end_date, the sequence is determined starting off the higher frequency to lower frequency of the product purchased. Similarly by the 20/60/20 distributive principle, assign a F value to a sales frequency for individual product; the most preceding 20% products are given a F value of 3, the middle 60% products are assigned a F value of 2 and the last 20% products are assigned a F value of 1.

Last attribute in RFM Model which is monetary, we apply the SQL scripts in Table 4 to generate some values. The SQL scripts in Table 4 generate a sequence containing the total amount of money spent on each product during the time period from start_date to end_date, the sequence is determined starting off the larger total amount of money spent on individual product to lower. Likewise the scoring method for assigning M value is by the 20/60/20 distribute principle, which assigns the most preceding 20% products a M value of 3, the middle 60% products a M value of 2 and the last 20% products a M value of 1.

Through the simple 20/60/20 principle in the above, the three attributes in RFM Model are assigned values separately, because each attribute has three scoring variations (1, 2, 3) and there will be 3^3=27 variations in

Table 8: Calculation on optimized quantity of production

Part numbers of final goods	Order cycle /Days (Tn)	Probability	Average quantity of delivery (Qn)	Optimized quantity of production
122006460015201	20	0.60	829	497
121007330015202	29	0.41	788	323
121001440015203	30	0.40	741	296
121004440015202	30	0.40	432	173
122008460015201	30	0.40	83	33
123006480015201	37	0.32	739	236
123018480015201	58	0.21	418	88
121001440015201	87	0.14	100	14
123017480015201	90	0.13	780	101
121007360015201	116	0.10	525	53
123029420015201	120	0.10	100	10

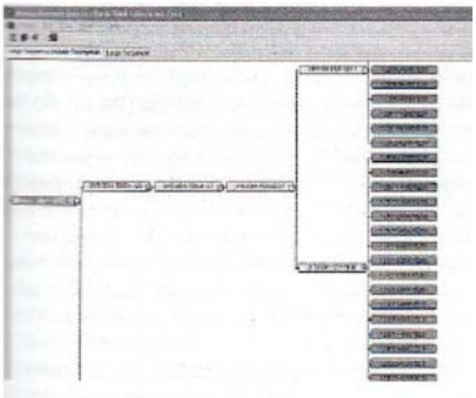

Fig. 6: Generation of frequent sequence rules of highly potential products

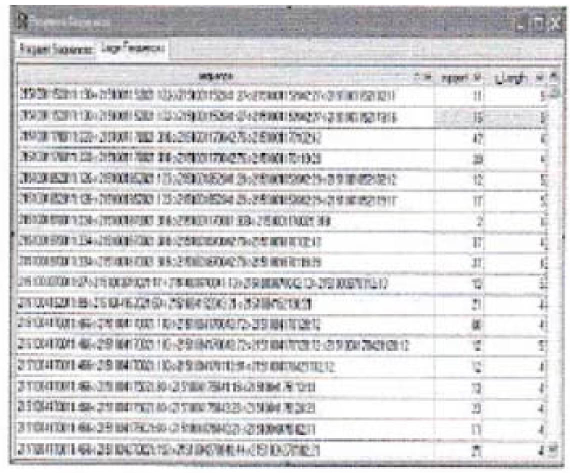

Fig. 7: Frequent sequence tree of highly potential products

total for the three attributes. Enterprises can pick a proper RFM score to be the threshold value depending on different industrial conditions and product characteristics in order for selection of a most probable product on which customers will continuously place order. For example, for products with a total score of RFM that is bigger than 6, store them in another RFM data table for BOM sequence pattern mining at the next step; these are the potential, highly-valued products for further analysis.

BOM sequence patterns mining phase: The BOM in Fig. 4 can be converted to a string of sequence on the value added production process as shown in Table 5. The Product Id records the number assigned to each final good in the BOM; and the Item Set records the branching paths of product structure tree, starting from the lowest level of parts purchase through the final and semi-finished goods part numbers by the sequence, in which each branch in the structure tree is regarded as a complete string of sequence. Those stickers, hexagonal screws, packaging, or paperboard boxes of G0, J0, J2 and K2 from Fig. 4 belong to low profit Type C materials in terms of ABC Classification, usually are monitored and controlled by some easier methods such as Visual Review System or Two-Bin System in planning[15], handout distributing, material receiving, checking and cost calculation. ABC Classification depends on the proportion of cost of raw materials over annual cost of purchase and classifies into A, B and C three categories. Type A items approximately take up 10-20% of total raw materials and 50-80% of total cost of annual purchase, Type B items 20-30% of total raw materials and 15-20% of total cost of annual purchase and Type C items 50-70% of total raw materials and 5-10% of total cost of annual purchase[5]. Raw materials for Type C that have more variety yet cost less can be monitored and controlled appropriately and easily, although Type C items are purchased in large quantity which lead to higher inventory cost but do not account for higher percentage in total cost hence more room for price negotiation and unnecessary management cost can be avoided. Therefore our method ignores the purchased materials that are either single phased or insufficiently phased that these materials will not be used for sequence calculation of common products at the succeeding phase. This way, we can prevent from the mined sequence rules of BOM being too short while these rules may be publicly known or some trivial concept rules. Take the packaging of each specification as an example, the packaging paper box must be the largest common part of all products because every product must be packaged before delivery so that these sequence rules which are too short apparently have no reference value or meaning with respect to improving the production flexibility.

There are many past sequence pattern mining algorithms whether they are Generating Candidate-and-Test or Pattern-Growth methods, they must consider and test before composing the items in frequent patterns if the variation in items combinations results in a new frequent pattern[27]. What our method is going to deal with is a string of sequence established from the conversion of tree structure in BOM, differing from the past sequence pattern mining algorithms mainly in two parts: One is the characteristic of enforced sequence, the other is characteristic of unique path. The two characteristics are explained in the following:

[Characteristic of enforced sequence]: The string of sequence $S=<s_1,s_2, \ldots, s_n>$ moves along the time line by the production process. Raw materials are constantly added values in which their status and specification are both different from previously. To clearly identify the difference in these characteristics, each different material therefore in ERP and MRP information systems requires a unique part number to represent every product of different characteristic so as to ensure clear discernment of the difference between part numbers. Therefore in the string of sequence {A4, A3, A2, A1, A0} from Table 5, the sequence of appearance of each item has interrelated, positional relationship in which enforcement $(s_1<s_2<,\ldots<s_n)$ exists, because in real production there are limitations of feasibility related to physics, technology, cost and so on, therefore the sequence of processing must maintain and guarantee a certain order of priority. One of the purposes in BOM is also to ensure the same products being produced at different time points have same specification and quality. For example, if there exists the processing sequence {A4, A3, A2, A1, A0} in Table 5, there will exist no sequence {A3, A4, A2, A1, A0} because A3 is the resulting part number after rough processing on A4 so that the sequence A3->A4 cannot exist in the converted sequence data. Thus, from [characteristic of enforced sequence] we know the string of sequence formed by the order of prior processing will only yield one unique sequence of appearance of the composed items that no other sequence of items exists. It's therefore not necessary to take account of all large items and if other possible permutational combinations are frequent sequences during sequence pattern mining of BOM. For example in Table 5, the production sequence A4, A3, A2, A1, A0 is the only sequence thus it's not necessary to consider the sequence such as A3, A4, A2, A0, A1 and if the other 2^5-1 sequences of combination variation are frequent sequences as well as other permutational combinations of the items being frequent patterns or not during the sequence pattern mining of BOM.

[Characteristic of unique path]: If the string of sequences $S1=<s_1,s_2, \ldots, s_n>(s_1<s_2<,\ldots<s_n)$ and $S2=<s_a,s_b, \ldots, s_m>(s_a<s_b<,\ldots<s_m)$ which are two frequent sequences, then S1 will not be equal to S2 and the only path of completing s_n must be s_a->s_b-> …-> s_m. For example in Table 5, the string of sequence {A4, A3, A2, A1, A0} and {B4, B3, B2, B1, B0}, if the semi-finished good A0 is to be made, it will have to go the path from A4->A3->A2->A1 where the path from B4->B3->B2->B1->A0 will not work here because the size of B1 in terms of its physical property has already been determined back when the forging raw materials of B4 is purchased thus B1 will not be transformed into A0 no matter whatever processing is applied to it. According to [characteristic of unique path], it is deducted that if $S=<s_1,s_2, \ldots, s_n>$ is the frequent sequence that has its support bigger than the minimum support, then $s_1,s_2, \ldots, s_n$ of all processing sequences must be the large 1 item set whose support is bigger than the minimum support; if s_4 is the large 1 item set, then the s_1, s_2 and s_3 prior to s_4 must also be the large 1 item sets and s_1->s_2-> …->s_4 of all processing sequences must be the frequent sequence due to that the path to s_4 is only one that s_4 must go through the s_3 processing and s_3 is completed from the s_1->s_2 goods.

Also from the research by Agrawal *et al.*[21,22], any one string of sequence whose frequency of occurrence is high, then each of its composed item in the sequence must have its number of occurrences bigger than the minimum support. Consequently in the step of sequence pattern mining of BOM, the SQL scripts in Table 6 are used to find the part number in which its number of occurrences of child part number is bigger than the minimum support.

As well from the [characteristic of enforced sequence], it's known that the part numbers which occur in the processing sequence bear the characteristic of enforced sequence so it's not necessary to consider the combinations variation between the part numbers. And from [characteristic of unique path], the processing sequence has this characteristic that path is unique so the processing sequences in BOM exist individually without affecting each other. Accordingly in our modification of Apriori Like sequence rule mining algorithm, we need to continually generate candidate aggregates and verify the candidate aggregates that lead to combination variation are the frequent patterns then propose Generate-Frequent-Sequences algorithm in Table 7. Along the tree structure in mother part number of BOM, we move from the direction of low-level to high-level to find all child part numbers. If the low-level part numbers of BOM are smaller than the given Min_Support, we will neglect this processing sequence and move on to the next processing sequence. If the low-level part numbers of BOM are bigger than the given Min_Support, we will add it to our Temp_Sequence. If this processing sequence has searched up to the 0 phase and that the counts of the part numbers that compose processing sequence are smaller than the given minimum sequence, we will ignore this sequence; if the counts of the part numbers that compose processing sequence are bigger than the given minimum sequence, we will add it to our frequent sequence aggregates.

Generate sequence rules phase: Fig. 6 is the resulting frequent sequence rule generated from the algorithm in Table 7. For example, for sequence rules 2151001152011:130-> 2151001152021:122-> 2151001152041:27->2151001152042:27->2151001152102:11, the sequence rules are shown from left to right by the order of processing and the value compartmentalized by colon in each part number represents the number of times that this part number of all products has been commonly used. The length of this sequence rule is 5 meaning 5 processes and the value to the very right of colon in the sequence rule means from the most potential products that customers will continually purchase there are 11 products that will use this common production process. The sample data we used comes from a bike cranks manufacturer; in its ERP system, the parts numbering method employs meaning rule which allows our method to generate the largest common part sequence rules that are easier to comprehend and apply even though the length of part numbers are somewhat longer and not good for rules presentation so we choose to keep its manipulation. In fact from Fig. 6, we know that the first four part numbers are the same so are their supports in the first two sequence rules. The difference takes place after the fourth process that there are two processes to choose from: one generates the part number 2151001152102 and the other 2151001152119 and there are 11 and 16 types of products which use these two processes, respectively. As a matter of fact, the common process from 2151001152011 to 2151001152042 is used by those two sequence rules, that includes the length of this process is 4 so the sum of supports in these two sequence rules are 27. Thus using sequence rule alone to provide as a reference to the production management personnel doesn't seem instinctually easy to comprehend. For this reason our method additionally uses from Fig. 7 a more instinctual and easier for comprehension tree structure and present the common sequence rules in visualization. The oval block at the rightmost in Fig. 7 shows the part numbers of this process, the rectangular block shows the frequent sequences. The production management personnel can easily understand that 2151001152102 and 215100115211916 are commonly used by the 11th and 16th products, respectively. During the extension of production materials, rest assured when magnifying the production multiple for these two part numbers, or the production management personnel can choose to magnify 215100115204227 which is the common production multiple used by these two sequence rules for proper inventory. It's this way that increases the ability of raw materials purchase and price negotiation for outsourcing, stabilizes the product quality during the processing and even minimizes the various factors that may occur as a result of multiple production of same processing in less quantity.

Optimize produce multiple of semi-finish-good: From the largest common processing sequence rule obtained in the above procedure, we know how many highly potential products are available and the common process used during the production. Although the commonness of semi-finished goods of all products of high frequency in customers' purchases is high (bigger than minimum support), these semi-finished goods can be produced for inventory in a very safe sense without overstocking which may result in money issues. However, using the largest common processing sequence rule alone is still insufficient to accurately determine how much semi-finished goods are to be produced. In spite of their high commonness of semi-finished goods, more than needed stocking will be a burden and hurt the enterprises' capital in view of cost of materials, processing and even inventory management. Consequently for optimized production multiple, our method proposes a reliable and safe method to calculate the optimized quantity of production for these semi-finished goods and make a balance between production flexibility and inventory cost.

The total available final goods (n) which utilize these common semi-finished goods processing (Fig. 6 and 7) can be known from the largest common processing sequence rule. Therefore from the sales record at the first phase, we can calculate the average cycle of sales on the final goods separately and set these final goods as $T_1, T_2, \ldots, T_n$ and their average quantity of sales as $Q_1, Q_2, \ldots, Q_n$. Next is to find the part number from the final goods in the longest order cycle, set the number of days as standard cycle T_{base} and assign to it a α value which represents the probability value of delivery of this final good during the expected length of time (T_{base}). Assume the order cycle of the n-1th final good is T_{n-1} and the average quantity of delivery is Q_{n-1}, knowing that products of higher delivery frequency have higher next-delivery frequency, we can calculate the probability value of delivery on the n-1th product by Formula 1 during the average cycle of delivery, thus the multiple of probability value on the n-1th product can be pre-determined. Multiply the result from Formula 1 by Q_n, we get the count of production inventories on the n-1th product (Formula 2). Adding up all counts of productions from 1st to n-1th products by Formula 2, we get the optimized counts of production inventories for this common processing sequence rule (Formula 3).

$T_{base}/T_{n-1} * \alpha$ .... Formula 1
$(T_{base}/T_{n-1} * \alpha)*Q_{n-1}$ .... Formula 2
$\sum(T_{base}/T_{n-1} * \alpha)*Q_{n-1}$ .... Formula 3

Table 8 is shows the trial calculation of optimized quantity of production on the sequence rules 2151001152011:130-> 2151001152021:122-> 2151001152041:27->2151001152042:27->2151001152102:11 from Fig. 6. The part numbers of final goods in the column shows 11 part numbers that use this common process; and the order cycle (T_n) of each product and the average quantity of delivery (Q_n) are recorded in the second and fourth columns of Table 8. Let the probability of on-time delivery on the product 123029420015201 of longest order cycle be 0.10(α) within 120 days (T_{base}) and multiply the probability which is 0.1 by the historically average quantity of delivery which is 100 we get 10. 10 is the optimized quantity of production in which this product is expected to contribute to the sequence rule. The part number of final good is 122006460015201, the order cycle is 20 days and the average quantity of delivery is 829. Now for all products which use this common process, take the longest number of days of order cycle which is 120, divide 120 by the order cycle which is 20 and multiply by its probability which is 0.1, we get the probability of delivery of 0.6 for 122006460015201 with respect to 123029420015201. Multiply 0.6 by the average quantity of delivery for 122006460015201 which is 829, we get an optimized quantity of production which is 497 in which this product is expected to contribute to the sequence rule. We can use the above method to calculate the optimized quantity of production from all products that use this common sequence rule. Then we add the optimized quantity of production from all 11 products together, we get an optimized quantity of production for 2151001152011:130? 2151001152021:122? 2151001152041:27?2151001152042:27?2151001152102:11 by this sequence rule with which we end up is 1824. When production management personnel in the ERP system plan on production schedule for the orders, if the semi-finished good 2151001152102 is needed, they can properly magnify the production multiple on 2151001152102 in terms of this value of optimized production in order to enhance their ability on price negotiation with the suppliers, maintain a stable and desirable production quality and better respond to uncertain delivery demand from customers.

Our proposed method takes advantage of RFM analysis for initial selection and filtering on potential products which customers would continually like to purchase. Those common processing sequence rules generated by our method not only dynamically but also instantly respond quite well to current market demands and customers who purchase the products. We present these common processing sequence rules in visualization and allow newly product management personnel or assisting consultants to quickly extract knowledge or information about process flow of the processing that is used and commonness of materials. It's also in this way that we lower the threshold for new comers so they can quickly get a hang of the production schedule in the enterprise, for example; additionally, we minimize and prevent incidents from happening such as screwed-up production schedule and unpunctual delivery due to migration of workers. Another advantage of using the common processing sequence rule is it serves to developers as a guide when they are developing new products that they can enhance the commonness and modularization of products. For general managers, a visualized common processing sequence rule provides a more instinctual as well as farseeing insight to check for commonness in all product lines, modularization, possible bottlenecks in capability as well as resources conflict in order to make sure everything works efficiently and flexibly in resource distribution strategy, manpower demand, overtime, equipment, new outsourcing and what not.

There are two possible limitations as follows: (1) For industries with their production processing being too short or their commonness of materials being low, despite using our method, they will not be able to get useful common processing sequence rules for their needs. (2) Whether our generated largest common processing sequence rule can completely shorten the lead time of delivery is up to the type of industry and property of product; if the mined common processing sequence rule is used for production on just some small parts, it's still need to wait until the processing of other products of lower commonness is finished before the assembly and delivery. Even though shortening the total amount of time by our method before the delivery is up to the property of product, our method basically can prevent resource conflicts from happening, decrease the percentage of defective production, increase the fluency of production schedule and enhance the flexibility to respond to customers no matter what category of product we are dealing with.

In our future work to improve our method, we will extend our research then put our work on the time point of executing the algorithm, modification of the optimized quantity of production and automation of feedback mechanism. Of ERP or MRP systems, after modeling Mater Production Schedule (MPS) and Rough-Cut Capacity Planning (RCCP), proceed to extension of Material Requirement Planning (MRP). If our method can combine with MPS and analyze first the largest common processing materials before the extension of MRP then send the optimized quantity of production as feedback to the ERP system to even reduce the time and cost for which the production management personnel have to modify the optimized quantity of production themselves according to the common processing rule would be. In addition, if we are considering automation, we have to integrate into the ERP system error-free and the efficiency part of full extension of BOM and the time point of executing the algorithm may be specially attended to agree the instant efficiency requirement of automation.

REFERENCES

1. Burgess, T.F., H.K. Gules, J.N.D. Gupta and M. Tekin, 1998. Competitive priorities, process innovations and time-based competition in the manufacturing sectors of industrializing economies: The case of Turkey. Benchmarking, for Quality Manage. Technol., 5: 304-316.
2. Samadhi, T.M.A. and K. Hoang, 1995. Shared computer-integrated manufacturing for various types of production environment. Int. J. Oper. and Product. Manag., 15: 95-108.
3. Donk, D.P., 2001. Make to stock or make to order: the decoupling point in the food processing industries. Int. J. Product. Econom., 69: 297-306.
4. Olhager, J., 2003. Strategic positioning of the order penetration point. Int. J. Product. Econom., 85: 319-329.
5. Partovi, F.Y. and M. Anandarajan, 2002. Classifying inventory using an artificial neural network approach. Computers and Industrial Engineering, 41: 389-404.
6. Vlist, P.V., J.E.M. Hoppenbrouwers and H.M.H. Hegge, 1997. Extending the enterprise through multi-level supply control. Int. J. Product. Econom., 53: 35-42.
7. Gfrere, H. and G. Zapel, 1995. Hierarchical model for production planning in the case of uncertain demand. European J. Oper. Res., 86: 142-161.
8. Tang, O. and R.W. Grubbstrom, 2002. Planning and replanning the master production schedule under demand uncertainty. Int. J. Product. Econom., 78: 323-334.
9. Zapel, G., 1996. Production Planning in the case of uncertain individual demand extension for an MRP II concept. Int. J. Product. Econom., 46-47: 153-164.
10. Zapel, G., 1998. Customer-order-driven production: an economical concept for responding to demand uncertainty. Int. J. Product. Econom., 56-57: 699-709.
11. Huiskonen, J., P. Niemi and T. Pirttila, 2003. An approach to link customer characteristics to inventory decision making. Int. J. Product. Econom., pp: 255-264.
12. Bertrand, J.W.M., M. Zuijderwijk and H.M.H. Hegge, 2000. Using hierarchical pseudo bills of material for customer order acceptance and optimal material replenishment in assemble to order manufacturing of non-modular products. Int. J. Econom., 66: 171-184.
13. Olsen, K.A., P. Saetre and A. Thorstenson, 1997. A procedure-oriented generic bill of materials. Computers Industry Engineer, 32: 29-45.
14. Yeh, C.H., 2000. A customer-focused planning approach to make-to-order production. Industrial Management and Data Systems, 100: 180-187.
15. Lee, J.K. and P.R. Larry, 1999. Operations Management: Strategy and Analysis. 5th Edn.
16. Carol, P.A. and E. Schragenhieim, 2000. ERP: tools, techniques and applications for integrating the supply chain. CRC Press LLC.
17. Miglautsch, J.R., 2001. Thought on RFM scoring. International Society for Strategic Marketing, 27: 1-7.
18. Colombo, R. and W. Jiang, 1999. A stochastic RFM model. J. Interactive Market., 13: 2-12.
19. Viaene, S., B. Baesens, T. Gestel Van, J.A.K. Suykens, D.V.D. Poel, J. Vanthienen, B.D. Moor and G. Dedene, 2001. Knowledge discovery in a direct marketing case using least squares support vector machines. Int. J. Intelligent Systems, 16: 1023-1036.

20. Chen, M.S. and J.H.P.S. Yu, 1996. Data mining: an overview from a database perspective. IEEE Transactions on Knowledge and Data Engineering, 8: 866-883.
21. Agrawal, R., T. Imieliki and A. Swami, 1993. Mining association rules between sets of items in large databases. Proceedings of the 1993 International Conference on Data Management of ACM SIGMOD., 22: 207-216.
22. Agrawal, R. and R. Srikant, 1995. Mining sequential patterns. Proceedings of the Eleventh International Conference on Data Engineering, pp: 3-14.
23. Chen, Y.L., S.S. Chen and P.Y. Hsu, 2002. Mining hybrid sequence patterns and sequence rules. Information Systems, 27: 345-362.
24. Guoli, J., G. Daxin and F. Tsui, 2003. Analysis and implementation of BOM of tree-type structure in MRPII. J. Materials Process. Technol., 139: 535-538.
25. Stonebraker, P.W., 1996. Restructing the bill of material for productivity: A strategic evaluation of product configuration. Int. J. Product. Econom., 45: 251-260.
26. Pei, J., J. Han, M.A. Behzad, P. Helen, Q. Chen and M.C. Hsu, 2001. Prefix span: mining sequential patterns efficiently by prefix-projected pattern growth. Proceedings of 17th International Conference on Data Engineering, pp: 215-224.
27. Pei, J., J. Han, H. Lu, S. Nishio, S. Tang and D. Yang, 2001. H-mine: hyper-structure mining of frequent patterns in large databases. Proceedings IEEE International Conference on Data Mining (ICDM 2001), pp: 441-448.