An Improved Algorithm of Extracting Fault Diagnosis Rules Based on Rough Sets

Juanli Li and Zhaojian Yang[*]

College of Mechanical Engineering, Taiyuan University of Technology, Taiyuan, 030024, China

Abstract: Rough set is a new data analysis theory. It is often used to deal with fuzzy and uncertain problems. Attribution reduction is the key step in obtaining the knowledge by utilizing rough set. An improved heuristic reduction algorithm of attribute significance is proposed in the study based on analyzing the classic knowledge acquisition method of rough set theory. The algorithm corrects the discernibility matrix and redefines the calculation method of attribute importance. Then it fuses the both, gets the core by using the revised method of discernibility matrix and calculates the attribute importance by using the weighted method and then the algorithm is applied to extract the rules of the hoist fault diagnosis. Verified by the experiment, using the algorithm, it can excavate high reliability diagnosis rules from existing history diagnosis knowledge and expert knowledge. This method can provide reasonable basis for fault diagnosis.

Keywords: Attribute importance, discernibility matrix, fault diagnosis, rules extraction, rough set.

1. INTRODUCTION

Using reasoning and judgment method, Mechanical equipment fault diagnosis system can diagnose the fault accurately and effectively, the key factor to form a fault diagnosis system is a complete knowledge base [1]. Knowledge acquisition is the process of extracting useful knowledge and experience from the specific domain knowledge source. The traditional way of obtaining knowledge is that the experts analyze and sum up the experienced knowledge, then the diagnosis system developers summarize them into design rules and set up the corresponding knowledge base. There are the following questions to get knowledge through this way: First of all, because every individual expert is different, leading to a domain expert knowledge has strong randomness. Secondly, the subjective knowledge has bias and errors and is derived from the lack of scientific theory basis. Combined with the complexity and difficulty of the communication with the domain experts, the knowledge obtained is often incomplete. Thus knowledge acquisition problem is recognized as the "bottleneck" problem in fault diagnosis system [2, 3].

Because the rough set theory has a superior ability to extract the uncertainty information, it has been widely applied in fault diagnosis for mechanical equipment, especially it has made great achievements in extracting the rules [4-8]. Attribute reduction is one of the main contents of rough set theory. Commonly used method of attribute reduction contained the reduction algorithm based on discernibility matrix; attribute importance and collection approximation quality etc. Ref. [9] presented a large rotating machinery fault diagnosis model based on rough set. Based

on the original information, the decision table was constructed, and attribute was reduced using discernibility matrix, and then the diagnosis rules were obtained. Ref. [10] studied the method of fuzzy rules extraction based on rough set in depth, proposed a new heuristic reduction algorithm based on fuzzy membership. It corrected the fuzzy membership function, then it was applied into the chemical process fault diagnosis and the validity of the method was verified. The reduction algorithms have their own drawbacks and advantages. So, in this study, the existing classic rough set algorithm was improved and fused and a new fault diagnosis knowledge acquisition method was put forward based on the theory of rough set. It can find useful knowledge patterns from data and generate the diagnostic rules. These rules provides a strong theoretical and data support for fault diagnosis.

2. METHOD OF FAULT DIAGNOSIS KNOWLEDGE ACQUISITION BASED ON ROUGH SET

Rough set theory is a kind of tools with inexact research and uncertain knowledge. It was initially put forward by Poland scientist Z. Pawlak in 1982. It does not need any auxiliary information, but is only based on the information provided by data itself. Rough set theory can simplify the data and obtain the minimum expression of knowledge under the premise of retaining the key information, then generate decision-making rules and find implicit knowledge from the given data set. So rough sets describes or processes the uncertainty problem which can be said to be more objective [11, 12]. The basic idea of rough set is that the domain is divided based on the existing problems of the given knowledge and then the extent of the collection that belongs is divided for the division of each part of the set. That is, the decision or classification rules are exported by knowledge reduction under the same classified premise.

Based on rough set theory, the knowledge acquisition model of fault diagnosis rule is built and is shown in Fig. (**1**).

Firstly, discrete historical diagnostic data sets and field experience knowledge create the decision table by the sample database, then removes the redundant rules and attributes by using the algorithm of attribute reduction and values reduction in rough set. Thereby generating the diagnostic rules and then evaluates the rule and metric its importance. Finally, puts the high reliability rules into the rule knowledge base. In the process of fault diagnosis, rules knowledge acquisition and attribute reduction are critical. The reduction results decide the confidence of gaining rules.
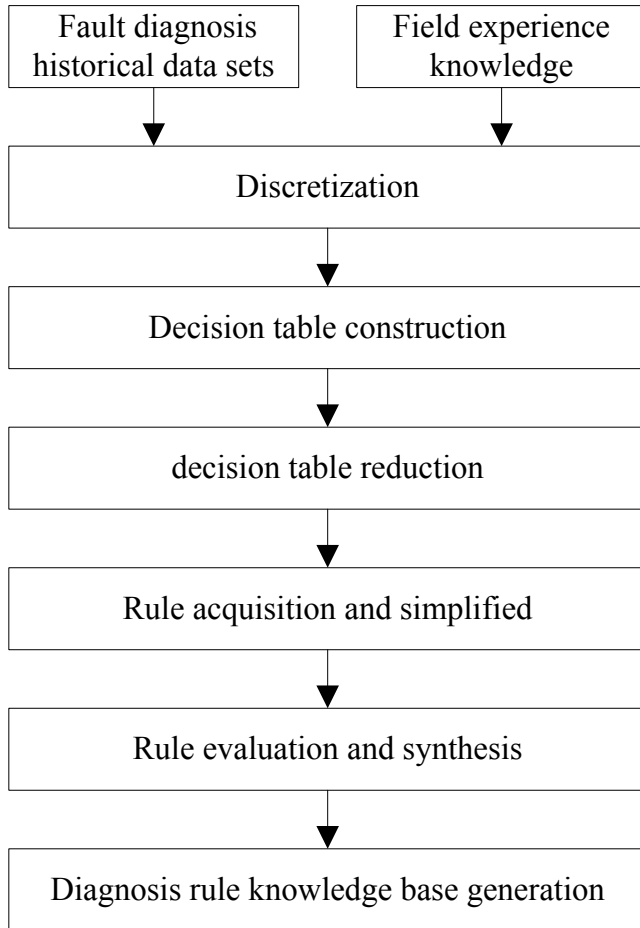

Fig. (1). The fault diagnosis knowledge acquisition model.

3. ATTRIBUTE REDUCTION BASED ON ROUGH SETS MODEL

In rough set system, the attributes are not equally important, some are even redundant. Attribute reduction is to delete the irrelevant or unimportant attributes under the conditions that does not change the system's classification ability, thereby improving the clarity of potential knowledge in the system.

3.1. Method of Attribute Reduction Based on Iscernibility Matrix

In decision table $S = (U, C \cup D, V, \rho)$, if the same equivalence class record in U / C has the same decision value, then any record in the equivalence class is a deterministic record, else is an uncertainty one. If all records of the decision table are deterministic records in S, then S

is called as the consistent decision table, else it is called a non-consistent one.

In order to formally describe the condition attribute reduction and core computation, Polish mathematician A. Skowrno proposed the method of attribute reduction using discernibility matrix. The method is suitable for the reduction of consistent decision tables [13]. The discernibility matrix is a $n \times n$ matrix, n is the number of objects in the decision table, the element c_{ij} of the matrix represents object u_i and u_j has the different attribute values, the definition of descernibility matrix as follows:

Let

$$S = (U, C \cup D, V, \rho),$$

discernibility matrix

$$C(S) = \left[c_{ij} \right]_{n \times n},$$

$$c_{ij} = \begin{cases} c \in C : \rho(u_i, c) \neq \rho(u_j, c) \\ \Phi \end{cases} \tag{1}$$

Core can be defined as a set of matrix entries which only has one element in discernibility matrix as follows:

$$CORE(C) = \left\{ c \in C, \delta(u_1, u_2) = \{c\}, \exists (u_1, u_2) \in U \right\}$$

where $\delta(u_1, u_2)$ represents object u_i and u_j has the different attribute values.

It has the advantage that expression of knowledge is intuitive and easy to understand using discernibility matrix, especially it can easily calculate the reduction and core, but also has the defect that calculation results are not unique.

3.2. The Heuristic Reduction Algorithm Based on Attributes Significance

In decision table, the importance of different attribute sets is different. If removed, the property has a great influence on classification. It has the importance of the properties of high or low [14].

Respectively, defines the attribute significance from the view of approximation set and information theory. The important degree of attributes can be classified by two calculation methods based on the dependency degree and entropy.

- The attribute importance based on the dependency degree.

It can be defined by using the positive region. In decision table S , let condition attributes set is C, decision attribute set is D, $R \subset C$, $a \in C - R$ is a attribute ,based on the definition of dependency degree, the importance of attribute a can be defined as follows:

$$SGF_d(a, R, D) = \gamma(R \cup \{a\}, D) - \gamma(R, D) \tag{2}$$

$$SGF_d(a, R, D) > 0$$

Above equation shows that the attribute a influences the classification ability of R to D and the larger the value, the stronger the ability to influence.

- The attribute importance based on the entropy

Suppose $S = (U, C \cup D, V, \rho)$ is a decision table, the division of C and D is X and Y on U, $X = \{X_1, X_2, ..., X_m\}$, $Y = \{Y_1, Y_2, ..., Y_n\}$, according to the definition of probability distribution, joint probability distribution of knowledge, the entropy and conditional entropy of information theory, the entropy $H(D)$ of knowledge D, the conditional entropy $H(D|C)$ of knowledge D relative to knowledge C and the mutual information $I(C, D)$ of knowledge C and knowledge D are defined as follows:

$$H(D) = -\sum_{i=1}^{n} p(Y_i) \log_2 p(Y_i)$$

$$H(D|C) = -\sum_{i=1}^{m} p(X_i) \sum_{j=1}^{n} p(Y_j | X_i) \log_2 p(Y_j | X_i)$$

$$I(C, D) = H(D) - H(D|C)$$

$$p(X_i) = CARD(X_i) / CARD(U), i = 1, 2, ..., m,$$

$$p(Y_i) = CARD(Y_i) / CARD(U), i = 1, 2, ..., n$$

In information theory, the size of the information content is measured by the size of the eliminated uncertainty. The attribute importance of condition, attributes relative to the decision attributes can be measured by adding the size of the mutual information which caused by a condition attribute in the attribute reduction of decision table.

In decision table S, for the known attribute set $R \subset C$, after adding a property $a \in C - R$, the increment of mutual information can be defined as

$$\Delta I = I(R \cup \{a\}; D) - I(R; D)$$
$$= H(D|R) - H(R \cup \{a\})$$

So, the importance $SGF_s(a, R, D)$ of attribute $a \in C - R$ can be defined as follow:

$$SGF_s(a, R, D) = H(D|R) - H(R \cup \{a\}) \qquad (3)$$

The bigger the $SGF_s(a, R, D)$ value indicates that the attribute a is more important on the Decision D in the condition of Known as the R value. For decision attribute sets D R, $H(D)$ and $H(D|R)$ are certain, therefore in the conditions of known R, the most important conditional property a can be described as follows:

$$a \in C - R, \forall b \in C - R,$$
$$H(D|R \cup \{a\}) \leq H(D|R \cup \{b\})$$

3.3. The Improved Heuristic Reduction Algorithm Based on Attributes Importance

In order to obtain the better reduction result, based on the detailed analysis of the above two kinds of classical rough set methods, combining the two together, this study computes the core by using corrective discernibility matrix and gets the attribute importance by using the weighted attribute importance algorithm.

- The discernibility matrix modificatory

Using the discernibility matrix in attribute reduction can effectively reduce the amount of calculation, easily calculate the core and reduction of knowledge system, improve the efficiency of solving kernel. But because of lack of evidence when the minimal complete reduction set is selected, the calculation results are not exclusive and not the best reduction. In addition, attribute reduction method of discernibility matrix is suitable for consistent attribute set. In fact, inconsistent data objects exists in the fault diagnosis data as always. In this case, the method of discernibility matrix can not get the right core. So it is necessary to modify the original discernibility matrix.

The definition of discernibility matrix for consistent attribute set and inconsistent one is as follows:

Let $S = (U, C \cup D, V, \rho)$, discernibility matrix $C(S) = [c_{ij}]_{n \times n}$,

$$c_{ij} = \begin{cases} c \in C : \rho(u_i, c) \neq \rho(u_j, c) \quad \text{when} \\ d_i \neq d_j, u_i \in POS_R(U), u_j \in POS_R(U) \\ \\ c \in C : \rho(u_i, c) \neq \rho(u_j, c) \quad \text{when} \\ u_i \in POS_R(U), u_j \in NEG_R(U) \\ \Phi \end{cases} \qquad (4)$$

- The redefinition of attribute importance

The existing definition of attribute measure methods is not complete, for example:

Let $S = (U, C \cup D, V, \rho)$, $CARD(U) = 16$

condition attributes $c_1, c_2 \in C$, decision attribute is d, the decision table is shown as Table **1**, then Let $V_{\{c_1\}} = \{0, 1, 2, 3\}$, $V_{\{c_1\}} = \{0, 1, 2, 3\}$, $V_{\{d\}} = \{0, 1\}$, the decision table can converse into the follow sub decision table $S_1 = (U, \{c_1\} \cup \{d\}, V, \rho)$ and $S_2 = (U, \{c_2\} \cup \{d\}, V, \rho)$, as shown in Tables **2** and **3**:

According to the definition of attribute importance based on the dependence degree, the importance of c_1 and c_2 is:

$$SGF_d(c_1, R, D) = \gamma(R \cup \{c_1\}, D) - \gamma(R, D)$$

=4/16-0=0.25

$$SGF_d\left(c_2, R, D\right) = \gamma\left(R \cup \left\{c_2\right\}, D\right) - \gamma\left(R, D\right) = 0$$

Table 1. Decision table.

U/C	c_1	c_2	d
u_1	0	1	0
u_2	0	0	0
u_3	1	0	0
u_4	1	1	1
u_5	2	0	1
u_6	0	0	0
u_7	2	0	0
u_8	3	0	0
u_9	2	1	1
u_{10}	0	0	0
u_{11}	1	1	1
u_{12}	2	1	1
u_{13}	3	0	0
u_{14}	2	1	1
u_{15}	1	1	1
u_{16}	2	1	0

Table 2. Sub decision table S_1

	$c_1=0$	$c_1=1$	$c_1=2$	$c_1=3$	$\sum$
d=0	4/16	1/16	1/16	2/16	8/16
d=1	0	3/16	3/16	2/16	8/16
$\sum$	4/16	4/16	4/16	4/16	16/16

Table 3. Sub decision table S_2

	$c_2=0$	$c_2=1$	$\sum$
d=0	7/16	1/16	8/16
d=1	1/16	7/16	8/16
$\sum$	8/16	8/16	16/16

According to the definition of attribute importance based on the information entropy, the importance of c_1 and c_2 is:

$$SGF_s\left(c_1, R, D\right) = H\left(D \mid R\right) - H\left(R \cup \left\{c_1\right\}\right)$$
$$= H\left(D\right) - H\left(D \mid \left\{c_1\right\}\right)$$

$$= 2 \times 8 / 16 \cdot \log_2\left(16 / 8\right) -$$
$$2 \times 4 / 16 \cdot \left(1 / 4 \log_2 4 + 3 / 4 \log_2 4 / 3\right) = 0.3444$$
$$- 4 / 16 \cdot \left(1 / 2 \log_2 2 + 1 / 2 \log_2 2\right)$$

$$SGF_s\left(c_1, R, D\right) = 0.4564$$

As you can see from the above results, according to the calculation method of attribute importance based on dependence, the important degree of attribute c_1 is higher than c_2. However, according to the standard of attribute importance based on information entropy, the important degree of attribute c_2 is higher than, c_1, this is clearly a contradiction. The cause of this problem in the final analysis is the algorithm characteristics. The former only considers the set of certain elements, but probability distribution information of boundary element is not considered. The latter considers the information which is provided by uncertain elements in boundary domain and does not consider the importance of deterministic elements. Therefore, it is not sufficient to calculate only by these two methods the importance of attribute. So, this study analyzes the advantages and disadvantages of both and proposed the following algorithm:

In a complete information system, decision table $S = \left(U, C \cup D, V, \rho\right)$ should not contain uncertainty elements, namely, $R^-\left(D\right) = U$. By rough set theory, the $R-$ boundary domain of set D is:

$$B_R\left(D\right) = R^-\left(D\right) - R_-\left(D\right) = U - R_-\left(D\right)$$

The ratio of deterministic element and uncertainty element of attribute R are defined as:

$$\omega_1\left(R, D\right) = CARD\left(POS_R\left(D\right)\right) / CARD\left(U\right)$$

$$\omega_2\left(R, D\right) = CARD\left(B_R\left(D\right)\right) / CARD\left(U\right)$$

From the above three equations, $\omega_2\left(R, D\right)$ can be simplified as:

$$\omega_2\left(R, D\right) = 1 - \omega_1\left(R, D\right)$$

Synthesize the above two kinds of calculation method of attribute importance, because $0 \le \gamma\left(R, D\right) \le 1$, $0 \le H\left(D \mid R\right) \le \log_2 n$, the value of $\gamma\left(R, D\right)$ is bigger, the better, the $H\left(D \mid R\right)$ is as small as possible, standardized the $H\left(D \mid R\right)$ as follows:

$$H'\left(D \mid R\right) = 1 - \frac{H\left(D \mid R\right)}{\log_2 n},$$

After standardization, constructing weighted function as follows:

$$STD(D,R) = \omega_1(R,D) \cdot \gamma(R,D) + \omega_2(R,D) \cdot H'(D \mid R)$$

So, for any attribute $a \in C - R$, the importance degree can be defined as follows:

$$SGF_c(c_1, R, D) =$$
$$\omega_1(R \cup \{c_1\}, D) \cdot (\gamma(R \cup \{c_1\}, D) - \gamma(R, D)) + \quad (5)$$
$$\omega_2(R \cup \{c_1\}, D) \cdot (1 - H(D \mid R \cup \{c_1\}) / \log_2 n)$$

According to the definition, the importance degree of c_1 and c_2 are:

$$\omega_1(c_1, D) = 0.25 \ , \quad \omega_2(c_1, D) = 0.75 \ , \quad \omega_1(c_2, D) = 0 \ ,$$
$$\omega_2(c_2, D) = 1$$

$$SGF_c(c_1, R, D)$$
$$= \omega_1(R \cup \{c_1\}, D) \cdot (\gamma(R \cup \{c_1\}, D) - \gamma(R, D)) +$$
$$\omega_2(R \cup \{c_1\}, D) \cdot (1 - H(D \mid R \cup \{c_1\}) / \log_2 n)$$
$$= \omega_1(c_1, D) \cdot \gamma(c_1, D) + \omega_2(c_1, D) \cdot H'(D \mid c_1)$$
$$= 0.25 \times 0.25 + 0.75 \times 0.8361 = 0.69$$

$$SGF_c(c_2, R, D) = 0.8641$$

The above calculation results shows that the importance of attribute c_1 is higher than c_2 according to the new definition of attribute importance.

- The heuristic reduction algorithm based on the improved attribute significance

According to the modified definition of discernibility matrix and attribute importance algorithm proposed above, the study proposes a heuristic reduction algorithm based on improved attribute importance.

Input: decision table S, the properties set $CUSTOM(C)$ that users are especially concerned.

Output: Reduction set $RED(C)$.

Steps:

1) Constructing the discernibility matrix according to the formula (4)

2) Calculating correlation coefficient $\gamma(C,D)$ of condition attributes and decision attributes in decision table

3) Calculating core $CORE(C)$

4) $RED(C) = CORE(C) \cup CUSTOM(C)$

5) $C' = C - RED(C)$

6) According to the formula (5), computing the attribute importance of each attribute and sorting by it

7) Calculating $\gamma(RED(C), D)$ and $\gamma(C, D)$

8) While $(\gamma(RED(C), D) \neq \gamma(C, D))$ do

9) {Select the most important attribute c_i

10) $RED(C) = RED(C) + \{c_i\}, C' = C' - \{c_i\}$

11) Calculating $\gamma(RED(C), D)$ }

12) $M = |RED(U)|$

13) For $i = 0$ to $M - 1$ do

14) {If (c_i is not in $CORE(C)$) then

15) { $RED(C) = RED(C) - \{c_i\}$

16) Calculating $\gamma(RED(C), D)$

17) If $(\gamma(RED(C), D) \neq \gamma(C, D))$

Then $RED(C) = RED(C) + \{c_i\}$ }}

4. INSTANCES OF FAULT DIAGNOSTIC KNOWLEDGE EXTRACTION

This study takes the mine hoist braking system fault as an example to verify the effectiveness of the algorithm.

Hoist is large complex mining machinery; it has lots of fault feature parameters. In the study, gate 2 of the hoist brake system is taken as the object of study. The common fault feature parameters are analyzed by selecting the five representative characteristic parameters as samples for testing, namely, the oil pressure of hydraulic station, braking torque, free time, brake shoe clearance and spring force. The fault types can be predicted including the long free time, emergency brake failure, long braking time, over wind, slipping and open gate failure. Simulating the six faults and collecting 80 groups of the original sample data as shown in Table **4**.

Discretize and express the six types fault collection with $D_i (i = 1, 2, ... 6)$ and the failure attribute set using $C_i (i = 1, 2, ... 5)$. Among them: D_1 - long free time, D_2 - emergency brake failure, D_3 - long braking time, D_4 - overwind, D_5 - slipping, D_6 - open gate failure, C_1 - Maximum brake oil pressure is lower than 6.3MPa or minimum value is greater than 0.5MPa, C_2 - Brake shoe clearance is greater than 2mm, C_3 - The brake torque is less than 40kN·m, C_4 - Free time is greater than 0.3s, C_5 - Spring force is less than 26kN. Fault attribute has a value of 1 or 0, among them, the attribute value is 1 means failure has occurred and 0 means no failure has occurred. Take H as index number, K as sample amount; establish the braking system fault decision table as shown in Table **5**.

Table 4. The original data sample table.

	Oil Pressure (mpa)	Braking Torque (kn·m)	Free Time (s)	Brake Shoe Clearance (mm)	Spring Force (kn)
long free time	6.83	47	0.44	3.55	28.32
over wind	0.64	48	0.23	1.58	29.77
sliping	0.54	54	0.32	4.21	28.67
open gate failure	0.57	53	0.18	3.09	28.22
emergency brake failure	6.82	37	0.17	1.65	23.32
long braking time	6.76	39	0.22	1.79	25.33
…	…	…	…	…	…
sliping	0.58	49	0,39	2.76	28.43
open gate failure	0.62	50	0.13	3.28	27.44
long braking time	6.73	37	0.29	1.99	24.32
over wind	0.58	51	0.19	1.77	28.64
sliping	0.55	51	0.43	3.98	27.77

Table 5. The fault diagnosis decision table.

No		Condition Attributes					Decision Attributes
H	K	C1	C2	C3	C4	C5	D
1	18	0	0	1	1	0	d1
2	20	1	0	1	0	0	d2
3	21	1	0	1	1	0	d3
4	16	1	0	0	1	0	d4
5	1	0	1	0	0	1	d5
6	1	0	1	0	0	1	d6
7	2	0	1	1	0	1	d1
8	1	1	0	0	0	0	d2

Build the discernibility matrix, such as in Table **6**. Based on the discernibility matrix, calculating the core attributes, the results are $\{C_1, C_3, C_4\}$. Computing respectively the attribute importance by using the algorithm of the above section, we can get the results: $SGF_c(c_1, R, D) = 0.872$, $SGF_c(c_3, R, D) = 0.869$, $SGF_c(c_4, R, D) = 0.697$.

Therefore, the attributes with higher important degree are C_1 and C_3, then, the rules are shown in Table **7**.

The redundant attributes are eliminated effectively in the process of the reduction and the most simple decision rules are formed eventually. The rule set shows that:

- A too long air travel time fault occurs when the braking torque is less than 40 kN·m or the air travel time is greater than 0.3s.

- Over wind fault appears when residual pressure is higher for braking or brake torque is less.

- Slide problem occurs when the residual pressure is too high and the brake torque is too small and empty travel time is greater than 0.3s.

- Breaking open failure appears when the maximum oil pressure does not reach the standard and empty travel time is greater than 0.3s.

Table 6. The table of clear matrix.

	1	2	3	4	5	6	7	8
1		$C_1 C_4$	C_1	$C_1 C_3$	$C_2 C_3 C_4 C_5$	$C_2 C_3 C_4 C_5$	$C_2 C_4 C_5$	$C_1 C_3 C_4$
2			C_4	$C_3 C_4$	$C_1 C_2 C_3 C_5$	$C_1 C_2 C_3 C_5$	$C_1 C_2$	C_3
3				C_3	$C_1 C_2 C_3 C_4 C_5$	$C_1 C_2 3 C_3 C_4 C_5$	$C_1 C_2 C_4 C_5$	$C_3 C_4$
4					$C_1 C_2 C_4 C_5$	$C_1 C_2 C_4 C_5$	$C_1 C_2 C_3 C_4 C_5$	C_4
5							C_3	$C_1 C_2 C_5$
6							C_3	$C_1 C_2 C_5$
7								$C_1 C_2 C_3 C_5$
8								

Table 7.   Reduction rule set.

No.	Condition Attributes			Decision Attributes
	C_1	C_3	C_4	
1	0	1	1	d_1
2	1	1	0	d_2
3	1	1	1	d_3
4	1	0	1	d_4
5	0	0	0	d_5
6	0	0	0	d_6
7	0	1	0	d_1
8	1	0	0	d_2

The first three simplest rule sets show that in the braking process due to high residual pressure and brake fault, leads to braking slowly, which can not be normal parking. The last rule shows that because of a shortage of oil or air travel time it takes too long to brake while in drive., leading the gate to be unopened, thus causing the wear of brake discs and affecting speed.

CONCLUSION

- The study sets up the model of fault diagnosis knowledge acquisition based on rough set and describes the whole process of fault diagnosis knowledge acquisition.

- The author analyzed in detailed the two kinds of classic rough set reduction method which are based on the discernibility matrix and the attribute importance. Pointed out some drawbacks in algorithm and then proposed an improved heuristic reduction algorithm of attribute importance. The algorithm firstly corrected the discernibility matrix, then calculated the properties' core by using the matrix and then constructed the weighted function to calculate the attribute importance and completed the attribute reduction.

- At the end of the article, the improved attribute importance of rough set knowledge acquisition method was verified by the extraction instance of hoist diagnosis knowledge; the results show the validity and correctness of the method.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

REFERENCES

[1]   A.C.M. Fong and S.C. Hui, "An intelligent machine fault diagnosis system," *Computing and Control Engineering Journal*, vol.12, pp. 217-223, 2001.

[2]   Z. Zhao, X.F. Hu, D.k. He and F.L. Wang, "Knowledge acquisition of fault diagnosis expert system of hydraulic tube tester," *Journal of Northeastern University,* vol. 29, pp.1677-1680, 2008.

[3]   J.Q. Xuan, Q.D. Li and J.H. Jiang, "Knowledge acquisition techniques of fault diagnosis system based on test event graph," *Journal of Shanghai Jiaotong University*, vol. 45, pp.179-183, 2011.

[4]   X.Z. Wang, E.C.C. Tang and S.Y. Zhao, "Learning fuzzy rules from fuzzy samples based on rough set technique," *Information Sciences*, vol. 177, pp. 4493-4514, 2008.

[5]   M. Inuiguchi, and T. Miyajima, "Rough set based rule induction from two decision tables," *European Journal of Operational Research*, vol. 191, pp.1540-1553, 2007.

[6]   Y.H. Qian, and J.Y. Liang, "Positive approximation and rule extracting in incomplete information systems," *International Journal of Computer Science and Knowledge Engineering*, vol. 2, pp. 51-63, 2008.

[7]   Y. Leung, and D.Y. Li, "Maximal consistent block technique for rule acquisition in incomplete information systems," *Information Sciences*, vol.153, pp. 85-106, 2003.

[8]   Z. Zheng and G.Y. Wang, "RRIA: A rough set and rule tree based incremental knowledge acquisition algorithm," *Fundamenta Informaticae*, vol. 59, pp. 299-313, 2004.

[9]   H.J. Wang, J.M. Zhang and X.L. Xu, "Rules acquisition based on rough set for rotating water injection machinery fault diagnosis," *Chinese Journal of Mechanical Engineering*, vol. 42, pp. 135-138, 2006.

[10]   Z.Q. Geng and Q.X. Zhu, "Rough set based fuzzy rule acquisition and its application for fault diagnosis in petrochemical process *Industrial and Engineering Chemistry Research*, vol. 48, pp. 827-836, 2009.

[11]   M. Liu, M.W. Shao, W.X. Zhang and C. Wu, "Reduction method for concept lattices based on rough set theory and its application," *Computers and Mathematics with Applications*, vol. 53, pp. 1390-1412, 2007.

[12]   D. Xiao, J.F. Zhang, and S.S. Hu, "Generalization rough set theory," *Journal of Donghua University*, vol. 25, pp. 654-658, 2008.

[13]   X.B. Zheng, P. Li, J.Q. Sun and K.L. "Method of determining avionics system's weighting coefficient based on rough set theory," *Transaction of Beijing Institute of Technology*, vol. 27, pp. 609-612, 2007.

[14]   M.H. Shi, and C.L. Zhou, "Approach to knowledge discovery for capturing experts' reasoning based on rough set theory," *Journal of Computational Information Systems*, vol. 1, pp. 887-895, 2005.