Fault Diagnosis of Mechanical Machines Based on Symbolic Value Partition Technique and the Generalized Distribution Table

Hossam A. Nabwey and Mahdy S. El-Paoumy

Abstract— the task of condition monitoring and fault diagnosis of rotating machinery is both significant and important but is often cumbersome and labour intensive. Automating the procedure of feature extraction, fault detection and identification has the advantage of reducing the reliance on experienced personnel with expert knowledge. Various diagnostics methods have been proposed for different types of mechanical machines. This paper presents a method to extract fault diagnosis rules for mechanical machines. First, a decision table for fault diagnosis is obtained by discretization of continuous symptom attributes from original data; second, the discretized fault symptom attributes are reduced using rough set methodology; finally, a set of maximally generalized decision rules is generated by using a rule induction algorithm based on the symbolic value partition technique and the Generalized Distribution Table (GDT). The proposed method effectively reduces both the number of attributes and the size of attributes domains. Furthermore, it help computing smaller rule sets with better coverage and better classification accuracy rates compared with those of the attribute reduction approaches which only reduce the number of attributes

Index Terms— Rule Induction, data mining, knowledge discovery in database, Fault diagnosing, rough set theory, the Generalized Distribution Table, attributes reduction.

1 INTRODUCTION

THE fault progression process of mechanical systems usually consists of a series of degraded states due to component wear and fatigue during the operating process. Early detection of incipient faults and foretelling the future states of mechanical systems can minimize the costs of unnecessary maintenance, avoid unplanned breakdown and enable maintenance actions to be scheduled more effectively. Hence, the availability and reliability of machine can be increased. Consequently, machine fault diagnosis has been important subject of research in the recent years.

Machine fault diagnosis is the ability to detect fault, isolate failed component, and decide on potential impacts of the failed component on the system health. Due to the costs of implementing, only critical machine components, whose failures drastically affect the breakdown, are frequently examined. In this Paper, Mechanical Machines are considered, due to their indispensable role in several industrial applications. The faults of Mechanical Machines may not only cause interruption of product operation, but also increase costs, decrease product quality and effect safety of operators. Consequently, fault diagnosis in Mechanical Machines has been the subject of serious studies in the recent years.

- *Hossam A. Nabwey is currently Assistant Professor, Department of Mathematics, Faculty of Science, Salman Bin Abdulaziz University, Al-Kharj, Saudi Arabia. The permanent address is Department of Engineering Basic science, Faculty of Engineering, Menofia University, Menofia, Egypt. E-mail: eng_hossam21@yahoo.com*
- *Mahdy S. El-Paoumy , Department of Mathematics, Faculty of Science, Salman Bin Abdulaziz University, Al-Kharj, Saudi Arabia.*

In order to detect/diagnose faults, system identification and parameter estimation [1], as well as other techniques [2],[3],[4],[5] have been proposed. These techniques required expensive equipment or accurate mathematical models, which are challenging, to describe the faults of Mechanical Machines.

Generally, the data obtained from measurements are of high dimensionality and have a large amount of redundant features. If the data are directly input into the classifier, the performance will be significantly decreased. Feature extraction and selection have been utilized for reducing dimension of data by selecting important features, with feature extraction and transformation of existing features into a lower dimensional space [6] . Nevertheless, each feature set contains many redundant or irrelevant features; along with salient features in feature space after feature extraction has been done. Consequently, there is a need for a feature selection tool to achieve good learning, classification accuracy, compact and easily understood knowledge-base, and reduction in computational time [7].

Recently, Rough set theory is a relatively new mathematical and AI technique introduced by Pawlak and Skowron [8], [9]. The technique is particularly suited to reasoning about imprecise or incomplete data and discovering relationships in this data. The main advantage of rough set theory is that it does not require any preliminary or additional information about data-like probability in statistics, basic probability assignment in DS theory or the value of possibility in fuzzy set theory

This paper introduces a method to extract diagnosis rules for mechanical machines. The proposed method consists of two stages. First, using the symbolic value partition technique, which divides each attribute domain of a data table into a fam-

ily of disjoint subsets, and construct a new data table with less attributes and smaller attribute domains? Second, using the Generalized Distribution Table (GDT) as a hypothesis search space and combining the GDT with the rough set methodology to generate decision rules from the new data table.

2 PRELIMINARY

2.1 Preliminary concepts of RST

This section recalls some essential definitions from RST that are used for feature selection. Detailed description and formal definitions of the theory can be found in [10]. The notion of information table has been studied by many authors as a simple knowledge representation method. Formally, an information table is a quadruple $S=(U,A,V,f)$, where: U is a nonempty finite set of objects, A is a nonempty finite set of features, V is the union of feature domains such that $V=\bigcup_{a\in A} V_a$ for Va denotes the value domain of feature a, any $a\in A$ determines a function fa : U $U\rightarrow V_a$ Va,where Va is the set of va ues of a.

With any $B\subseteq A$, there is an associated equivalence relation $IND(B)$

$$IND(B) = \{(x,y)\in U \mid \forall a\in B, a(x)=a(y)\} \quad (1)$$

In order to classify an object based only on the equivalence class in which it belongs , we need the concept of set approximation. Given an information system, $S=(U,A)$, and a subset of attributes $B\subseteq A$, we would like to approximate a set of objects , X , using only the information contained in B. We define:

B-lower approximation of X:

$$\underline{B}X = \{x \mid [x]_B \subseteq X\} \quad (2)$$

The lower approximation is the set containing all objects for which the equivalence class corresponding to the object is a subset of the set we would like to approximate. This set contain all objects which with certainty belong to the set X.

B-upper approximation of X:

$$\overline{B}X = \{x \mid [x]_B \cap X \neq \phi\} \quad (3)$$

The upper approximation is the set containing all objects for which the intersection of the object's equivalence class and the set we would like to approximate is not the empty set Φ. This set contains all objects which possibly belong to the set X.

B-boundary region of X:

$$BN_B(X) = \overline{B}X - \underline{B}X \quad (4)$$

This set contains the objects that can not be classified as definitely inside X nor definitely outside X.

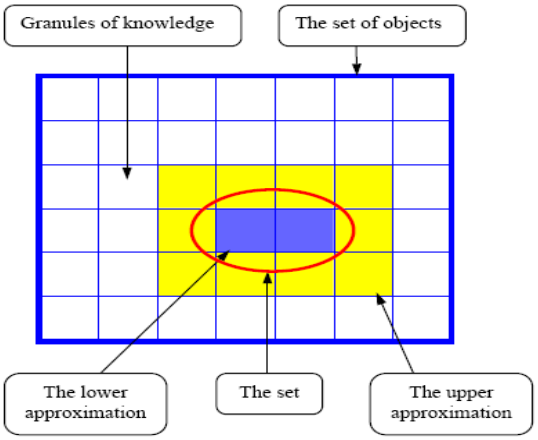

Fig. 1 . The graphical illustration of the set approximations

The goal of feature reduction is to remove redundant fe tures so that the reduced set provides the same quality of classification as the original.A reduct is defined as a subset R of the conditional feature set C such that $\mu_R(D)=\mu_C(D)$. A given decision table may have many feature reducts, the set of all reducts is defined as

$$\text{Red} = \left\{R\subseteq C \mid \mu_R(D)=\mu_C(D) \forall B\subset R, \mu_B(D)\neq\mu_C(D)\right\} \quad (5)$$

In rough set feature reduction,a reduct with minimal cardinality is searched for.An attempt is made to locate a single element of the minimal reduct set $R_{\min}\subseteq \text{Red}$

3 SCALING

In some theories, especially Formal Context Analysis [11], there is a need to transform a many-valued attribute into a number of binary valued attributes. This process is called scaling. The scaling process has no influence on the indiscernibility relations or positive regions of attribute sets. Here we require that the decision attribute not to be changed in the scaling process.

Definition 1

Given a decision table S = (U, C, {d}), the set of scaled attributes of Q ⊆ C is:

$$Q_B = \{(a,v) \mid a\in Q, v\in V_a\} \quad (6)$$

Where (a, v) : U → {0, 1} and $(a,v)(u)=\begin{cases}1 & \text{if } a(u)=v\\ 0 & \text{otherwise}\end{cases}$ (7)

IJSER

4 GENERALIZED DISTRIBUTION TABLE

Generalized Distribution Table (GDT) is a table in which the probabilistic relationships between concepts and instances over discrete domains are represented [12].
Any GDT consists of three components: possible instances, possible generalizations of instances, and probabilistic relationships between possible instances and possible generalizations.

The possible instances, which are represented at the top row of GDT, are defined by all possible combinations of attribute values from a database, and the number of the possible instances is

$$\prod_{i=1}^{m} n_i \qquad (8)$$

Where m is the number of attributes, n is the number of different data values in each attribute.

The possible generalizations for instances, which are represented by the left column of a GDT, are all possible cases of generalization for all possible instances, and the number of the possible generalizations is

$$\left(\prod_{i=1}^{m}\left(n_i+1\right)\right) - \left(\prod_{i=1}^{m} n_i\right) - 1 \qquad (9)$$

A wild card ` * ' denotes the generalization for instances, for simplicity, the wild card will sometimes be omitted in the paper. For example, the generalization $a_0 * c_0$ means that the attribute b is superfluous (irrelevant) for the concept description. In other words, if an attribute b takes values from $\{b_0, b_1\}$ and both $a_0b_0c_0$ and $a_0b_1c_0$ describe the same concept, the attribute b is superfluous, i.e. the concept can be described by a_0c_0 . Therefore, the generalization a_0*c_0 used to describe the set $\{a_0b_0\ c_0, a_0\ b_1\ c_0\}$

The probabilistic relationships between possible instances and possible generalizations, represented by entries G_{ij} of a given GDT, are defined by means of a probabilistic distribution describing the strength of the relationship between every possible instance and every possible generalization. The prior distribution is assumed to be uniform if background knowledge is not available. Thus, it is defined by

$$G_{ij} = p(PI_j \setminus PG_i) = \begin{cases} \frac{1}{N_{PG_i}} & \text{if } PG_i \text{ is a generalization of } PI_j \\ 0 & \text{otherwise} \end{cases} \qquad (10)$$

Where

PI_j is the j th possible instance,

PG_i is the Ith possible generalization,

And N_{PG_i} is the number of the possible instances satisfying the ith possible generalization, that is,

$$N_{PG_i} = \prod_{j}^{m} n_j \qquad (11)$$

Where j = 1,. . . , m, and j # the attribute that is contained by the ith possible generalization (i.e., j just contains the attributes expressed by the wild card) .

Rule Strength
In this paper, the rules are expressed in the following form: $X \rightarrow Y$ with S That is, "if X then Y with strength S". Where:
X: denotes the conjunction of the conditions that a concept must satisfy,
Y: denotes a concept that the rule describes, and
S: is a "measure of strength" of which the rule holds.

The strength of a given rule reflects the incompleteness and uncertainty in the process of rule inducing influenced by both unseen instances and noise. It is defined by

$$S\left(X \rightarrow Y\right) = s(X)\ .\ \left[1 - r\left(X \rightarrow Y\right)\right] \qquad (12)$$

Where:
s(X) : The strength of the generalization X and r : noise rate function .
s(X) : The strength of the generalization X (i.e., the condition of the rule) it represents explicitly the prediction for unseen instances.

$$s(PG_i) = \sum_j p(PI_j \setminus PG_i) = \frac{N_{\text{ins-rel,i}}}{N_{PG_i}} \qquad (13)$$

Where $N_{\text{ins-rel,i}}$ is the number of the observed instances satisfying the ith generalization.

5 THE PROPOSED ALGORITHM

5.1 Algorithm Structure

The main structure of the algorithm is repeating the three steps, namely, scaling, reduction and converting back. Until the new decision table is unpartitionable. In other words, a single group partition reduct (SGPR) of the new decision table should be computed recursively and new values such as 2, 3 . . should be assigned to k in Equation (14). By doing so an optimal partition reduct can be obtained, where the optimal metric is defined by the cardinality sum of new attribute domains [13].

Definition 2
A partition scheme P = $[P_1 \ldots\ldots P_{|C|}]$ of S is called a single group partition scheme (SGPS) if for any i $\in$ {1, . . . , |C|}, $|W_{ai}| = 1$.
Given an SGPS P = $[P_1, \ldots, P_{|C|}]$, for any i $\in$ {1, . . . , |C|}, P_i

essentially divides V_{ai} into two disjoint subsets $V_{a_i}^F$ and $V_{a_i}^G$,

and $P_i(v) = \begin{Bmatrix} v & \text{if } v \in V_{a_i}^F \\ k & \text{if } v \in V_{a_i}^G \end{Bmatrix}$ where $k \in W_{ai}$ (14)

5.2 Algorithm Structure

The pseudo code of the introduced algorithm is listed below:

Algorithm
{Input: A decision table S.**}**
{Output: A set of decision rules**}**
//Initialize. M^i is used for M-relative reduct.
Step 1. $M^1 = \phi$
//The initial partition scheme P^0. In fact $S^{P^0} = S$.
Step 2. $P^0 = \left[P_1^0, ..., P_{|C|}^0\right]$ Where $P_i^0(v_i) = v_i$ for any $i \in \{1, \ldots, |C|\}$ and $v_i \in V_{ai}$;
//Initialize unprocessed attribute-values pairs for each attribute
//Now all attribute-values pairs are unprocessed.
Step 3. For (i = 1; i ≤ |C|; i++) $H_i^0 = \{a_i\}_B$;
//Attack the OSVP-problem through attacking the OSGP-problem recursively.
Step 4. For (i = 1; ; i++) **begin**
//**scaling.**
Step 4.1 compute $S_B^{P^{i-1}}$;
//**Reduction.**
Step 4.2 R^i = an optimal M-relative reduct of $S_B^{P^{i-1}}$ where $M = M^i$;
Step 4.3 $M^{i+1} = M^i$;//Initialize M^{i+1}.
Step 4.4 for (j = 1; j ≤ |C|; j ++) **begin**
//Compute P^i.
Step 4.4.1 $\forall(a_j, v) \notin H_j^{i-1} - R^i, P_j^i(v) = P_j^{i-1}(v)$;
Step 4.4.2 $\forall(a_j, v) \notin H_j^{i-1} - R^i, P_j^i(v) = i$;
Step 4.4.3 $H_i^j = H_j^{i-1} \cap R^i$ //Remove processed attribute-values pairs
//Compute M^{i+1}
Step 4.4.4 if $\left(H_i^j \neq \phi\right)$ $M^{i+1} = M^{i+1} \cup \{(a_j, i)\}$;
End//of for j;
//**Converting back to a "normal" decision table **
Step 4.5 compute S^{P^i} where $P^i = \left[P_1^i, ..., P_{|C|}^i\right]$;
//See if all attribute-values pairs have been processed
Step 4.6 if $H^i = \bigcup_{j=1}^{|C|} H_j^i = \phi$ break; **end** //of for i
Step 5 P = P^i, return P;
Step 6 Create the GDT;
Step 7 Simplify the GDT;
Step 8 Group the generalizations;
Step 9 Rule Selection;

6 Case Study

In this section we will descuss briefly the application of the proposed algorittm to the rotary clinker kiln control. Fig. 2 shows the simplified schame of the kiln.

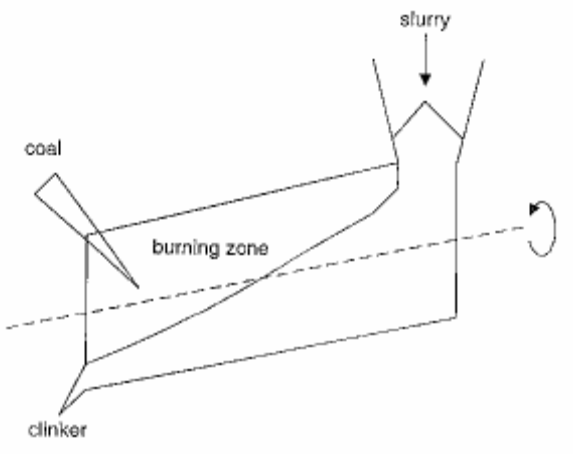

Fig. 2 . Rotary clinker kiln

The aim of the control is to mimic the behavior of the stoker of the kiln. To this end the control algorithm (set of control rules) has been generated from the analysis of the stoker behavior.

The stoker observes the burning zong of the kiln and identifies the state of the kiln by evaluate the following parameters, (condition attributes):

C_1 - burning zone temperature
C_2 - burning zone color
C_3 - clinker granulation in burning zone
C_4 - inside color of the kiln

Values of these parameters range are given below in table 1:

TABLE 1
VALUES OF THE CONDITION ATTRIBUTES

Parameter	values
C_1	{1,2,3,4} where $1 = (1380°C - 1420°C)$ $2 = (1421°C - 1440°C)$ $3 = (1441°C - 1480°C)$ $4 = (1481°C - 1500°C)$
C_2	{1,2,3,4,5}, where 1=scarlet, 2=dark pink, 3=bright pink, 4=decidedly bright pink, 5=rosy white
C_3	{1,2,3,4}, where 1=fines, 2=fines with small lumps, 3=distinct granulation, 4=lumps

IJSER

C_4	{1,2,3}, where 1=distinct dark streaks, 2=indistinct dark streaks, 3=no dark streaks

Note that condition attribution are both quantitative (*burning zone temperature*) and qualitative (*burning zone color, clinker granulation in the burning zone and inside color of the kiln*).

After identification of the kiln state, determined by the condition attributes, the stoker using his knowledge and experience acts accordingly. His control decisions consist in setting values of the following control parameter (decision attribute): *d - coal worm revolutions*. The Values of this parameter range is {1, 2, 3, 4}, where 1=0[*rpm*], 2=15[*rpm*], 3=20[*rpm*], 4=40[*rmp*].

In Table 2 (the decision table) control decisions of the stocker during one shift are given.

TABLE 2
THE DECISION TABLE (PROTOCOL OF STOCKER DECISIONS)

U	C_1	C_2	C_3	C_4	d
x1	3	1	3	2	4
x2	3	2	3	2	3
x3	3	1	3	2	4
x4	4	2	3	2	2
x5	4	2	4	2	2
x6	4	2	4	3	2
x7	4	1	4	3	2
x8	4	1	3	3	2
x9	4	1	3	2	2
x10	4	3	3	2	2
x11	3	2	3	2	4
x12	3	1	3	2	4
x13	3	1	3	2	4
x14	3	3	3	2	3
x15	3	3	2	2	4
x16	3	1	2	2	4
x17	3	2	2	2	4
x18	3	2	3	2	3
x19	3	2	3	2	3
x20	4	2	3	2	2
x21	4	2	4	2	2
x22	4	2	4	3	2
x23	4	1	4	3	2
x24	4	1	3	3	2
x25	4	1	3	2	2

According to the proposed method in the paper, the corresponding rule sets obtained are in Table 3.

TABLE 3
THE CORRESPONDING RULE SETS GENERATED

Rules
if (C_1, 3) *and* (C_4, 1) **then** (d, 4)
if (C_1, 3) *and* (C_4, 2) **then** (d, 4)
if (C_4, 2) **then** (d, 4)
if (C_3, 2) *and* (C_4, 3) **then** (d, 3)
if (C_3, 3) *and* (C_1, 2) **then** (d, 2)

7 CONCLUSION

Rough set theory proved to be a very well suited candidate, beside fuzzy sets, neural networks and other soft computing methods, for intelligent industrial applications. Particularly challenging areas of applications of rough sets in industrial environment are material sciences, intelligent control, machine diagnosis and decision support.

Rough set approach has many advantageous features like, identifies relationships that would not be found using statistical methods, allows both qualitative and quantitative data and offers straightforward interpretation of obtained results

The Proposed method is both efficient for the space and time complexities and tends to give suboptimal results. It helps obtaining small rule sets with good performance for most mechanical machines

ACKNOWLEDGMENT

We thank the President of Salman Bin Abdulaziz University, Deanship of Scientific Research at Salman Bin Abdulaziz University, Dean of College of Science and Humanitarians studies and all the Staff of Department of mathematics for their continuous support, encouraging and for their useful and fruitful discussion. This paper was supported by Salman Bin Abdulaziz University under the grant number 1433/ت /15

REFERENCES

[1] R. Isermann and P. Balle , "Trends in the application of model-based fault detection and diagnosis of technical processes," Control Eng. Practice 5(5), pp. 709-719, 1997

[2] B.S. Yang and K.J. Kim "Application of Dempster-Shafer theory in fault diagnosis of induction motors using vibration and current signals". Mechanical Systems and Signal Processing 20, 403-420, 2006

[3] B.S.Yang, S.K. Jeong, Y.M. Oh, and A.C.C. Tan, "Case-based reasoning with Petri nets for induction motors fault d agnosis", Expert Systems with Applications 27 (2) 301-311, 2004

[4] R.Casimir, E. Boutleux, G. Clerc, and A.Yahoui, "The use of feature selection and nearest neighbors rule for faults diagnosis in induction motors", Engineering Applications of Artificial Intelligence 19, 169-177,2006

[5] A.Widodo, B.S. Yang, and T. Han, "Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors", Expert Systems with Applications 32 (2) 299-312, 2006

[6] B.S.Yang, T. Han, and Z.J. Yin "Fault diagnosis system of induction motors using feature extraction, feature selection and classification algorithm", JSME Int. J. (C) 49 (3) 734-741, 2006

[7] R.Kumar, V.K. Jayaraman, and R.D. Kulkarni, "An SVM classifier incorporating simultaneous noise reduction and feature selection: illustrative case examples", Pattern Recognition 38, 41-49, 2005

[8] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. Advances in Knowledge Discovery and Data Mining. AAAI Press/MIT Press, 1996.

[9] Hossam Abd Elmaksoud Mohamed, "An Algorithm for Mining Decision Rules Based on Decision Network and Rough Set Theory". T.-h. Kim et al. (Eds.): UCMA 2011, Part I, CCIS 150, pp. 44–54, 2011.

[10] Pawlak, Z., Skowron, A.: Rough sets and Boolean reasoning. Information Sciences 177, 41–73 , 2007

[11] B. Ganter, R. Wille (Eds.), Formal Concept Analysis: Mathematical Foundations, Springer-Verlag, New York, 1999.

[12] Z. Weihua , Wei Zhang, Yunqing Fu , An Incomplete Data Analysis Approach using Rough Set Theory , Proceedings of the 2004 international Conference on Intelligent Mechatronics and Automation, Chengdu , China August 2004

[13] Hossam A. Nabwey, "A Probabilistic Rough Set Approach to Rule Discovery". International Journal of Advanced Science and Technology Vol. 30, May, 2011

IJSER