

## Transformer Fault Diagnosis Method Based on Rough Set and Generalized Distribution Table

Shaaban M. Shaaban\*, Hossam A. Nabwey

Basic Engineering Science Department, Faculty of Engineering , Menofia University, Shebin Elkom, Egypt

\* Corresponding author's Email: [shabaan27@gmail.com](mailto:shabaan27@gmail.com)

---

**Abstract:** Transformers are considered as significant equipments in electrical power systems, once failure ,the economic operation will be lost. To overcome this difficulty and to maintain economic operation of facilities, diverse diagnosis methods are developed to implement fault forecasting. According to intelligent complementary ideas, a fault diagnosis is proposed when there is a missing failure symptom of transformer. The core of the proposed approach is a soft hybrid induction system called the Generalized Distribution Table and Rough Set System (GDT-RS) to discover classification rules. The system is based on a combination of Generalized Distribution Table (GDT) and the Rough Set methodologies. The proposed approach is applied into transformer fault diagnosis and the results indicate that it is very effective and accurate.

**Keywords:** Rough set theory; Generalized Distribution Table (GDT); Fault diagnosis; Transformer

---

### 1. Introduction

High-safety and high-capacity to deliver electric power energy are considered the basic requirements for modern power system. To achieve this objective, the scale of the electric power systems becomes bigger and bigger, the construction of the power delivery network becomes more and more complicated, which ,to a certain degree ,enlarges the scope of breakdown. Transformer is an important apparatus and its operating condition directly affects the operation of the whole electrical system .In order to avoid immense economic loss caused by the breakdown of transformer and to define the failure type, location and its nature,it is significant for the electrical system to diagnose the potential failure of transformer as early as possible.Exa-mining operating condition and failure diagnosis of transfor-mer is a very complicated problem. Many factors such as transformer capacity, voltage grade, performance, working surrounding conditions, loading records and its manufacturer data, di-

rectly affect the results of diagnoses. Frequently there are also random indistinctness and indefiniteness about the cause, phenomenon and mechanism of transformer faults. The complexity of transformer failure diagnosis problem inspires the application of artificial intelligence and mathematical techniques. More complexities arise from efficient and accurate diagnose especially the problem of correct diagnose when there is a loss of failure information or missing failure symptoms of transformer. Many artificial intelligence techniques such as neural network [1], wavelet analysis [2], gray clustering [3], decision tree [4], Petri network [5], information fusion [6] have been applied to transformer diagnosis and have produced some useful results. However, transformers are complex systems with uncertain factors and information, and these methods have different shortages. For example, Petri network puts domain knowledge into a series of producing rules,which can solve fault diagnosis problems. But when new fault or new information is coming, it will lead to matching collision and combination blast

because of the slow speed of Petri network resulted from vast rules.

Classical rough set theory developed by Professor Z. Pawlak in 1982 has made a great success in knowledge acquisition in recent years [7]. In Rough set theory, knowledge is represented in the form of information systems. An information system is a data set represented in a table called decision table [8]. Each row in the table represents an object for a case or an event. Each column in the table represents an attribute, for instance a variable, an observation or a property. To each object, there are some attribute values assigned.

The main drawback of rough set theory is its dependence on complete information systems that the decision table to be processed must be complete and all object values must be known [9]. However, this is not the case in real-life applications. Due to many factors such as measurement errors, miscomprehension, access limitation and misoperation in register information, systems with missing values often occur in knowledge acquisition. In which conditions, where the decision tables are incompletely specified, they are called incomplete information systems [10].

This paper aims to introduce a rough set approach to discover diagnosis rules of transformer failure when there is a missing failure symptom of transformer. The decision table is built by using the observed or surveyed symptoms of the transformer failure as conditional attributes of failure classification and takes the actually existing failure (circle short circuit) as decision attribute as shown in Table 1. The core of the approach is a soft hybrid induction system called the Generalized Distribution Table and Rough Set System (GDT-RS) for discovering classification rules. The system is based on a combination of Generalized Distribution Table (GDT) and the Rough Set methodologies.

## 2. Rough Set and missing Attribute Values

Missing attribute values are commonly existing in real world data set due to the data collecting process, redundant diagnosis tests, unknown data and so on. Since the main concern about learning from examples, and an example with a missing decision value is useless [11], it will be assumed that only attribute values may be missing. Discarding all data containing the missing attribute values cannot fully preserve the characteristics of the original system. Therefore, in data analysis two main strategies are used to deal with missing attribute values in decision tables.

*The former strategy* is based on conversion of in-

complete data sets into complete data sets and then acquiring knowledge. The process is carried-out via a technique called completeness of data set. Multiple approaches of filling-in the missing attribute values are introduced [12,13], such as selecting the “most common attribute value”, the “concept most common attribute value”, “assigning all possible values of the attribute restricted to the given concept”, “ignoring examples with unknown attribute values”, “treating missing attribute values as special values”, “event covering method” and so on. In this strategy conversion of incomplete data sets into complete data sets is a preprocessing to the main process of data mining.

*In the later strategy*, knowledge is acquired from incomplete data sets taking into account that some attribute values are missing. The original data sets are not converted into complete data sets. The later strategy is exemplified by the C4.5 approach to missing attribute values [14] or by a modified LEM2 algorithm [15,16]. In both algorithms original data sets with missing attribute values are not preprocessed.

This paper will concentrate on the later strategy used for rule induction. It will be assumed that the rule sets are induced from the original data sets. The next basic assumption is that there are three approaches to missing attribute values [17]:

*The first approach* is that an attribute value, for a specific case, is lost. For example, originally the attribute value was known; however, due to a variety of reasons, currently the value is not available. Maybe it had been recorded but later it would be erased.

*The second approach* is that an attribute value was not relevant, the case was decided to be a member of some concepts, i.e., was classified, or diagnosed, in spite of the fact that some attribute values were not known. Since such missing attribute values do not matter for the final outcome, we will call them “do not care” conditions.

*The third approach* is a partial “do not care” condition; we assume that the missing attribute value belongs to the set of typical attribute values for all cases from the same concept. Such a missing attribute value will be called an attribute-concept value. Calling it concept “do not care” condition would perhaps be better, but this name is too long in the sequel that it is assumed that all decision values are specified. Also, all missing attribute values are denoted either by “?” or by “\*”, or by “-”, lost values will be denoted by “?”, “do not care” conditions will be denoted by “\*”, and attribute-concept value will be denoted by “-”. Additionally, it is assumed that for each case at least one

Table 1 An incompletely specified decision table

Transformer	Attributes			Decision
	O	E	U	D
1	$a_1$	$b_1$	$c_1$	yes
2	-	$b_0$	$c_0$	yes
3	$a_1$	$b_0$	$c_1$	no
4	$a_1$	$b_0$	*	yes
5	$a_0$	$b_0$	$c_0$	no
6	$a_0$	?	-	no
7	$a_1$	?	$c_1$	yes
8	$a_1$	$b_1$	$c_0$	yes

attribute value is specified.

An example of an incompletely specified table is presented in Table 1. Obviously, in rough set theory any decision table defines a function  $\rho$  that maps the set of ordered pairs (case, attribute) into the set of all  $v$  alues [18]. For example, in Table1,  $\rho(1, O) = a_1$ .

Where:

$a_1 = b_1 = c_1$  =Symptom failure,  $a_0 = b_0 = c_0$  =not appearing

O: Over current protection.

E: Exceeding of winding insulation resistance.

U: Unbalance of three-phase winding direct current resistance.

D: Circle short circuit.

Rough set theory [19] is based on the idea of an indiscernibility relation. The indiscernibility relation  $IND(B)$  is an equivalence relation. Equivalence classes of  $IND(B)$  are called elementary sets of  $B$  and are denoted by  $[x]_B$ . The indiscernibility relation  $IND(B)$  may be computed by using the idea of blocks of attribute-value pairs. Let  $a$  be an attribute and let  $v$  be a value of  $a$  for some cases. For complete decision tables if  $t = (a, v)$  is an attribute-value pair then a block of  $t$ , denoted  $[t]$ , is a set of all cases from  $U$  that for attribute  $a$  has value  $v$ .

For incomplete decision tables, the definition of a block of an attribute-value pair must be modified as follows:

- If for an attribute  $\mathbf{a}$  there exists a case  $x$  such that  $\rho(x, \mathbf{a}) = ?$ , then the case  $x$  should not be included in any blocks  $[(a, v)]$  for all values  $v$  of attribute  $\mathbf{a}$ .
- If for an attribute  $\mathbf{a}$  there exists a case  $x$  such that  $\rho(x, \mathbf{a}) = *$ , then the corresponding case  $x$  should be included in blocks  $[(\mathbf{a}, v)]$  for all specified values  $v$  of attribute  $\mathbf{a}$ .

- If for an attribute  $\mathbf{a}$  there exists a case  $x$  such that  $\rho(x, \mathbf{a}) = -$ , then the corresponding case  $x$  should be included in blocks  $[(\mathbf{a}, v)]$  for all specified values  $v$  of attribute  $\mathbf{a}$  that are members of the set  $V(x, \mathbf{a})$ , where

$$V(x, a) = \{\rho(y, \mathbf{a}) | y \in U, \rho(y, d) = \rho(x, d)\}$$

and  $d$  is the decision.

Thus for table 1,

$$[(O, a_1)] = \{1, 2, 3, 4, 7, 8\},$$

$$[(O, a_0)] = \{5, 6\},$$

$$[(E, b_1)] = \{1, 8\},$$

$$[(E, b_0)] = \{2, 3, 4, 5\},$$

$$[(U, c_1)] = \{1, 3, 4, 6, 7\},$$

$$[(U, c_0)] = \{2, 4, 5, 6\}.$$

These modifications of the definition of the block of attribute-value pair are consistent with the interpretation of missing attribute values [17] lost, “do not care” conditions, and attribute-concept values. Also, note that the attribute-concept value is the most universal, since if  $V(x, a) = \phi$ , the definition of the attribute-concept value is reduced to the lost value, and if  $V(x, a)$  is the set of all values of an attribute  $\mathbf{a}$ , the attribute-concept value becomes a “do not care” condition.

### 3. Generalized Distribution Table

Generalized Distribution Table (GDT) is a table in which the probabilistic relationships between concepts and instances over discrete domains are represented [20], [21]. 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

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

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

$$\left( \prod_{i=1}^m (n_i + 1) \right) - \left( \prod_{i=1}^m n_i \right) + 1 \quad (2)$$

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  $a_0 b_0 c_0$  and  $a_0 b_1 c_0$  describe the same concept, the attribute  $b$  is superfluous, i.e. the concept can be described by  $a_0 c_0$ . Therefore, the generalization  $a_0 * c_0$  is used to describe the set  $\{a_0 b_0 c_0, a_0 b_1 c_0\}$

**The probabilistic relationships** between possible instances and possible generalizations, represented by entries  $G_{i,j}$  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 the background knowledge is not available. Thus, it is defined by

$$G_{i,j} = p(PI_j \setminus PG_i) = \begin{cases} \frac{1}{N_{PG_i}} & \text{if } PG_i \text{ is a Generalization of } PI_j \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where

$PI_j$  is the  $j$ th possible instance,

$PG_i$  is the  $i$ th possible generalization,

and  $N_{PG_i}$  is the number of the possible instances satisfying the  $i$ th possible generalization, that is,

$$N_{PG_i} = \prod_j^m n_j \quad (4)$$

where  $j = 1, \dots, m$ , and  $j\#$  the attribute that is contained by the  $i$ th possible generalization (i.e.,  $j$  just contains the attributes expressed by the wild card).

### Rule Strength

In this approach, 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 induction influenced by both unseen instances and noises. It is defined by

$$s(X \rightarrow Y) = s(X) \cdot [1 - r(X \rightarrow Y)] \quad (5)$$

where  $s(X)$ : The strength of the generalization  $X$  and  $r$ : noise rate function.

### $s(X)$ : The strength of the generalization $X$

It represents explicitly the prediction for unseen instances. It is given by Eq. (7).

$$s(PG_i) = \sum_j p(PI_j \setminus PG_i) = \frac{N_{ins-rel,i}}{N_{PG_i}} \quad (6)$$

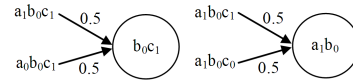


Figure 1 The relationship among generalizations

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

### $r$ : noise rate function

It shows the quality of classification measured by the number of the instances satisfying the generalization  $X$  which cannot be classified into class  $Y$ . The user can specify an allowed noise level as a threshold value. Thus, the rule candidates with a noise level larger than the given threshold value will be deleted. It is defined by,

$$r(X \rightarrow Y) = \frac{N_{ins-rel}(X) - N_{ins-class}(X, Y)}{N_{ins-rel}(X)} \quad (7)$$

Where  $N_{ins-rel(x)}$  is the number of the observed instances satisfying the generalization  $X$ ,  $N_{ins-class}(X, Y)$  is the number of the instances belonging to the class  $Y$  within the instances satisfying the generalization  $X$ . From the GDT, we can see that a generalization is 100% true if and only if all of instances belonging to this generalization appear. Let us use the example shown in Table 1. Considering the generalization  $\{b_0, c_1\}$ , if instances both  $\{a_0 b_0 c_1\}$  and  $\{a_1 b_0 c_1\}$  appear, the strength  $s(\{b_0, c_1\})$  is 1; if only one of  $\{a_0 b_0 c_1\}$  and  $\{a_1 b_0 c_1\}$  appears, the strength  $s(\{b_0 c_1\})$  is 0.5, as shown in Figure 1.

Obvious that one instance can be expressed by several possible generalizations and several instances can be also expressed by one possible generalization. For the example shown in Table 1, the instance  $\{a_1 b_0 c_1\}$  can be expressed by  $\{a_1 b_0\}$ ,  $\{b_0 c_1\}$ , ..., or  $\{c_1\}$ .

Every generalization in upper levels contains all generalizations related to it in lower levels. That is,

$$\begin{aligned} \{a_1\} &\supset \{a_1 b_0\}, \{a_1 c_1\}, \\ \{a_1 b_0\} &\supset \{a_1 b_0 c_1\} \end{aligned}$$

In other words, if the rule  $\{a_1\} \rightarrow y$  is true, the rule  $\{a_1 b_0\} \rightarrow y$  and  $\{a_1 c_1\} \rightarrow y$  are also true. Otherwise, if  $\{a_1 b_0\} \rightarrow y$  or  $\{a_1 c_1\} \rightarrow y$  is false, the rule  $\{a_1\} \rightarrow y$  is also false.

Figure 2 gives the relationship among generalizations.

A generalization that contains the instances with different classes is contradictory, and it cannot be used as a rule. In contrast, a generalization that contains the instances with the same class is consistent, so From Table 1, we can see that the generalizations can be di-

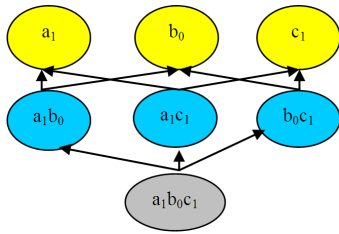


Figure 2 The relationship among generalizations

vided into three groups: contradictory, belonging to class y, and belonging to class n.

#### 4. Searching Algorithm for an Optimal Set of Rules

We now outline the idea of a searching algorithm for a set of rules based on the GDT-RS methodology. a sample decision table shown in Table 1 is used to illustrate the idea.

##### Algorithm

##### Step 1. Create the GDT.

Since:

$$O \in \{a_0, a_1\} \Rightarrow n_1 = 2,$$

$$E \in \{b_0, b_1\} \Rightarrow n_2 = 2,$$

$$U \in \{c_0, c_1\} \Rightarrow n_3 = 2.$$

Hence:

the number of attributes(m)=3,

from Eq.(1) number of the possible instance is 8,

from Eq.(2) number of the possible generalizations is 18.

**Step 2. Simplify the GDT.** By deleting all of the instances and generalizations un-appeared in the example database shown in Table 1. From Table 1, the instances appeared with respect to cases 1, 3, 5, 8 are  $\{a_1 b_1 c_1\}$ ,  $\{a_1 b_0 c_1\}$ ,  $a_0$ ,  $b_0$ ,  $c_0$ ,  $a_1 b_1 c_0$ , respectively.

From Eq.(1) and Table 1, the instance appeared with respect to case 2 is  $\{a_1 b_0 c_1\}$ ;

From Eq.(1) and Table 1 the instance appeared with respect to case 4 may be one of  $\{\{a_1 b_0 c_0\}, \{a_1 b_0 c_1\}\}$ ;

From Table 1, the instance appeared with respect to case 6 may be one of  $\{\{a_0 b_0 c_0\}, \{a_1 b_0 c_1\}, \{a_0 b_1 c_0\}, \{a_0 b_1 c_1\}\}$ ;

Similarly, the instance appeared with respect to case 7 may be one of  $\{\{a_1 b_0 c_1\}, \{a_1 b_1 c_1\}\}$  however  $\{a_1 b_0 c_1\}$  is not consistent with Table 2(see Appendix). So the appeared instance is  $\{a_1 b_1 c_1\}$ . So the simplified GDT is shown in Table 3 (see Appendix).

##### Step 3. Group the generalizations

Generalizations can be divided into three groups contradictory, belonging to class yes, and belonging to

Table 4 The generalizations belonging to class yes

	$a_1 b_0 c_0$	$a_0 b_1 c_0$	$a_1 b_1 c_1$
$a_1 * c_0$	1/2	1/2	
$a_1 b_1 *$		1/2	

Table 5 The generalizations belonging to class no

	$a_0 b_0 c_0$	$a_0 b_0 c_1$	$a_0 b_1 c_0$	$a_0 b_1 c_1$
$a_0 * c_0$	1/2		1/2	
$a_0 * c_1$		1/2		1/2
$a_0 b_0 *$	1/2	1/2		
$a_0 b_1 *$			1/2	1/2
$a_0 **$	1/4	1/4	1/4	1/4

class no. The contradictory generalizations, containing the instances belonging to different decision classes, cannot be used as the rules. Hence they are ignored. In other words, we are just interested in the generalizations belonging to class yes or no, which will be selected as the rules.

##### Step 4. Rule Selection

There are several possible ways for rule selection. For example:

- Selecting the rules that contain as many instances as possible.
- Selecting the rules in the levels of generalization as high as possible according to the number of "\*" in a generalization.
- Selecting the rules with larger strengths.

Since the purpose is to simplify the decision table and simpler results of generalization (i.e., more general rules) are preferred, the first priority to the rules will be what contains more instances, then to the rules corresponding to an upper level of generalization and the third priority to the rules with larger strengths. Thus, from Table 4 and Table 5 the final rule set is  $a_1 c_0 \rightarrow \text{yes}, \text{with } S = 1$   
 $a_1 b_1 \rightarrow \text{yes}, \text{with } S = 2$   
 $a_0 \rightarrow \text{no}, \text{with } S = 1$

##### Results

The induced Rules can be written as:

- **If** (over-current protection, symptom of failure) and (unbalance of three-phase winding direct current resistance, not appearing) **then** (circle short circuit, yes)

Table 2 the GDT for the decision table shown in Table 1

	$a_0 b_0 c_0$	$a_0 b_0 c_1$	$a_0 b_1 c_0$	$a_0 b_1 c_1$	$a_1 b_0 c_0$	$a_1 b_0 c_1$	$a_1 b_1 c_0$	$a_1 b_1 c_1$
$*b_0 c_0$	1/2	L	L	L	1/2			
$*b_0 c_1$		1/2				1/2		
$*b_1 c_0$			1/2				1/2	
$*b_1 c_1$				1/2				1/2
$a_0 * c_0$	1/2		1/2					
$a_0 * c_1$		1/2		1/2				
$a_1 * c_0$					1/2		1/2	
$a_1 * c_1$						1/2	1/2	
$a_0 b_0*$	1/2	1/2						
$a_0 b_1*$			1/2	1/2				
$a_1 b_0*$					1/2	1/2		
$a_1 b_1*$							1/2	1/2
$**c_0$	1/4		1/4		1/4		1/4	
$**c_1$		1/4		1/4		1/4		1/4
$a_0 **$	1/4	1/4	1/4	1/4				
$a_1 **$					1/4	1/4	1/4	1/4
$*b_0*$	1/4	1/4			1/4	1/4		
$*b_1*$			1/4	1/4			1/4	1/4

Table 3 The simplified GDT for the decision table shown in Table

	$a_0 b_0 c_0$	$a_0 b_0 c_1$	$a_0 b_1 c_0$	$a_0 b_1 c_1$	$a_1 b_0 c_0$	$a_1 b_0 c_1$	$a_1 b_1 c_0$	$a_1 b_1 c_1$
$*b_0 c_0$	1/2(n)				1/2(yy)			
$*b_0 c_1$		1/2(n)				1/2(ny)		
$*b_1 c_0$			1/2(n)				1/2(y)	
$*b_1 c_1$				1/2(n)				1/2(yy)
$a_0 * c_0$	1/2(n)		1/2(n)					
$a_0 * c_1$		1/2(n)		1/2(n)				
$a_1 * c_0$					1/2(yy)		1/2(y)	
$a_1 * c_1$						1/2(ny)	1/2(yy)	
$a_0 b_0*$	1/2(n)	1/2(n)						
$a_0 b_1*$			1/2(n)	1/2(n)				
$a_1 b_0*$					1/2(yy)	1/2(ny)		
$a_1 b_1*$							1/2(y)	1/2(yy)
$**c_0$	1/4(n)		1/4(n)		1/4(yy)		1/4(y)	
$**c_1$		1/4(n)		1/4(n)		1/4(ny)		1/4(yy)
$a_0 **$	1/4(n)	1/4(n)	1/4(n)	1/4(n)				
$a_1 **$					1/4(yy)	1/4(ny)	1/4(y)	1/4(yy)
$*b_0*$	1/4(n)	1/4(n)			1/4(yy)	1/4(ny)		
$*b_1*$			1/4(n)	1/4(n)			1/4(y)	1/4(yy)

- **If** (over-current protection, symptom of failure ) and ( Exceeding of winding insulation resistance, symptom of failure) **then** (circle short circuit, yes) ”
- **If** (over-current protection, not appearing ) **then** (circle short circuit, no)

## 5. Power Transformer Fault Diagnosis

In the normal operation of the transformer, the released gases are methane ( $CH_4$ ), ethane ( $C_2H_6$ ), Hydrogen ( $H_2$ ), ethylene ( $C_2H_4$ ), and acetylene ( $C_2H_2$ ) and so on. When there is an abnormal situation such as occurring a fault, some specific gases are produced more than in the normal operation and the amount of them in the transformer oil increases. The increase in the amount of gases results in saturation of the transformer oil and no more gas can be dissolved in oil. From the point of rough set theory, the fault diagnosis table of power transformer is also a decision table. In this section, an example is given to show how the proposed algorithm can be used to generate diagnostic rules for the power transformer. According to the historical fault data of the power transformer, the fault decision table is shown in Table 1. Here, the condition attributes are concentrations (ppm by volume) of dissolved gases in the insulation oil, such as  $H_2$ ,  $CH_4$ ,  $C_2H_6$ ,  $C_2H_4$ , and  $C_2H_2$ . The decision attribute (D) is the fault class of the transformer, where “0” represents the fault of local discharge, “1” represents the fault of low-energy discharge, “2” represents the fault of high-energy discharge, “3” represents the fault of low-temperature superheat, “4” represents the fault of medium-temperature superheat and “5” represents the fault of high-temperature superheat.

According to rough set theory and generalized distribution table algorithm, the fault diagnosis rules are:

- If  $H_2 = M$ ,  $CH_4 = H$ ,  $C_2H_6 = H$ ,  $C_2H_4 = H$  and  $C_2H_2 = L$ , then  $D=3$ .
- If  $H_2 = M$ ,  $CH_4 = L$ ,  $C_2H_6 = L$ ,  $C_2H_4 = M$  and  $C_2H_2 = M$ , then  $D=1$ .
- If  $H_2 = H$ ,  $CH_4 = L$ ,  $C_2H_6 = L$ ,  $C_2H_4 = L$  and  $C_2H_2 = M$ , then  $D=2$ .
- If  $H_2 = M$ ,  $CH_4 = L$ ,  $C_2H_6 = L$ ,  $C_2H_4 = L$  and  $C_2H_2 = L$ , then  $D=0$ .

Table 6 The decision table for the transformer fault

Object	$H_2$	$CH_4$	$C_2H_6$	$C_2H_4$	$C_2H_2$	D
$X_1$	H	L	L	L	M	2
$X_2$	M	L	*	M	M	1
$X_3$	*	M	H	M	L	5
$X_4$	M	H	H	H	-	3
$X_5$	L	L	L	L	L	0
$X_6$	L	M	M	-	L	5
$X_7$	H	H	?	H	L	3
$X_8$	M	*	M	M	L	4
$X_9$	M	*	L	M	H	1
$X_{10}$	L	M	H	M	L	5
$X_{11}$	M	L	L	M	M	1
$X_{12}$	-	L	L	L	L	0
$X_{13}$	H	L	L	L	*	2
$X_{14}$	H	H	H	H	L	3
$X_{15}$	-	M	H	M	L	3

- If  $H_2 = M$ ,  $CH_4 = M$ ,  $C_2H_6 = M$ ,  $C_2H_4 = M$  and  $C_2H_2 = L$ , then  $D=4$ .
- If  $H_2 = L$ ,  $CH_4 = M$ ,  $C_2H_6 = M$ ,  $C_2H_4 = M$  and  $C_2H_2 = L$ , then  $D=5$ .
- If  $H_2 = L$ ,  $CH_4 = M$ ,  $C_2H_6 = H$ ,  $C_2H_4 = M$  and  $C_2H_2 = L$ , then  $D=5$ .

## 6. Conclusions

This paper has presented an effective and efficient approach to extract diagnosis rules from incomplete and redundant data set of power transformers by using rough set theory and generalized distribution table. The extracted diagnosis rules can effectively reduce space of input attributes and simplify knowledge representation for fault diagnosis. The fault diagnosis decision table is first built through discretized attributes. Next, the generalized distribution table is created. Finally, the rule simplification process is adapted to achieving the decision rules derived from incomplete and redundant information. Main features of that methodology can be summarized as follows:

- (1) It can discover If-Then rules from very large, complex databases.
- (2) It represents explicitly the uncertainty of a rule including the prediction of possible instances in the strength of the rule.
- (3) Lost values are considered during the process of rule induction.
- (4) It can flexibly select biases for search control.
- (5) It can effectively handle noisy data, missing data.

## References

- [1] X. Yu and H. Zang, "Transformer Fault Diagnosis Based on Rough Sets Theory and Artificial Neural Networks," *2008 International Conference on Condition Monitoring and Diagnosis*, Beijing, China, April 21-24, 2008.
- [2] Y. C. Huang, H. C. Sun, and H. J. Lu, "Decision Rule Generation Using Data Mining Approach," *paper submitted to 2009 International Conference on Advanced Information Technologies (AIT)*, Taichung, Taiwan, April 24-25, 2009.
- [3] H. Su, "Multi-Source Fuzzy Information Fusion Method Based on Bayesian Optimal Classifier", *Acta Automatic Sinica*, 34(3), pp.282-287, 2008.
- [4] Sun H., Li D., et al., "Electric Power Transformer Fault Diagnosis Using Decision Tree". *Chin. Soc. for Elec. Eng.* 21 50-55, 2001.
- [5] Sun X., Guo J., et al, "Study on Fault Diagnose Method of Transformer DGA with Fuzzy Model Hierarchy Classification". *Chin. Soc. for Elec. Eng.* 21 37-41, 2001.
- [6] Wang J., Ji Y., "Application of fuzzy Petri nets knowledge representation in electric power transformer fault diagnosis". *Proceedings of the CSEE*. 23, 121-125, 2003.
- [7] Guoyin Wang, "Extension of Rough Set under Incomplete Information Systems", *National Science Foundation of china* ( No. 69803014 ), PD program of P.R. China, 2002.
- [8] Grzymala-Busse J. W., "Three Approaches to Missing Attribute Values - A Rough Set Perspective ", *Workshop on Foundations of Data Mining*, associated with the fourth IEEE International Conference on Data Mining, Brighton, UK, November 1-4, 2004
- [9] Grzymala-Busse J. W. and Wang A. Y., "Modified algorithms LEM1 and LEM2 for rule induction from data with missing attribute values". *Proc. of the Fifth International Workshop on Rough Sets and Soft Computing (RSSC'97) at the Third Joint Conference on Information Sciences (JCIS'97)*, Research Triangle Park, NC, March 2-5, 69-72, 1997.
- [10] Grzymala-Busse J. W. and M. Hu, "A comparison of several approaches to missing attribute values in data mining". *Proceedings of the Second International Conference on Rough Sets and Current Trends in Computing RSCTC'2000*, October 16-19, Banff, Canada, 340-347, 2000.
- [11] Grzymala-Busse J. W., "Data with missing attribute values: Generalization of indiscernibility relation and rule induction". *Transactions on Rough Sets*, Lecture Notes in Computer Science Journal Sub line, Springer-Verlag, vol. 1, 78-95, 2004.
- [12] Grzymala-Busse J. W., "On the unknown attribute values in learning from examples". *Proc. of the ISMIS-91, 6th International Symposium on Methodologies for Intelligent Systems*, Charlotte, North Carolina, October 16-19, 1991, 368-377, *Lecture Notes in Artificial Intelligence*, vol.542, Springer-Verlag, Berlin, Heidelberg, New York, 1991.
- [13] Grzymala-Busse J. W., "Rough Set Strategies to Data with Missing Attribute Values" , *Proceedings of the Workshop on Foundations and New Directions in Data Mining*, associated with the third IEEE International Conference on Data Mining, November 19-22, Melbourne, FL, USA, 56-63, 2003.
- [14] Marte Skarstein Bjanger , "vibration Analysis in Rotating Machinery using Rough Set theory and ROSETTA" . *Tech. report*, Univ. of Norwegian, 1999./
- [15] Zhong N. and Ohsuga S., "Using Generalization Distribution Tables as a Hypotheses Search Space for Generalization", *Proc. 4th International Workshop on Rough Sets, Fuzzy Sets, and Machine Discovery (RSFD-96, 396-403)*, 1996.
- [16] Zhong N., Dong J. Z., and Ohsuga S., "Discovering Rules in the Environment with Noise and Incompleteness", *Proc. 10th International Florida AI Research Symposium (FLAIRS-97) edited in the Special Track on Uncertainty in AI*, 186-191, 1997.
- [17] Zhong N., Dong J. Z., and Ohsuga S., "Soft Techniques to Rule Discovery in Data", *Proc. the Fifth Congress on Intelligent Techniques and Soft Computing (EUFUT-97) edited in the Invited Session on Soft Techniques in Knowledge Discovery*, 212-217, 1997.
- [18] Zhong N., Fujitsu S. and Ohsuga S, "Generalization Based on the Connectionist Networks Representation of a Generalization Distribution Table", *Proc. First Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD-97)*, World Scientific, 183-197, 1997.
- [19] H.Su, and Q.Li, "Application Study for Fault diagnosis of Transformer Based on concept of Lattice", *High Voltage Engineering*, Vol.32, No2, pp.12-17, 2008.
- [20] Weihua Zhu , Wei Zhang, Yunqing Fu , "An Incomplete Data Analysis Approach using Rough Set Theory", *Proceedings of the 2009 international Conference on Intelligent Mechatronics and Automation*, Chengdu , China August 2009.
- [21] Pawlak Z., Busse J. G., R. Slowinski and Ziarko W., Rough Sets, *Communication of the ACM*, vol.38:11, pp.89-95 , 1995.