International Journal of Data Mining Techniques and Applications
Volume: 04 Issue: 01 June 2015 Page No.15-18
ISSN: 2278-2419

Manufacturing Process Optimization using Data Mining Techniques

R Shankar[1], M Sundararajan[2]
[1]Research Scholar, Computer Science, Bharathiar University, Coimbatore,Indai
[2]Asst. Professor, Dept. of Computer science,Govt. Arts and Science College for Men,Nandanam, Chennai
Email:shankarramchanda@hotmail.com, drmsrajan23@yahoo.com

Abstract - Manufacturing, being referred to, as an integrated management process has been radically transformed with the advent of latest technologies and advancements. Data generated by machines are continuously encapsulated and archived in storage systems for analysis and futuristic applications.Such storage necessitates enormous warehouse capacities thus vitalizing methods deployed for storage. Apart from vitality, they serve as critical parameters in improving efficiency, as well. With regard to engine assembly plants, these data can comprehend subtle characteristics of the engine assembly and further their testing processes, as well.Simply put, each engine assembly comprises a minimum of 50 major assembly process parameters and 15 pivotal testing parameters meant for observation and record. Of course, they are of great significance to restore lineage and traceability.Competition demands the use of such legacy data repositories of automotive engine manufacturing companies owing to market demand. Hence trigger the need to deploy appropriate techniques to secure business insights for effective decision-making.Data Mining (DM) is one such subject that mandates a thorough and detailed study. Here in this paper it is handled with the clustering analysis of the subject DM with a sample data volume of five hundred engines' performance test result files.

Keywords:Clustering Analysis, Manufacturing management,Automotive Engine testing Process.

I. INTRODUCTION

Minim ax, there exist two requirements necessitating maintenance of manufacturing process data in an automotive engine manufacturing organization, via,

- Government regulations aligning with customerreservations;
- Demand for improvement of critical parameters that drive customer expectations!

The purpose of this paper is to justify the utility value of DM models in consuming process parameter data, amassed for extrapolating events, for applying control measures, for simulating trials of control range and for pursuing span of control.In the context of presenting this paper, the objective is to draw inference from cluster sets drawn against scatter graphs, plotted of data sets from process parameter repositories, to bring out the hidden pattern of Specific Fuel Consumption (SFC) of engines for differential torque valuesusing clustering analysis of DM from a selected volume of 500 datasets. The remaining of this paper is organized as follows, Section II reveals the relevant literature inline with concept of this paper which is the use of DM techniques in the process data generated in an automotive engine manufacturing system and the volume of real time data being generated and stored in data repositories. Section III explains the schema of an automotive engine manufacturing system and the rough estimate of data that are generated during the process of testing of the engines before assembling them in to the vehicles. The details about the 2 steps involved during the analysis of the data usingclustering technique has been explained in section IVand finally the section IV concludes the analysis activity.

II. LITERATURE SURVEY

Data mining for manufacturing and service applications is discussed by KUSIAK in his paper [5]. In this paper basic concepts of machine learning and data mining are introduced. Machine learning algorithms supports yielding knowledge from diverse data bases that can be used to build decision-making systems. For example, based on the operational engineering data, equipment faults can be detected, the number of items to be ordered can be predicted, optimal control parameters can be determined.Same way, the use of DM in improving the manufacturing Quality has widely been discussed by LIOR ROKACH and ODED MAIMON in their paper. Here they emphasize the importance and the necessity of DM techniques towards the identification of the hidden patterns of the abundant information and data which are produced during manufacturing activity. They specifically focus on mining the quality related data by the useofclassifier models towards the achievement of improvement in manufacturing quality [7].The need of DM techniques in identifying the inherent interdisciplinary character of the automobile development and manufacturing process has been apparently analyzed by RUDOLF KRUSE, MATHIAS STEINBRECHER and CHRISTIAN MOEWES in their paper which emphasizes the use of DM methodologies in automotive industries [8]. The authors reiterate the use of DM tools towards getting a valuable means to manage the process which produces overwhelmingly large information during the manufacturing process in an automotive industry.

III. ASSEMBLY AND TESTING PROCESS OF AN AUTOMOTIVE ENGINE

Fig.1 represents the schema of Engine Manufacturing System. Consequent to process time complexities, DM techniques surpassed other models meant for effective and efficient use of data collected over substantial periods.Performance starts with the mandate to conform to theoretical design specifications. Though, this depends on various factorsas supplier parts, handling of materials, equipment etc., and the most blatant fact remains that quality of the final product is gauged over real-time performance.Testing, though being performed in the end, it still endorses the product on its availability to the customer,

Volume: 04 Issue: 01 June 2015 Page No.15-18
ISSN: 2278-2419

promising on its delivery commitments. Testing qualifies the engine for delivery and sale, thereof. The number of production Units per Hour (UPH) / Jobs per Hour (JPH) in the assembly lines span between15 - 60 UPH / JPH. Most certainly, this conditions the manufacturer to maintain data volumes ranging between 1000 - 4000 Data Sets / Hour [1]. This obviates collection of about 16,000 - 64,000 Data-sets / Day and expands to about 48, 00,000 - 1, 92, 00,000 Data-sets / Year. This is simply huge.As the collection of data from the testing of engines by real time basis during their manufacturing from the production houses lies large, this paper encourages the DMpathway.Quoting SAP America Inc., [2] this paper brings to contention that enterprises world-wide, use just 10% of data compiled for analytics and decision-making. What is baffling is this insensitivity to corporate data assets that sail through time, uncared for, remaining as non-performing assets. This is where DM adds vitality.Most manufacturing processes running process control systems deploy various types of sensors and feedback devices to track parameters such as pressure, flow rate, temperature, variables and factors of efficiency needed to exert control over processes [6]. When collected, these, as huge data repositories, are meant for proper and purposeful utilization for better performance.What suits this scenario fittingly is DM clustering analysis.

IV. THE CLUSTERING TECHNIQUE AND ANALYSIS

The process of partitioning sets of data or objects into sets of meaningful sub-classes begets clusters [3]. These throw light on the nature of grouping and the resultant structure in data-sets.This paper examines, a sample volume "N", being test result data-sets of 500 engines which are ASCII files generated as test results from the Engines Dynamometer testing equipment.

$$N \rightarrow 500$$

Dynamometer testing equipment is special testing equipment in automotive testing process of engines which applies load to the engines during its running conditions as specified in Pass-off standards given in the engine design specifications. During the process of applying load to the engines the resultant characteristics are collected as data-sets against the engine serial number and stored in repositories mainly for genealogy and traceability purposes.Real-time data-sets (ASCII files) gathered of manufactured engines have been corroborated. Applying the clustering technique of DM, data-sets thus obtained during Engine Performance Testing for every single engine on Power, Torque, SFC, Temperature etc., have been tabulated as shown in the Table: 1.This is for inference and improvement.Factors, for this paper, for the study of deployment of clustering are SFC and Torque. Results obtained as multitudes of specific SFC values against particular Torque ratings over clusters are comprehended. Even visual inspection reveals the intensity range of clusters where SFC scores are in maximum level against certain Torque numbers, on clustering.This enables to identify correct Torque values for which SFC remains high. This extends in establishing appropriate process design for better SFC production.

Table 1: Data table view

Sl. No	Test Bed	Date	Time	Engine Number	Engine Type	REPORT command	Corr. Torque	Corr. Power	SFC	Operator Comment/ Result
1	3	9-Jan-13	10:32:19	CDH010783P	A2N03900	Step 1200 Rpm	114.7	14.4	282	OK
2	3	9-Jan-13	10:32:19	CDH010783P	A2N03900	Step 2200 Rpm	167	38.4	215	OK
3	3	9-Jan-13	10:32:19	CDH010783P	A2N03900	Step 3300 Rpm	127.8	44.1	238	OK
4	1	9-Jan-13	10:46:44	CDH010965P	A2N10000	Step 1200 Rpm	120.9	15.2	281	OK
5	1	9-Jan-13	10:46:44	CDH010965P	A2N10000	Step 2200 Rpm	162.6	37.5	221	OK
6	1	9-Jan-13	10:46:44	CDH010965P	A2N10000	Step 3300 Rpm	126.4	43.7	241	OK
7	2	9-Jan-13	10:59:29	CDH010968P	A2N10000	Step 1200 Rpm	117.2	14.7	241	OK
8	2	9-Jan-13	10:59:29	CDH010968P	A2N10000	Step 2200 Rpm	160.4	37	206	OK
9	2	9-Jan-13	10:59:29	CDH010968P	A2N10000	Step 3300 Rpm	123.5	42.6	239	OK
10	4	9-Jan-13	11:04:43	CDH010966P	A2N10000	Step 1200 Rpm	123.1	15.5	256	OK
11	4	9-Jan-13	11:04:43	CDH010966P	A2N10000	Step 2200 Rpm	164.4	37.9	212	OK
12	4	9-Jan-13	11:04:43	CDH010966P	A2N10000	Step 3300 Rpm	127	43.9	234	OK
13	1	9-Jan-13	11:05:39	CDH010949P	A2N03900	Step 1200 Rpm	119.5	15	260	OK
14	1	9-Jan-13	11:05:39	CDH010949P	A2N03900	Step 2200 Rpm	161.4	37.2	223	OK
15	1	9-Jan-13	11:05:39	CDH010949P	A2N03900	Step 3300 Rpm	124.8	43.1	242	OK
16	3	9-Jan-13	11:16:11	CDH010969P	A2N10000	Step 1200 Rpm	124.4	15.6	253	OK
17	3	9-Jan-13	11:16:11	CDH010969P	A2N10000	Step 2200 Rpm	171.3	39.4	209	OK
18	3	9-Jan-13	11:16:11	CDH010969P	A2N10000	Step 3300 Rpm	131.6	45.4	228	OK
19	2	9-Jan-13	11:21:47	CDH010971P	A2N10000	Step 1200 Rpm	118.8	14.9	246	OK
20	2	9-Jan-13	11:21:47	CDH010971P	A2N10000	Step 2200 Rpm	161.7	37.3	223	OK
21	2	9-Jan-13	11:21:47	CDH010971P	A2N10000	Step 3300 Rpm	123.5	42.7	218	OK
22	1	9-Jan-13	11:24:25	CDH010970P	A2N10000	Step 1200 Rpm	120	15	264	OK
23	1	9-Jan-13	11:24:25	CDH010970P	A2N10000	Step 2200 Rpm	163.4	37.6	219	OK
24	1	9-Jan-13	11:24:25	CDH010970P	A2N10000	Step 3300 Rpm	126.1	43.6	240	OK
25	4	9-Jan-13	11:28:56	CDH010973P	A2N10000	Step 1200 Rpm	121.1	15.2	261	OK
26	4	9-Jan-13	11:28:56	CDH010973P	A2N10000	Step 2200 Rpm	165.1	38	216	OK
27	4	9-Jan-13	11:28:56	CDH010973P	A2N10000	Step 3300 Rpm	126.1	43.6	238	OK
28	3	9-Jan-13	11:40:30	CDH010976P	A2N10000	Step 1200 Rpm	125.1	15.7	252	OK

Why Torque? Torque, for commercial vehicle, is the determining factor for serving horse power to engine. More horse power implies better load-bearing capacity, speed. The experiment hence determines the range over which SFC reaches its maximum.By modifying the assembly process settings, it is possible to manufacture engines by delimiting performance variances over Torque and SFC values. What is fitting for SFC and Torque can be extrapolated to befit other critical parameters, as well for more improvements. In doing so, such deviations, limitations and variances can be leveraged and process design be made robust.Using tabulated values as shown in Table: 1.A sample cluster has been obtained using DM which is explained below in steps 1 and 2.Step- 1-Exploring the data from Table: 1.Graphically plotting, SFC values along Y axis and the respective engine numbers in X axis, we have Fig:2.thusarrived.Interpretation of results obtained during testing as process parameter data-sets:-The green line on the graph represents the safe limit of SFC which is 233 g / KWh.If SFC value is ≤ 233 g / kWh at a fixed engine speed of 2200 rpm, then the process is conforming to engine process design specifications; this is traced with the help of green colour arrow in the graph.The overall trend reveals that almost 99.4% of engines are within specified limits of SFC with regard to engine process design specification [8].However, the customer survey conducted as post sales feedback revealed that 30% of customers are not happy with the performance with the product in terms of fuel economy of the vehicle which is contrary to the obtained test result shown in Fig: 2.This is where this paper builds logic and substance for improvement towards the data analysis methodology.

International Journal of Data Mining Techniques and Applications
Volume: 04 Issue: 01 June 2015 Page No.15-18
ISSN: 2278-2419

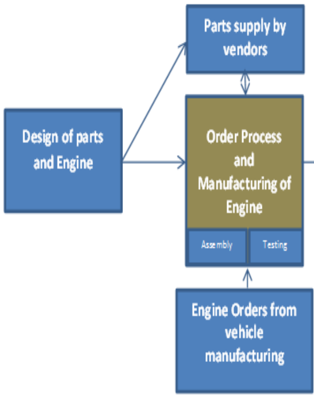

Figure 1

To isolate areas of concern from the rest and to distinguish each such collected set of data based on performance, in most manufacturing problems, it is necessary to project the summarized data in descriptive terms along with overall picture.However, due to limited linkage with real-time data, the uses of tools for all purposeful decision-making processes have been minimal, in practice [7]. This has been whipping and constraint-bound, in the past.DM has weeded out such inadequacies and has made deep inroads. Putting DM to practice, business decisions are implemented [4]. These have proven to be quite informed, transparent and autonomous.Step -2 -Compounding relevance of attributes, suffice to their differences, clustering has been computed on the design process parameter data-set tabulated above with different sets of Torque values ending with different volumes of SFCs, but at the same rpm.In other words, the experiment has been repeated for different Torque values at the same rpm for all 500 engines and results captured. Fig.3 represents SFC - Torque relation graph with identified clusters.From the identified clusters, it has been inferred that results mainly fell into 5 major categories, discarding anomalies that ended up beyond the design specifications of torque and SFC

Specifications:

Torque	$\geq$	150 Nm
SFC	$\leq$	233 g / kWh

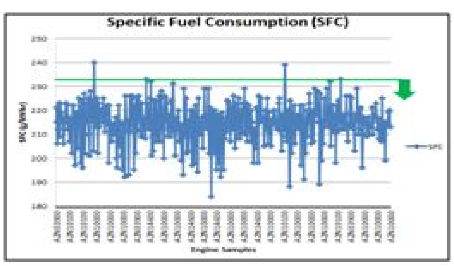

Figure 2: Graph of SFC results of tested engines

The five major clusters are classified as,

C1 Engines with Very good Torque and Good SFC
C2 Engines with Very Good SFC and Good Torque
C3 Engines with Good Torque and Good SFC
C4 Engines with Nominal SFC and Nominal Torque
C5 Engines with Nominal SFC and Good Torque

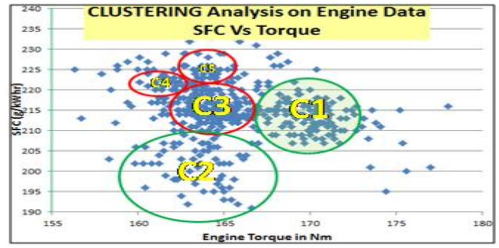

Figure 3:Clustered SFC - Torque relation Data

From the worst to the best these clusters may be sequenced into *C4, C5, C3, C1 and C2*Which is for further analysis as mentioned in Fig: 4. Firstly, this paper warms up DM into Engine Design Process Parameter Settings and extends into deploying clustering for study. Secondly, the paper infers the contour, over which influences of Torque and SFC are realized. As a third step, this paper, enables conditions for optimality with regard to fuel consumption, isolating constraints.As a result, a comprehensive and conclusive analysis towards the final solution can be taken up for further study. The authors are working on the same data-sets, applying Association Rule Mining Technique, one more of DM Application to analyze the associated parts of the engine to move closer to arrive at the conditions of optimality [5]. As a result, a comprehensive and conclusive analysis towards the final solution can be taken up for further study. The authors are working on the same data-sets, applying Association Rule Mining Technique, one more of DM Application to analyze the associated parts of the engine to move closer to arrive at the conditions of optimalitywith regard to process design parameter settings for optimal consumption of fuel at specific values of control parameters.

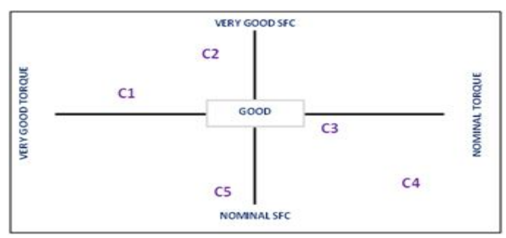

Figure 4:Clustered SFC - Torque relation Data

V. CONCLUSION

Some trivial states of production results revealed especially when it is dealt with real-time reports of“engine performance process test result analytics” where the *Process capability constant*(Cpk) establishment becomes very tough and critical.Therefore, theoretical events may be simulated towards achieving the same by using Data-Mining techniques. This attempt is a significant step towards expressing the relationship between SFC and Torque standards since the work has been consistent and within specified design limits.This attempt has revealed many interesting patterns and brought out facts about test result anomalies. This is where, manufacturers shall have to pay due attention, obtain insight into business and end up improving quality standards of their products to increase their market shares.

REFERENCES

Volume: 04 Issue: 01 June 2015 Page No.15-18
ISSN: 2278-2419

[1] Data Mining Technologies Inc.Web. 14 Nov. 2014. < http://www.data-mine.com/white_papers/Manufacturing >.

[2] SAP Inc. Web. 20 Dec 2014. < http://www.sap.com/asset/detail.2014-04-apr.what-is-predictive-analytics-data-mining-mp4.html >.

[3] Jiawei Han, Michelin Kamber, JianPei. “Data Mining Concepts and Techniques” -3rd Edition,pp. 443-445.

[4] Waldemar Wójcik and Konrad Gromaszek (2011). “Data Mining Industrial Applications” , “Knowledge-Oriented Applications in Data Mining”, Prof. Kimito Funatsu (Ed.), ISBN: 978-953-307-154-1, InTech, Available from http://www.intechopen.com/books/knowledge-oriented-applications-in-data-mining/data-mining-industrialapplications, pp. 435 – 440.

[5] Kusiak, A. “Data mining: manufacturing and service applications”, International Journal of Production Research, Vol. 44 (2006), pp. 18-19, ISSN: 4175-4191.

[6] Salim Charaniya, Huong Le, Huzefa Rangwala, Keri Mills, Kevin Johnson, George Karypis, Wei-Shou Hu, “Mining manufacturing data for discovery of high productivity process charecteristics” Journal of Biotechnology 147 (2010), pp. 186–197.

[7] Lior Rokach, Oded Maimon, “Data mining for improving the quality of manufacturing: a featureset decomposition approach” pp. 26-27.
Rudolf Kruse, Matthias Steinbrecher, Christian Moewes, “Data Mining Applications in the Automotive Industry” proceedings of 4th International Workshop on Reliable Engineering Computing (REC 2010), pp. 25-26.