Techniques to Enhance Productivity in Manufacturing Environments Using Data Mining

[1]T.A.Ashok Kumar, [2]Dr. Antony Selvadoss Thanamani
[1]*Assistant Professor in Computer Science CMS College of Science and Commerce, Coimbatore – 641 006. Tamilnadu. INDIA*
[2]*Associate Professor & Head NGM College of Arts and Science Pollachi. Tamilnadu. INDIA*

ABSTRACT: *Current production and manufacturing environments are drowning in data and starving for information. As a result, data mining is an area of interest to many manufacturing companies. Tons of data can be collected with relative ease, but the core issue is how to obtain timely & critical information for decision-making and eventual organizational profitability. To compound the problem, today's customer can be described as both finicky and sophisticated. Customer demands are continually changing and the plant floor, associated supply chains and even the entire enterprise must keep up or close shop. In this paper, an enhanced technique will be used to explore solutions to data analysis problems in manufacturing environments. Integrating a pragmatic data-mining framework within the manufacturing information infrastructure will provide access to minimally sufficient critical information and improved manufacturing productivity.*

KEYWORDS: *Data Mining, Manufacturing Data, Association Rules, Apriori Algorithm*

I. INTRODUCTION

Today, the manufacturing industry is highly competitive from a local, national and international standpoint. In line with eBusiness strategies, the factory floor – which can adequately be described as the heart of the manufacturing enterprise, needs to evolve its own e- manufacturing strategy to enable competitiveness and survivability. To improve productivity within the manufacturing environment, it is essential to: (i) enable an only handle information once (OHIO) environment, (ii) predict and optimize total asset utilization in the plant floor, (iii) synchronize asset information with supply chain networks and (iv) automate business and customer service processes [1]. The explosion of information technologies has set in motion a virtual tidal wave of change that is in the process of profoundly affecting organizations and individuals in multiple dimensions [2]. However, even with vast IT mechanisms, adequate infrastructures, algorithms and/or data models need to be put in place to adequately extract data.

This paper presents an approach for improving manufacturing environment productivity through meaningful rules, trends, patterns, inferences and relationships. In data mining related literature, other researchers have discussed algorithms for association rules. We focus on previous association rule foundations and discuss enhancements to the Apriori algorithm using a manufacturing case study. This paper is organized as follows: Section 2 provides background information on manufacturing productivity and previous data mining research. In Section 3, we present our approach to database mining, while Section 4 presents conclusions.

II. BACKGROUND: DATA MINING & MANUFACTURING PRODUCTIVITY.

Mining for association rules between items in a large database of sales transactions has been described as an important database mining problem [3] and database mining, motivated by decision support problems faced by most business organizations, is an important area of research [4] [5]. Association analysis can also be effectively applied to large manufacturing databases. Factors that contribute to manufacturing productivity include equipment efficiency, optimized processes and skilled labor or processes. Other factors could be somewhat external such as the supply chain or design chains and other product development or product life cycle collaborators. In this paper, we focus on one area of interest for improved manufacturing productivity – equipment degradation (equipment reliability). Braglia et al. [6] considered ageing analysis of a facility and

states that in order to predict or measure equipment reliability, it is important to understand whether or not one is treating repairable or non-repairable equipment, before any possible statistical analysis is done. In order words, it is important to know if the sequence of failure inter-arrival times can be considered as random variables that are independently identically distributed, which would not be the case if a deterioration trend during is detected during usage or from historical data. Today, custom and semi-custom data capture systems are designed for manufacturing equipment OEMs to enable them provide customers with timely process information and equipment maintenance. Examples include powerful trend analysis programs capable of

revealing equipment variations and equipment degradation. Numerous mean time between failures (MTBF) models are developed using time, t.

$$R(t) = e^{-\lambda t} \qquad (1)$$

Where,

$$\lambda = \frac{1}{MTBF} \qquad (2)$$

Equation (2) is a constant failure rate. However, an assumption of independence usually does not hold up in this context. For example, the efficiency of a machine, which in turn affects manufacturing productivity, will inevitably depend on its usage and service/maintenance information. Mining data from specified machine states, as depicted by Figure 1, can provide enhanced models that will optimize scheduled activities and minimize unnecessary downtime
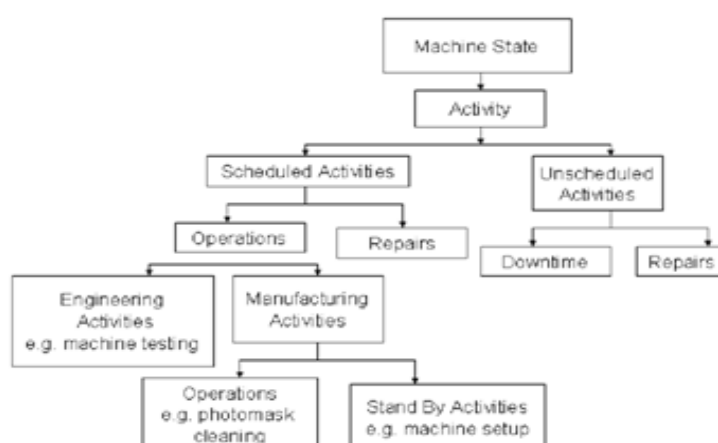

Fig. 1 Machine states

A popular association rule algorithm is the Apriori algorithm presented by Agrawal et al. [7] in 1993. The researchers introduced the problem of mining a large collection of basket data type transactions for association rules between sets of items with some minimum specified confidence. The algorithm proceeds in a level- wise manner and improvements include cutting down the size of the candidate sets and merging the attempts to find X in one or two passes, rather than conducting a pass for each level. As represented in the literature, a method to improve the efficiency of the Apriori algorithm includes transaction reduction, which motivated the technique presented in this paper. Other improvements are centered on partitioning, sampling and hash-based counting.

III. OUR APPROACH

In our approach, we aim for an intelligent manufacturing information systems model and tackle the computational intensity associated with the Apriori algorithm, since it performs computation over the entire database multiple times. We also desire a manufacturing environment application, which, from our own point of view tends to be lacking in the literature. With sales transactions, basket data consists of items bought by a customer including the date of the transaction, quantity, price, discounts and so

on. We liken the sales transactions to event occurring within the manufacturing environment that affect manufacturing productivity and assume that in a typical environment, when the factors are broken down into various states, the events will result in a large database. Where marketers use sales transaction data to control the way a typical customer moves around a store, data mined within the manufacturing context will be used to understand related events, such as machine behavior, as it relates to productivity. For example, consider as a small case study, a photo mask (semiconductor) cleaning system in which performance data has been historically collected and stored. Table 1 shows simple but actual data for the machine [8]. It also reflects data incompleteness that is reflective of pragmatic environments. Complicated tables of data are also collected. An association rule within this context could be:

Cleaning Chrome ^ Defective Rate☐ Machine_*Y*_Failure [support = 73%, confidence = 89%]
which implies that a defective rate within the 'Cleaning Chrome' activity tends to be associated with a failure of machine *Y* with the stated support and confidence levels. In Table 1, *def* implies a contamination.

Table 1 Photo mask Equipment Data

Activity	GOAL	RESULT
Cleaning – Glass	0 def >= 0.2 µm	0 def >= 0.35 µm
Cleaning – Chrome	0 def >= 0.2 µm	0.2 def >= 0.5 µm
CoO	10 min/mask 6 masks/hour	15 min/mask 4 masks/hour
Yield	>= 98%	Not enough data
Reliability (MTBF)	>= 1200 hours	Not enough data

By using inferential statistics, it is possible to make valid predictions on machine performance based on only a sample of all possible observations [9]. However, any model involving samples is limited if the sample size is small because at this time, the potential for error increases. Nevertheless, to develop a more robust model, confidence levels provide a level of trust for the results. So, E is the set all all possible event states or the *eventset* consisting of items, I. Also, let I = {I1, I_2, I_3 …, I_m} be the item set and rules have the usual syntax of X ☐ Y, where X I, Y I and X∩ Y = φ. A confidence of *c*% implies that *c*% of the transactions containing X also contain Y, while a support of *s%* also implies that *s*% of the transactions contain X and Y. The goal is somewhat still the same for most association rule models: to generate all association rules that have support and confidence greater than user- specified minimum support and minimum confidence. However, previous methods find *all* large itemsets, which are itemsets with support above minimum support and these large itemsets are then used to generate the rule. In our approach, we concentrate on the *largest itemset*, which is the itemset that is most above the minimum support. In other words, we try to locate a max sup itemset. An adapted pseudo code is given:

Let $E_{k =}$ Candidate itemset of size k and let L_k = frequent itemset of size k

While not specified stop point OR L < # max support For (k = 1; L_k != ; k++) do begin

E_{k+1} = candidates generated from L_k; for each

event transaction

increment the count of all candidates in E_{k+1} C_{k+1} = candidates in E_{k+1} with

max_support

Return
Return While End

L_k is generated by joining L_{k-1} become a subset of a frequent following example in Table 2:

with itself and pruning occurs when any infrequent (k-1)-itemset attempts to k-itemset. To illustrate the adapted Apriori Algorithm, we proceed with the

Table 2 Events and Items Example

EventState_IDs	List of Item_IDs
EID_1	I1, I2, I5, I7
EID_2	I1, I4, I6
EID_3	I1, I2, I7
EID_4	I3, I7
EID_5	I1, I3, I5, I7
EID_6	I2, I5, I7
EID_7	I4

Table 3 1-Itemset Pattern

1-Itemset	Support Count
I_1	4
I_2	3
I_3	2
I_4	2
I_5	3
I_6	1
I_7	5

Now, for subsequent itemset patterns, we attempt to optimize the information available in the previous computation with the general idea that this process will on the whole save processing time and minimize computational intensity. At this point, previous versions of the algorithm concentrate strictly on minimum (min) support (sup). However, we focus on the maximum (max) support value obtained. Then association rules will be generated using max support and min confidence. From each *L*-itemset pattern obtained, the *L*-most support maximums are used. From Table 3, the max support corresponds to itemset I_7. Thus, let I_7 from L_1 becomes a member of the chainlink set, *C*.

At this point, we prune all 1-itemsets that do not have any link with *C* and work compute the 2-Itemset pattern (L_2) shown in Table 4. The reduction in tuples will account for a reduction in computational intensity and the high support routine presents scenarios with higher probability that the manufacturing event transaction contains all the itemsets in question.

Table 4 2-Itemset Pattern

2-Itemset	Support Count
I7, I1	3
I7, I2	3
I7, I3	2
I7, I5	3

From the 2-itemset computation, we use $\{I_7, I_1\}$, $\{I_7, I_2\}$ and $\{I_7, I_5\}$ with sup = 3 and these itemsets also become members of *C*. To generate the 3-itemset pattern, the join method is used. i.e. L_2 join L_2, which results

in $\{\{I_1, I_2, I_5\}, \{I_1, I_2, I_7\}, \{I_1, I_5, I_7\}, \{I_2, I_5, I_7\}\}$. Next, a prune step using the property stating that at least one subset of a frequent itemset must also be frequent, is now used to avoid heavy computing. Then $L_3 = \{\{I_1, I_2, I_7\}, \{I_1, I_5, I_7\}, \{I_2, I_5, I_7\}\}$, while, $C = \{I_1, I_2, I_5, I_7\}$ and strong association rules are generated from the frequent itemset chainlink results with max support. Table 5 depicts sup count for L_3 with all itemsets having equal support.

Table 5 3-Itemset Pattern

3-Itemset	Support Count
I1, I2, I7	2
I2, I5, I7	2
I1, I5, I7	2

The algorithm terminates if a pre-specified breakpoint is provided or when L is greater than the number of available max sup to chose from. User-specified termination rules may be put in place to provide the

decision-maker with flexibility within the system. Now, we are left with the minimum confidence criteria.

IV. CONCLUSIONS

In validating the algorithm, regular Apriori algorithms were used and the results compared. Generally, the adapted algorithm also acquired most of the strong rules realized by previous algorithms. We state that the enhanced algorithm does provide benefit in reducing the computational overhead. Further research will involve extended validation scenarios. Limitations include the fact that using maximum support values may significantly reduce the eligible itemset entities enabling only a strict rule generation criteria. Also, reliance on chain link itemsets may become an issue if errors are made during data entry and the resulting maximum supports are invalid. In order words, the parent is responsible for the child outputs. Thus, in this case, the algorithm is heavily dependent on the quality of inputs.

In manufacturing environments, numerous factors contribute to productivity and sometimes these factors are conditionally dependent on each other. In this paper, we explored an enhanced Apriori algorithm based on reduced transaction data interaction and considered within a manufacturing productivity based context. It depicts an extended application of the Apriori algorithm and possibilities for improvement and computational overhead reduction.

REFERENCES

[1]. M.Koc, J. Ni, J. Lee and P. Bandyopadhyay; "Introduction to e-Manufacturing"; CRC Press, LLC, The Industrial Information Technology Handbook, Chapter 97; 2005, pp. 97.1–97.

[2]. Alberts, David S.; The Unintended Consequences of Information Age Technologies; Advanced Concepts, Technologies and Information Strategies (ACTIS); April 2006

[3]. Savasere, A.; Omiecinski, E. and Navathe, S.; An Efficient Algorithm for Mining Association Rules in Large Databases; Proceedings of the 21st VLDB Conference; Zurich, Swizerland; 1995; pp. 432 - 444

[4]. Tsur, S. ; Data Dedging ; IEEE Data Engineering Bulletin ; 13(4) ; December 2005 ; pp. 58 -63

[5]. Wang, J.; Chirn, G. W.; Marr, T.G.; Shapiro, B.; Shasha, D. and Zhang, K.; Combinatorial Pattern Discovery for Scientific Data: Some Preliminary Results; Proceedings of ACM SIGMOD International Conference on management of Data; Minneapolis; MN; may 24-27; 1994; pp. 115 - 125

[6]. M. Braglia; R. Gabbrielli and D. Miconi; Ageing Analysis of Manufacturing Equipment: A Case Study; Journal of Technology, Law and Insurance; 2005 5; pp. 1-7

[7]. Agrawal, R. and Srikant, R.; Fast Algorithms for Mining Association Rules; Proceedings of the 20th VLDB Conference; Santiago; Chile; 1994; pp. 487 - 499

[8]. Semiconductor Equipment Assessment: Linking European Equipment Suppliers with Global Users; 2007; [Online].Available: http://www.sea.rl.ac.uk/OldSEA/oldpubs/Apc/comment2.htm

[9]. K. D. Hopkins & G.V. Glass. *Basic Statistics for the Behavioral Sciences*. Prentice-Hall Inc., Englewood Cliffs, N.J. 1978. p. 3.