Ultra Scientist Vol. 26(2)B, 174-180 (2014).
www.ultrascientist.org

ISSN 2231-3478 (Print)
2319-8052 (Online)

Application of Association rule Mining for Waste Heat Recovery Boiler

[1]NEERAJ BHARGAVA, [2]ATIF AZIZ and [3]RAMESH PRASAD AHARWAL

[1]Associate. Prof., Department of Computer Science, MDS University Ajmer (INDIA)
[2]Research Scholar NIMS University
[3]Asstt. Prof., Department of mathematics, Govt. P.G. College, Damoh (M.P.) (INDIA)

(Acceptance Date 11th August, 2014)

Abstract

Data mining is a technique for discovering positive information from huge databases. This technique is currently being beneficially used by a number of industries. A common approach for information discovery is to identify association rules which make known relationships among different items. Our aim is to obtain association rules among various attributes of mechanical data. Entire database was used to obtain these association rules. This paper presents an application of Association Rule mining to discover precious association rules for data obtained from mechanical industry. The data used in this is of waste heat recovery boiler used in a fertilizer industry. In India only alittle work is done on mining of mechanical industrial data. In this paper we have used Apriori Algorithm with WEKA data mining tool.

Key words: Data Mining, Association Rule, KDD, WEKA.

Introduction

Association Rule Association rule mining, one of the most important and well researched techniques of data mining. It aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases. Association Rule Mining (ARM) is a technique used to discover relationships among a large set of variables in a data set. The concept of association rule mining firstly proposed by[1], Association Rule Mining refers to the discovery of relationships among a large set of variables. That is, given a database of records, each containing two or more variables and their respective values. ARM is also used to discover variable relationships, but each relationship also known as an association rule may contain two or more variables. Data mining is the key step in the knowledge discovery process, and association rule mining is a very important research topic in the data mining field[1]. There is a need to introduce datamining technique to mechanical industrial for the purpose of research and development because this technique saves both time and money and the data is real time data which is collected during full fledged working of the machines with shuting down the machine. The data used in this paper is of waste heat recovery boiler which uses dearator, economiser, superheater. The system uses the waste heat of the sulphuric acid plant in generation of low pressure steam. The relation between temperature and pressure is found out and hidden patterns were revealed through assoiation rule mining (ARM). The efficiency of carnot cycle depends on source temperature and sink temperature. With the help of Weka we have found the correlations with various parameters of waste heat recovery boiler. The data obtained from the industry is preprocessed to make it compatible with the software.

Knowledge Discovery and Database :

Knowledge Discovery in Databases (KDD) has become one of the fastest growing research topics in mathematics and computer science, because the ability to continually change and acquire new understanding is a driving force for its applications[6] Washio, 2007. The KDD process, specifically data mining techniques, is used to characteristically discover knowledge from data[9]. The data mining process extracts knowledge from an existing data set and transforms it into a human-understandable structure for further use[8]. Data mining techniques are required to help in identification of model characteristics important to capture and document in an enhancement context of the safety and reliability of complex engineering systems[7]. Data mining is a very important analysis activity of the Knowledge Discovery in Databases (KDD) process, which is an interdisciplinary field of computer science; this refers to a very broad process of finding knowledge in a large database. In order to find knowledge, a standard process has been developed, ‘‘The Knowledge Discovery in Databases process’’[5]: As seen on following Figure, the KDD process extracts knowledge from data in four different steps. The first step, selection, develops the understanding of the application domain, of the prior knowledge and the goals of the end-user. A target data is created; the selection of data in which the discovery will be performed. During the preprocessing step, the data is cleaned from noise and outliers; the necessary modeling information is collected.

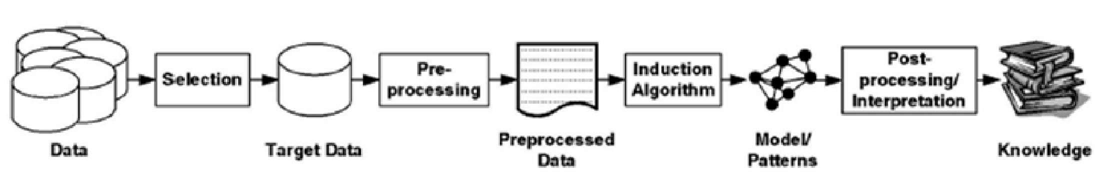

Figure 1. KDD Process

Basic Concepts & Basic Association Rules Algorithms :

Let I=I1, I2, … , Im be a set of m distinct attributes, T be transaction that contains a set of items such that T I, D be a database with different transaction records Ts. An association rule is an implication in the form of XY, where X, Y I are sets of items called itemsets, and X ∩ Y = φ. X is called antecedent while Y is called consequent, the rule means X implies Y. There are two important basic measures for association rules, support and confidence. Since the database is large and users concern about only those frequently purchased items, usually thresholds of support and confidence are predefined by users to drop those rules that are not so interesting or useful. The two thresholds are called minimal support and minimal confidence respectively. Support(s) of an association rule is defined as the percentage/fraction of records that contain X Y to the total number of records in the database. Suppose the support of an item is 0.1%, it means only 0.1 percent of the transaction contain purchasing of this item. Confidence of an association rule is defined as the percentage/fraction of the number of transactions that contain $X \bigcup Y$ to the total number of records that contain X. The support s of the rule is the prior probability of X and Y,

$$\mathrm{s} = \sup(\mathrm{x} \cap Y) = \frac{|X \cup Y|}{n}, and$$

The confidence c of the rule is the conditional probability of Y given X,

$$c = \frac{\sup(X \cup Y)}{\sup(X)} = \frac{|X \cup Y|}{|X|}$$

The Apriori algorithm:

This algorithm has emerged as one of the best association rule mining algorithms. It also serves as the base algorithm for most parallel algorithms. Apriori uses a complete, bottom-up search with a horizontal layout and enumerates all frequent itemsets. It is based on data passes. It identifies frequent "itemsets", subsets of items with a transaction, by performing as many data passes as specified by the user, or until there are no additional frequent itemsets to be identified.

The Apriori algorithm[2,3] finds frequent itemsets from databases by iteration. At each iteration *i* the algorithm attempts to determine the set of frequent patterns with *I* items and this set is engaged to generate the set of candidate itemsets of the next iteration. The iteration is repetitively performed until no candidate patterns can be discovered. It uses a bottom up approach, where frequent subsets are extended one item at a time. In the input datasets are referred as sequences composed of more or less items. The output of Apriori is a set of rules explaining the links these items have in their sets[4].

WEKA Data mining software:

Weka is developed at the University of Waikato in New Zealand. "WEKA" stands for the Waikato Environment of Knowledge Analysis. The system is written in Java, an object-oriented programming language that is widely available for all major computer platforms, and Weka has been tested under Linux, Windows, and Macintosh operating systems. Java allows us to provide a uniform interface to many different learning algorithms, along with methods for pre and post processing and for evaluating the result of learning schemes on any given dataset. Weka expects the data to be fed into to be in ARFF format. It is necessary to have information about each attribute which cannot be automatically deduced form the attribute values (Witten and Frank, 2000). Weka includes a variety of tools for preprocessing a dataset, such as attribute selection, attribute filtering and attribute transformation, feeding into a learning scheme, and analyze the resulting classifier and its performance. Weka is organized in packages that correspond to a directory hierarchy. The important packages of Weka are association, attribute selection, classifiers, clusterers, estimators, and filters packages. The association package has only one association rule.

Model Building :

In the initial experiment the researcher took 21 attributes for association rule model building purpose. The selection of attribute is made using subjective judgment. To build the association rule model, the arff format of the selected dataset was given to Weka, apriori algorithm. The following is the first twenty rules generated in this experiment.

Experimental setup :

=== Run information ===
Scheme: weka.associations.Apriori -N 20 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation: atifCSV PHD
Instances: 24
Attributes: 21
=== Associator model (full training set) ===
Apriori
=======
Minimum support: 0.45 (11 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 11
Generated sets of large itemsets:
Size of set of large itemsets L(1): 17
Size of set of large itemsets L(2): 31
Size of set of large itemsets L(3): 9

Screen shots during learning association rules from WEKA software

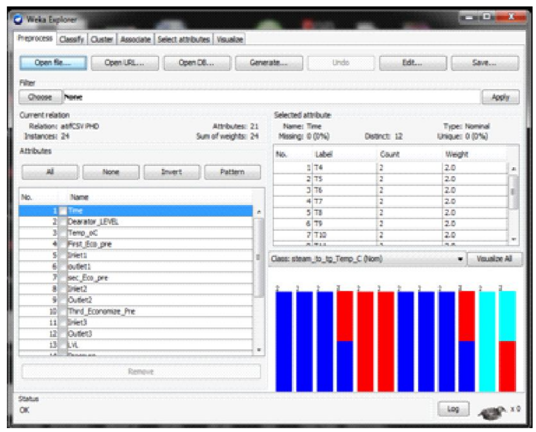

Figure 2 attributes list with weka explorer

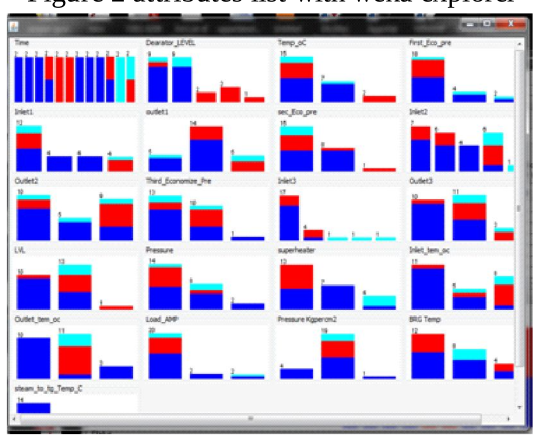

Figure 3 attributes statistics with weka explorer

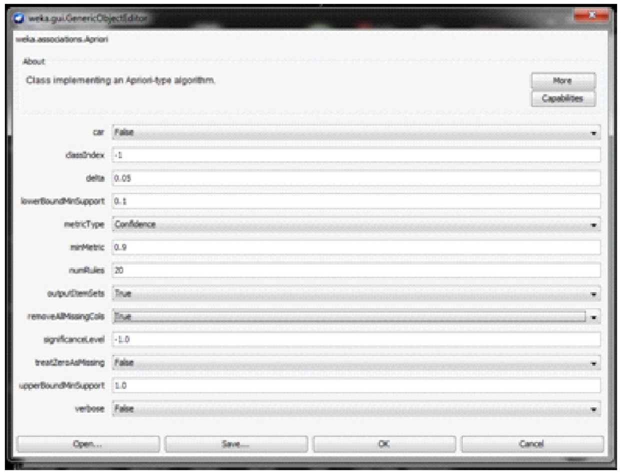

Figure 4 Parameters setting

Results and Discussion

S.No.	Association Rules	conf.
1	superheater=Normal 13 ==> First_Eco_pre=High 13	1
2	BRG Temp =Low 12 ==> First_Eco_pre=High 12	1
3	Temp_oC=Normal Inlet3=Extremely Low 12 ==> First_Eco_pre=High 12	1
4	Outlet_tem_oc=High 11 ==> Pressure Kgpercm2=Low 11	1
5	Temp_oC=Normal superheater=Normal 11 ==> First_Eco_pre=High 11	1
6	sec_Eco_pre=High Inlet3=Extremely Low 11 ==> First_Eco_pre=High 11	1
7	superheater=Normal Load_AMP=Low 11 ==> First_Eco_pre=High 11	1
8	Load_AMP=Low BRG Temp =Low 11 ==> First_Eco_pre=High 11	1
9	outlet1=Normal 14 ==> Inlet3=Extremely Low 13	1
10	steam_to_tg_Temp_C=Low 14 ==> Inlet3=Extremely Low 13	1
11	Pressure =Normal 14 ==> Load_AMP=Low 13	0.93
12	Third_Economize_Pre=Normal 13 ==> First_Eco_pre=High 12	0.92
13	LVL=High 13 ==> Load_AMP=Low 12	0.92
14	LVL=High 13 ==> Pressure Kgpercm2=Low 12	0.92
15	First_Eco_pre=High Inlet3=Extremely Low 13 ==> Temp_oC=Normal 12	0.92
16	Temp_oC=Normal First_Eco_pre=High 13 ==> Inlet3=Extremely Low 12	0.92
17	sec_Eco_pre=High Load_AMP=Low 13 ==> First_Eco_pre=High 12	0.92
18	First_Eco_pre=High sec_Eco_pre=High 13 ==> Load_AMP=Low 12	0.92
19	BRG Temp =Low 12 ==> Load_AMP=Low 11	0.92
20	First_Eco_pre=High BRG Temp =Low 12 ==> Load_AMP=Low 11	0.92

Table 1 Database table

Time	Dearator _LEVEL	Temp _oC	First _Eco _pre	Inlet1	outlet1	Sec _Eco _pre	Inlet2	Out-let2	Third_ Eco_ Pre	Inlet3	Outl-et3	Pres-sure	super-heater	Inlet _tem_oc	Load _AMP	Press-ure Kgper-cm2	BRG Temp	steam_ to_tg_ Temp_C
T4	Low	Normal	High	High	Low	High	Low	Low	Normal	Extremely Low	Low	Normal	Normal	Low	Low	Normal	Low	Low
T4	High	Normal	High	High	Normal	High	Low	Low	Normal	Extremely Low	Low	Normal	Normal	Low	Low	Normal	Low	Low
T5	Low	Low	Normal	Low	Normal	Normal	Normal	Normal	Low	Extremely Low	Low	Normal	Low	Low	Low	Low	Normal	Low
T5	Low	Low	Low	Normal	Normal	Normal	Normal	Normal	Low	Extremely Low	Normal	Normal	Low	Normal	Low	Low	Normal	Low
T6	Low	Normal	High	Normal	Normal	High	Low	Low	Normal	Extremely Low	Normal	Normal	Low	Normal	Low	Low	Normal	Low
T6	High	Normal	High	Normal	Normal	High	Extremely low	Low	Normal	Extremely Low	Normal	High	Normal	Normal	Low	Low	Low	Low
T7	Low	Low	High	Low	Low	High	Extremely low	Low	Normal	Extremely Low	Normal	High	High	High	Low	Normal	Low	Low
T7	Extremely Low	Normal	High	Extremely High	High	High	High	Low	Normal	Extremely Low	High	High	Normal	Low	Low	Low	Low	Normal
T8	Low	Normal	High	High	Normal	Normal	Low	Low	Normal	Low	Low	Normal	Normal	Normal	Low	Low	Low	Normal
T8	Extremely High	Normal	High	High	Normal	High	Normal	High	Low	Extremely Low	Normal	Normal	Normal	High	Low	Low	Low	Normal
T9	Extremely Low	High	High	Extremely High	High	High	High	High	Low	Low	Normal	Normal	Normal	High	Low	Low	High	Normal
T9	Extremely high	Normal	High	High	Normal	High	Normal	High	Low	Extremely Low	Normal	Normal	Normal	High	Low	Low	Low	Normal
T10	Low	Low	Normal	High	Normal	Normal	Normal	Normal	Low	Extremely Low	Low	Low	Low	Low	Low	Low	Normal	Low
T10	High	Normal	High	Normal	Low	Normal	Extremely low	High	Normal	Extremely Low	Normal	High	Normal	Low	High	Low	High	Low
T11	High	Normal	High	High	Normal	High	Low	Low	Normal	Extremely Low	Low	Normal	Normal	Low	Low	Normal	Low	Low
T11	High	Normal	High	High	Normal	High	Low	Low	Normal	Extremely Low	Low	Normal	Normal	Low	High	High	High	Low
T12	Low	Low	Normal	Low	Normal	Normal	Normal	Normal	Low	Extremely Low	Low	Low	Low	Low	Low	Low	Normal	Low
T12	High	Normal	High	High	Normal	Normal	Extremely low	High	Normal	Extremely Low	Low	High	Low	Low	Norma	Low	Low	Low
T1	High	Low	High	Low	Low	Normal	High	High	High	Low	Low	High	Low	Low	Low	Low	Low	Low
T1	Extremely high	High	High	Extremely High	High	high	High	High	Low	Low	High	Normal	Normal	High	Low	Low	High	Normal
T2	High	Normal	Normal	High	Low	High	extremely High	Low	Low	Normal	Normal	High	High	High	Norma	Low	Normal	High
T2	Low	Low	High	High	High	High	High	Normal	Normal	Extremely High	Normal	High	High	Normal	Low	Low	Normal	High
T3	High	Normal	Low	Extremely High	High	High	High	High	Normal	Extremely	High	Normal	High	High	Low	Low	Normal	High
T3	Extremely high	Normal	High	High	Normal	High	Low	High	Low	Extremely Low	Normal	Normal	Normal	High	Low	Low	Low	Normal

The results show a confidence level of 1 *i.e.* 100% which means that the rules generated will always give same results from 1-10. From 11-20 there is confusion as the confidence level is 92%. The superheater gives normal range of temperature when economizer 1st gives high temperature range.

Conclusion

This paper focused on application of Association rule mining on mechanical data. We have used apriori algorithm for learning association rule. Association rules were generated with the help of WEKA software which is shown above table.

References

1. Agrawal, R., Imielinski, T., and Swami, A. N., Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD *International Conference on Management of Data,* 207-216 (1993).
2. Agrawal, R. and Srikant, R., Fast algorithms for mining association rules. In Proc. 20th Int. Conf. Very Large Data Bases, 487-499 (1994).
3. Agarwal, R. Aggarwal, C. and Prasad V., A tree projection algorithm for generation of frequent itemsets. *In J. Parallel and Distributed Computing,* (2000).
4. Han, Jiawei and Kamber, Micheline, Data Mining: Concepts and Techniques and Applications. San Fransisico; Morgan Kufman Publishers (2001).
5. Fayyad, Usma, Piatetsky-shapiro, G. and Smyth, Padharic, From Data Mining to Knowledge Discovery in Databases. Available URL: http://citeseer.nj.nec.com.fayyad96from.html (1996).
6. Liao, S-H., Chu, P-H., & Hsiao, P-Y., Data mining techniques and applications – A decade review from 2000 to 2011. *Expert Systems with Applications, 39(12),* 11303–11311 (2012).
7. Saitta, S., Raphael, B., and Smith, I. F. C., Data mining techniques for improving the reliability of system identification. *Advanced Engineering Informatics, 19(4),* 289–298 (2005).
8. Witten, I. H., Frank, E., and Hall, M. A., Data mining: Practical machine learning tools and techniques (3 ed.). Elsevier. ISBN 978-0-12-374856-0 (2011).
9. Zhu, X., and Davidson, I., Knowledge discovery and data mining: Challenges and realities. New York: Hershey. ISBN 978-1-59904-252-7 (2007).