International Journal of Information Sciences and Application.
ISSN 0974-2255 Volume 3, Number 2 (2011), pp. 59-64

http://www.irphouse.com

Application of Data Mining in Manufacturing Industry

[1]V.K. Jha and [2]R.K. Singh

[1]Assistant Professor, Department of Information Technology,
BIT Mesra, Ranchi, India
[2]Professor & Head, Department of Information Technology,
MIT, Muzaffarpur, India

Abstract

The paper reviews applications of data mining in manufacturing engineering, in particular production processes, operations, fault detection, decision support, and product quality improvement. Data mining offers tools for discovery of relationships, patterns, and knowledge in large databases. The knowledge extraction process is computationally complex and therefore a subset of all data is normally considered for mining. In the case study reported in this paper, a data mining approach is applied to extract knowledge from a data set. The extracted knowledge is helpful for the prediction and prevention of manufacturing faults in engines.

Keywords: Data Mining, Manufacturing, Rapid Miner, JIT, Product development, Data interpretation, case study of automobile industry, Product Development, Software based Data analysis.

Introduction

Many organizations have based their aggressive strategies around automation and new production technology by adopting different applicable tools. Amongst these tools is Data Mining. The advancement in the computing technologies has made owning and running knowledge management systems and data warehouses or data marts easier and more cost efficient than it has ever been. This is particularly desired in order to keep pace with the changing customer's and business's needs. Managers are required to identify that the workforce knowledge and know-how collection must be strategically positioned to progress. An organization must have the ability to perform daily operations and continuous improvements. Different technologies can be used to track and monitor the organization's performance along several dimensions to ensure that the trends and values are on track. If not, critical signals can be inferred and actions can be initiated to improve the physical or resource value chains and their interactions with associated information and knowledge value chains.

Data Mining can be used in the following manufacturing domain:

- Data Mining in product design
- Data Mining in manufacturing lead time estimation
- Data Mining in quality
- Data Mining in supply chain management
- Data Mining in Just In Time manufacturing environment

Methodology

Data Collection

Preparing input for a data mining investigation usually consumes the bulk of the effort and time invested in the entire data mining process. When beginning work on a data mining problem, it is first necessary to bring all the data together into a set of instances. The data must be assembled, integrated, and cleaned up.

Data Mining tool

The Data Mining tool used on this paper is ***Rapid Miner*** (formerly YALE). Rapid Miner (fig 1)is the most comprehensive open-source software for intelligent data analysis, data mining, knowledge discovery, machine learning, predictive analytics, forecasting, and analytics in business intelligence (BI). Rapid Miner provides more than 400 data mining operators, a graphical user interface (GUI), an online tutorial with hands-on data mining applications, a comprehensive PDF tutorial, many visualization schemes (fig 2, fig 3) for data sets and data mining results, many different learning and meta-learning schemes ranging from decision tree and rule learners to neural networks, SVMs, ensemble methods, etc. Rapid Miner is implemented in Java and available under GPL (GNU General public License) as well as under a developer license (OEM license) for closed-source developers. The Data representation obtained are illustrated in fig 3 – a,b,c and d.

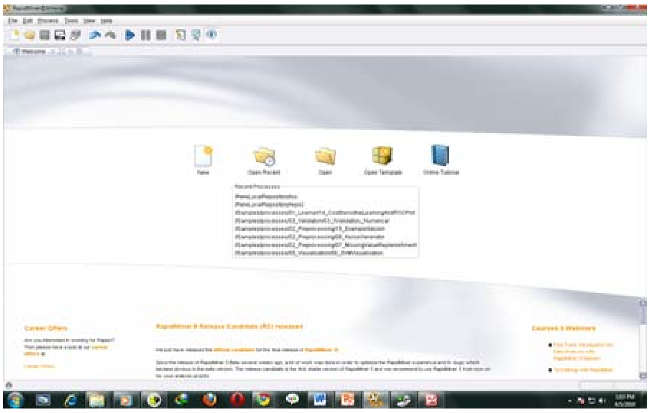

Figure 1: Rapid Miner.

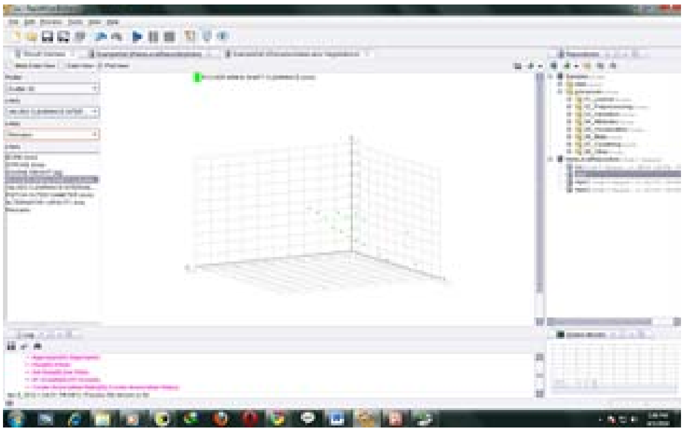

Figure 2: 3D Plot.

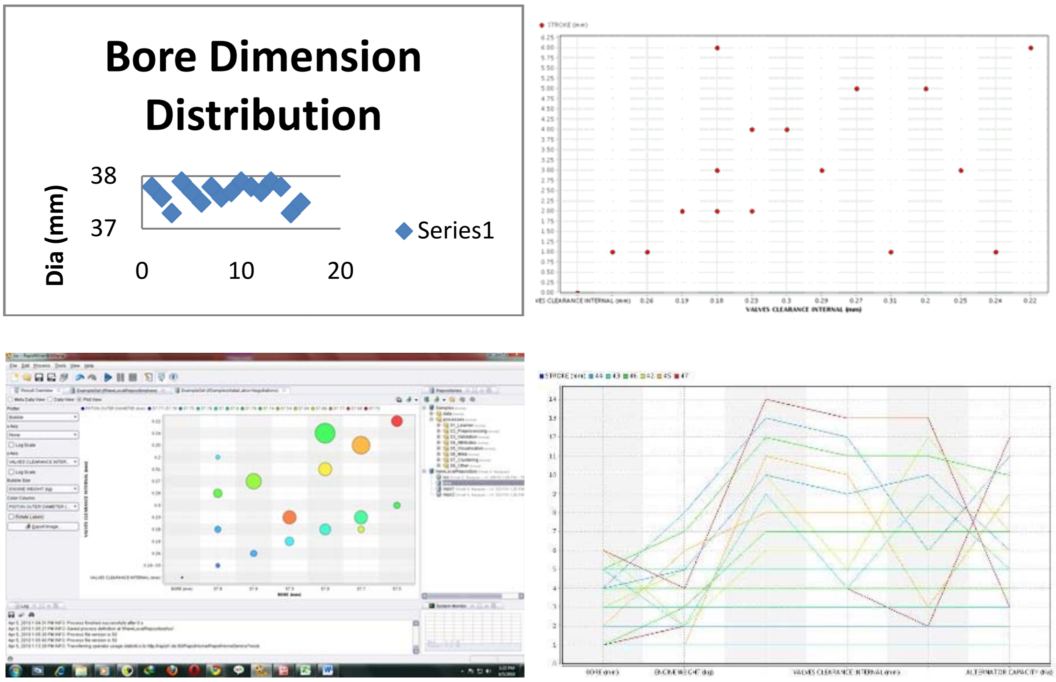

Figure 3(a-d): Data Representation.

Inferring Rules and Knowledge Representation

Most of the techniques used to present the output produce easily comprehensible descriptions of the structural patterns in the data. There are many different ways for representing the patterns that can be discovered by machine learning, and each one dictates the kind of technique that can be used to infer that output structure from data. The different types of techniques include:

- Decision Tables
- Decision Trees

- Classification rules
- Association rules
- Clusters

Algorithms Used

Decision Tables

The simplest, most rudimentary way of representing the output from machine learning is to make it just the same as the input—a decision table. creating a decision table might involve selecting some of the attributes. The problem is, of course, to decide which attributes to leave out without affecting the final decision. (fig 4). The no of iteration involved is dependent on available data and higher the iterations higher is the accuracy of results obtained.

ID	F1	F2	F3	F4	D
1	1.02	0.05	Yes	2.03	Accept
2	1.03	3.04	No	1.01	Reject
3	2.01	0.95	Yes	1.97	Accept
4	2.03	2.05	No	3.01	Accept
5	0.03	1.97	No	2.02	Reject
6	0.04	1.05	No	1.04	Reject
7	0.99	3.04	Yes	1.04	Accept
8	1.02	0.97	No	3.01	Reject

Figure 4(a): Decision Table.

Decision rule R1. (IF F3 = Yes) THEN (D = Accept); [1, 3, 7]
Decision rule R2. (IF F2 in [2.01, 2.55]) THEN (D = Accept); [4]
Decision rule R3. (IF F1 in [1.02, 1.52]) THEN (D = Reject); [2]
Decision rule R4. (IF F1 in [0.03, 0.52]) THEN (D = Reject); [5, 6]
Decision rule R5. (IF F2 in [0.96, 1.01]) THEN (D = Reject); [8]

Figure 4(b): Rule iteration.

Decision Trees

A "divide-and-conquer" approach to the problem of learning from a set of independent instances leads naturally to a style of representation called a *decision tree (fig 5).* Nodes in a decision tree involve testing a particular attribute. Leaf nodes give a classification that applies to all instances that reach the leaf. If the attribute that is tested at a node is a nominal one, the number of children is usually the number of possible values of the attribute.

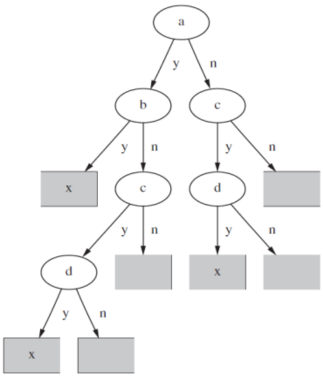

Figure 5: Decision Tree.

Case Study

Engine specifications were collected from a company from the quality department. The table-I below shows the data. The data shows the specification of 15 engines, which were tested for quality. The NOGO implies that those engines did not satisfy the quality tests.

Table I: Data Set.

BORE (mm)	STROKE (mm)	ENGINE WEIGHT (kg)	ROCKER ARM to SHAFT CLEARANCE (mm)	VALVES CLEARANCE INTERNAL (mm)	PISTON OUTER DIAMETER (mm)	ALTERNATOR CAPACITY (Kw)	REMARK
37.6	44	70	0.012	0.26	37.75	0.1	GO
37.3	43	68	0.014	0.19	37.78	0.099	NOGO
37.9	46	73	0.034	0.18	37	0.097	GO
37.7	43	72	0.035	0.23	37.8	0.104	GO
37.5	42	70	0.044	0.3	37.79	0.106	GO
37.8	46	68	0.037	0.29	37.74	0.103	GO
37.6	45	69	0.038	0.27	37.54	0.098	GO
37.7	43	70	0.027	0.18	37.69	0.104	NOGO
37.9	44	72	0.028	0.31	37.86	0.106	GO
37.8	45	71	0.031	0.2	37.78	0.109	GO
37.7	46	75	0.018	0.25	37.77	0.11	GO
37.9	44	76	0.019	0.24	37.79	0.095	NOGO
37.8	47	70	0.027	0.18	37.75	0.102	GO
37.3	42	72	0.028	0.23	37.68	0.103	GO
37.5	47	73	0.016	0.22	37.73	0.099	NOGO

Rule Sets

p1 = (37.3, 37.9)
p1 >37.4 && p2>43
p3>75 && p7>0.1
p5>2 || p6<37.86
p2>42
p6<p1
p5<0.3 && p5<0.16
p4>0.012 && p4<0.043
p5<p4

Rapid Miner was used to learn pattern from the data set given above. Since the volume of data was small the results could not be found properly. So if- then analysis was used to fine the trend in the data. Excel was used to learn the variation in the data on altering some of the parameters. The rule sets shown above were thus found out. (Where p1 is parameter 1, p2 is parameter 2 and so on as shown in the table.)

Conclusion

This paper outlines a framework for decision making based on the knowledge provided by the data mining techniques. Basic control limits and boundary conditions required for different attributes of an engine are specified. The origin of error can be found out using this method. It was also realized that the population density representation of the data needs to be much more. The data should be filtered and error in the data should be minimized.

The growing volume of data in manufacturing and service industries is a challenge that needs research on tools that discover unique properties of the data. Data mining is a discipline that offers tools for data analysis and knowledge discovery.

References

[1] Morgan Kaufmann Data Mining Practical Machine Learning Tools And Techniques - Ian H. Witten Department of Computer Science, University of Waikato and Eibe Frank Department of Computer Science, University of Waikato
[2] Data Mining: Techniques and Applications - Thaer Al-Nimer
[3] Data mining: manufacturing and service applications - A. KUSIAK, International Journal of Production Research, Vol. 44, Nos. 18–19, 15 September–1 October 2006, 4175–4191
[4] Survey of the Computational Intelligence Methods for Data Mining -Zdenˇek Buk
[5] Data mining - Wikipedia (http://en.wikipedia.org/wiki/Data_mining)
[6] http://www.the-data-mine.com/