Transformer Fault Diagnosis Based on Reasoning Integration of Rough Set and Fuzzy Set and Bayesian Optimal Classifier

HONGSHENG SU, HAIYING DONG
School of Automation and Electrical Engineering
Lanzhou Jiaotong University
Lanzhou 730070
P.R.CHINA
shsen@163.com

Abstract: - In accordance with intelligent complementary strategies, a new transformer fault diagnosis method is proposed based on rough set (RS) and fuzzy set (FS) and Bayesian optimal classifier in this paper. Through RS reduction, the diagnostic decision table is greatly simplified and fault symptoms information is compressed, dramatically, and the minimal decision rules can be obtained. In the light of the minimal decision rules, the complexity of Bayesian reasoning and difficulties of fault symptom acquisition are dramatically decreased. Moreover, probability reasoning may be realized applying Bayesian optimal classifier, it can be used to describe the characteristics of fault information and investigate the fault reasons of transformer. In the end, a practical application in transformer fault diagnosis indicates that the proposed method is very effective and intelligent and ubiquitous.

Key-Words: - Rough set; Fuzzy set; Bayesian optimal classifier; Fault diagnosis; Information entropy, Intelligent complementary; Transformer

1 Introduction

Transformers are considered to be the significant equipments in power supply and distribution systems, once the facilities are in failure, the enormous economic loss will be generated. To maintain normal operation of the facilities, all sorts of diagnosis methods are created to implement intelligent forecasting of the faults, recently. In [1-3] artificial neural networks (ANN) is applied to implement the fault diagnosis, it deals with the "bottleneck" difficulties in terms of self-learning and knowledge acquisition, effectively. However, when the samples space distribution is more complex, the convergence of artificial neural networks is quite difficult, which limits its farther applications, consequently. In [4-5] the connections between fault-sources and fault-symptoms could be established applying fuzzy technology, which effectively overcomes the "bottleneck" difficulties of expert knowledge acquisition to some extent. However, fuzzy membership functions as well as the shape parameters are constructed by man in advance, subjectively. Thus, the applicable object of fuzzy inference is a confirmed system before hand. In [6-7] rough set (RS) is used to mine diagnostic knowledge and discovery unknown rules from diagnostic knowledge library, thus, the simplified diagnostic decision table may be acquired and diagnostic efficiency may be improved, greatly. However, while fault information is deficient and indeterminate, RS approach becomes quite difficult to face them. In [8-9], expert systems (ES) are applied to perform fault diagnosis. It effectively emulated expert's reasoning and decision process. But the difficulties of the approach lie in that the reasoning error could happen if the constructed expert knowledge base isn't self-contained and rounded. In [10-12], Petri net is applied to simulate the faults mechanism, the diagnostic accuracy is improved, dramatically, but while the system scale is large and complex, Petri net model becomes quite complicated not to be well applied. In [13-15] multi-agent (MA) technology is applied to perform the distributed state monitoring and faults diagnosis beforehand, the investigation shows that they could cooperate with each knowledge, aim, technology and programming to accomplish each task through adopting joint action. But in distributed heterogeneous environments and under more complicated conditions, the study how to cooperate with its knowledge and technique is a complex task. Seen from the above analysis, due to the complexities of the transformer faults and the determination of the transformer operational surroundings, as well as the some deficiencies in accuracy and amount of the acquired data, the above diagnostic methods present some in-adaptabilities, whose diagnostic results is not satisfying. To change the state, new intelligent diagnosis methods are required to perform fault forecasting. In addition, single intelligent method possesses some flaws. Hence, it is necessary to incorporate with all kinds of intelligent methods to realize correctly diagnosis according to intelligent complimentary strategies.

The main thinking of RS is that the classification rules of the concepts are educed through knowledge reduction for same classification abilities as in [16-17]. Bayesian optimal classifier is considered to be able to make the likelihood of a new instance to be correctly classified up to maximum by incorporating with the posterior probabilities of all assumptions in terms of same hypothesis space and same observed data and same prior probabilities of these assumptions in [18]. However, Bayesian reasoning is more complex while the involved attributes are more. Hence, it is quite necessary to integrate the two together to reduce the complexities of Bayesian reasoning. In this paper, a new transformer faults diagnosis method is proposed based on rough set theory (RST) and Bayesian optimal classifier according to intelligent complementary fusion thinking, the constructed Bayesian model is applied to large transformer fault diagnosis, and the satisfying diagnostic results are achieved.

2 Rough Set Theory

The attributes set which is composed of the conditional attribute C and the decision attribute D, $A=C\cup D$, $C\cap D=\varnothing$, $a\in A, V=V_a$, V_a is scope of In RST, knowledge denotation system may be described by.

$$S=<U,A,V,F> \quad (1)$$

where U is the universe and expresses a set of the finite objects, A is ta, f: $U\times A\rightarrow V$ is a information function, it specifies attribute values of every object in U.

Information systems based on rough sets definition can be denoted by the use of table format, where columns express attributes and rows represent objects, and every row describes information of an object. The table therefore is called decision table, which can generalize the relationships among data and educe the classification rules of the concepts. In rough sets, binary indivisible relationship ind(R) determined by $R\subseteq A$ can be expressed by

$$\text{ind}(R)=\{(x,y)\in U\times U \mid \forall a\in A, f(x,a)=f(y,a)\} \quad (2)$$

It is very clearly that if $(x,y)\in \text{ind}(R)$, then x and y can't be differentiated in accordance with existing information, they are an equivalent relation in U.

Let $S=<U, C\cup D>$, if $C1\in C$, $C1\neq\varnothing$, and the following two conditions hold.

(1) $\text{ind}_{C1}(D)=\text{ind}_C(D)$

(2) $\forall C2\subseteq C1$, $\text{ind}_{C2}(D)\neq\text{ind}_{C1}(D)$

According to (1) and (2), we can say $C1$ is a reduction of C with regard to D, the intersection of all these reductions is called core, and defined as $\text{core}_D(C)=\cap\text{red}_D(C)$.

Through the above reduction we can get several reduction attribute sets. The best attributes combination is considered to possess the smallest average intervolving information, whose basic steps are described as follows.

1) Firstly, the intervolving information between all the two in each attribute combination is worked out, and the results are then totaled and averaged. The acquired average quantity is considered as the average intervolving information of the reduction attribute combination.

2) Secondly, the average intervolving information of all reduction attribute combinations are worked out, afterwards, the attribute combination possessing the smallest average intervolving information may be selected as the best reduction attribute set. The smallest means the appropriately inter-independent attribute combination. In [19], the intervolving information based on information entropy is defined as follows.

Definition1. The information entropy of the equivalent relation $G(U|I(G)=\{x_1, x_2,\ldots, x_n\})$ can be expressed by $H(G)$, the conditional entropy that the equivalent relation $Q(U|I(Q)=\{y_1, y_2,\ldots, y_n\})$ is relative to G can be expressed by $H(Q|G)$, then the involving information between G and Q is defined by

$$I(Q,G)=H(Q)-H(Q\mid G)=\sum_{x,y} p(x,y)\log_2\frac{p(x,y)}{p(x)p(y)} \quad (3)$$

where p expresses the probability.

3 Fuzzy Set Foundation

Fuzzy set was proposed by L. Zadeh in 1965, it could dispose fuzzy information, and tackle the "bottle neck" difficulty of fuzzy knowledge acquisition. In fuzzy set, fuzzy membership function was used to describe the fuzziness of faults information, and fuzzy algorithms were used to perform fuzzy operation.

Let V be an objects space, $\forall x\in V$, $A\subseteq V$. To study whether x belongs to A or not, a characteristic function $u_A(x)$ is defined, thus x, together with $u_A(x)$, constitutes a sequence couple $(x, u_A(x))$. Thus, fuzzy subset A in V may be defined as $A=\{x,u_A(x)|x\in V\}$, $u_A(x)$ is defined as fuzzy membership function of x to A, $u_A(x) \in[0,1]$.

Let A an B respectively express two fuzzy subsets in V, $u_A(x)$ and $u_B(x)$ express their membership functions, thus the basic fuzzy operations are defined as follows.

$$u_{A\cup B}(x)=\max[u_A(x), u_B(x)] = u_A(x) \vee u_A(x),\forall x \in V \quad (4)$$

$$u_{A\cap B}(x)=\min[u_A(x), u_B(x)]= u_A(x) \wedge u_A(x),\forall x \in V \quad (5)$$

$$u_{\bar{A}}(x)=1- u_A(x), \forall x \in V \quad (6)$$

The familiar fuzzy membership functions include Gauss, triangle, trapezoid, bell shape and so on.

4 Bayesian Optimal Classifier

Let $P(h)$ be the prior probability of the hypothesis h, $h\in H$, H is the hypothesis space, $P(D)$ specify the prior probability of the observed data D, $P(D/h)$ specify the likelihood of D occurrence when h is known, inversely, $P(h/D)$ be the likelihood of h occurrence while D is observed. $P(h/D)$ is called as the posterior probability of h, which reflects the influence degree of D to h. Thus, Bayesian law is described as follows.

$$P(h/D)=P(D/h)P(h)/P(D) \quad (7)$$

Since D is a constant and independent of h, hence

$$P(h/D) \propto P(D/h)P(h) \quad (8)$$

In this way, while a new instance D comes up, whose most possible classification $h\in H$ is called as the maximal posteriori (MAP) hypothesis. Definitely speaking, h_{MAP} can be called MAP hypothesis only when the following formula holds.

$$h_{\mathrm{MAP}} = \operatorname*{argmax}_{h\in H} P(h/D) \quad (9)$$

Up to now, what we discuss only is which one is its most possible hypothesis while D occurs. Actually, another interesting problem related to it is which one is the most possible classification while D is observed. For the latter, we may dispose it using MAP hypothesis to likely classifications of a new instance, that is,

$$c_{\mathrm{MAP}} = \operatorname*{argmax}_{c\in C} P(C/h_{\mathrm{MAP}}) \quad (10)$$

In (10) C is the possible classification space of the new instance; c is its possible classification, $c\in C$, c_{MAP} is the most likely classification. But in fact, we still have better algorithm, i.e., Bayesian optimal classifier. In general, the most likely classification of the new instance may be acquired by incorporating with the predictions of all assumptions.

Let $P(c_j/D)$ denote the probability of the likely classification of the new instance D, then

$$P(c_j \mid D) = \sum_{h_i\in H} P(c_j \mid h_i)P(h_i \mid D) \quad (11)$$

Thus, the optimal classification of the new instance D is c_j because it lets $P(c_j/D)$ up to the maximum, i.e.,

$$\operatorname*{argmax}_{c_j\in C} \sum_{h_i\in H} P(c_j \mid h_i)P(h_i \mid D) \quad (12)$$

Let us consider the possible classification of a new instance below.

Let the possible classification set of a new instance x be represented by $C= \{c_1, c_2\}$, the hypothesis space of which be described by $H= \{h_1, h_2, h_3\}$. Supposing that $P(h_1|x)$=0.3, $P(h_2|x)$=0.3, $P(h_3|x)$=0.4, where x is classified as c_1 by h_1 and h_2, and c_2 by h_3. Hence, $P(c_1|h_1)$=1, $P(c_2|h_1)$=0; $P(c_1|h_2)$=1, $P(c_2|h_2)$=0; $P(c_1|h_3)$=0, $P(c_2|h_1)$=1; Then according to (11), we have

$$\sum_{h_i\in H} P(c_1 \mid h_i)P(h_i \mid x) = 0.6$$

$$\sum_{h_i\in H} P(c_2 \mid h_i)P(h_i \mid x) = 0.4$$

From (12), we have

$$\underset{c_j \in \{c_1, c_2\}}{\operatorname{argmax}} \sum_{h_i \in H} P(c_j \mid h_i) P(h_i \mid x) = c_1$$

According to (8), we then have

$$\underset{c_j \in C}{\operatorname{argmax}} \sum_{h_i \in H} P(c_j \mid h_i) P(D \mid h_i) P(h_i) \qquad (13)$$

5 Bayesian Optimal Classifier Based on Fuzzy Set Theory

In practice, the observed information is usually fuzzy and indeterminate, for example, transformer faults symptom information. Hence, it will be very difficult to apply Bayesian optimal classifier to implement fault diagnosis, directly. According to the description in [20], fuzzy set is embedded into Bayesian optimal classifier to generate fuzzy Bayesian optimal classifier, which not only can deal with fuzzy information, effectively, but only tackle the "bottle neck" of fuzzy information acquisition in Bayesian optimal classifier. In (12), we apply fuzzy membership function $u_{hi}(D)$ to replace $P(h_i|D)$, then

$$\underset{c_j \in C}{\operatorname{argmax}} \sum_{h_i \in H} P(c_j \mid h_i) u_{hi}(D) \qquad (14)$$

where $u_{hi}(D)$ may be understood as: under the observed information D, it can be interpreted as the probability of the known classification h_i. Clearly, it is consistent with $P(h_i|D)$.

6 Transformer Fault Diagnosis Method Based on Rough Set and Fuzzy Set and Bayesian Optimal Classifier

The minimal decision rules can be obtained applying rough set to realize the knowledge reduction and compress the fault characteristics, and high-efficient fast fault diagnosis can be implemented applying reasoning abilities of Bayesian optimal classifier. Hence, the two possesses stronger complementary properties.

According to the descriptions in [21-23], the fault symptom set M of the transformer is described as shown in Table 1, and the fault set D is described as shown in Table 2, and the connection relation C between the fault source and the fault symptom information is described as shown in Table 3.

Tab.1 The fault symptom table

Symptom code	Symptom type
m_1	Heat fault characteristics based on three-ratio-code
m_2	Water capacities in transformer oil
m_3	Iron core earth current
m_4	Three-phase unbalance coefficient in direct current resistance of the winding
m_5	Local discharge
m_6	Discharge fault characteristics based on three-ratio-code
m_7	Absolute value of the deviation of the winding transformation ratio of the winding
m_8	CO/CO_2
m_9	Absorption ratio or polarization index

Tab.2 The fault source table

Fault code	Fault type
d_1	Multi-point earth or local short circuit in iron core
d_2	Leaking magnetism heating or magnetism shield overheat
d_3	Insulation aging
d_4	Insulation dampening
d_5	Tapping switch or down-lead fault
d_6	Suspending discharge
d_7	Winding distortion and circle short
d_8	Circle short and insulation damage
d_9	Encloser discharge

Tab.3 Connection intensity c_{ij}

	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9
d_1	0.32	–	0.9	–	0.07	0.05	–	–	–
d_2	0.28	–	–	–	0.09	0.08	–	–	–
d_3	0.08	0.19	–	–	–	–	–	0.27	–
d_4	–	0.51	–	–	–	–	–	–	0.75
d_5	0.26	–	–	0.87	–	0.06	–	–	–
d_6	–	–	–	–	0.22	0.24	–	–	–
d_7	0.06	–	–	–	0.18	0.19	0.5	0.24	–
d_8	–	–	–	–	0.22	0.14	0.5	0.22	–
d_9	–	0.3	–	–	0.22	0.24	–	0.25	–

Tab.4 Discretized decision table

	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9
d_1	1	0	2	0	1	1	0	0	0
d_2	1	0	0	0	1	1	0	0	0
d_3	1	1	0	0	0	0	0	1	0
d_4	0	2	0	0	0	0	0	0	2
d_5	1	0	0	2	0	1	0	0	0
d_6	0	0	0	0	1	1	0	0	0
d_7	1	0	0	0	1	1	2	1	0
d_8	0	0	0	0	1	1	2	1	0
d_9	0	1	0	0	1	1	0	1	0

Seen from RST, Table 3 is a sheet of decision table, whose conditional attributes are described by $\{m_1,m_2,\ldots,m_9\}$, and the decision attributes are described by $\{d_1,d_2,\ldots,d_9\}$. However, the data in Table 3 are the continuous values, and must be discretized according to RST. Therefore, let the connection intensity be 2 if $c_{ij}\geq 0.5$, and 1 if $0<c_{ij}<0.5$, and 0 otherwise. Thus, the conditional attributes in Table 3 are discretized as shown in Table 4.

Through rough set reduction, the minimal attributes sets are described by $\{m_1, m_2, m_3, m_4, m_7\}$, $\{m_1, m_2, m_3, m_4, m_8\}$, $\{m_1, m_2, m_3, m_5, m_7\}$, $\{m_1, m_2, m_3, m_5, m_8\}$, and $\{m_1, m_3, m_5, m_7, m_8\}$. According to (3), the average involving information of the above groups can be expressed by 0.084, 0.110, 0.153, 0.093, and 0.087. Clearly, since the average information of $\{m_1, m_2, m_3, m_4, m_7\}$ is the smallest, we select it as the minimal reduction set. The simplified decision table is gained as shown in Table 5. Clearly, the attributes now become 5 from initial 9, and the redundant information is for that ignored. But for comparison here, we also select the second minimal reduction set $\{m_1, m_3, m_5, m_7, m_8\}$ as reference as shown in Table 6. In accordance with Table 5 and Table 6, the networks diagrams of Bayesian reasoning are described as shown in Figure 1 and Figure 2, where the fault symptom type m_i is the father node, and the fault type d_i is the children node.

Tab.5 Simplified decision table

	m_1	m_2	m_3	m_4	m_7
d_1	1	0	2	0	0
d_2	1	0	0	0	0
d_3	1	1	0	0	0
d_4	0	2	0	0	0
d_5	1	0	0	2	0
d_6	0	0	0	0	0
d_7	1	0	0	0	2
d_8	0	0	0	0	2
d_9	0	1	0	0	0

Tab.6 Simplified decision table

	m_1	m_3	m_5	m_7	m_8
d_1	1	2	1	0	0
d_2	1	0	1	0	0
d_3	1	0	0	0	1
d_4	0	0	0	0	0
d_5	1	0	0	0	0
d_6	0	0	1	0	0
d_7	1	0	1	2	1
d_8	0	0	1	2	1
d_9	0	0	1	0	1

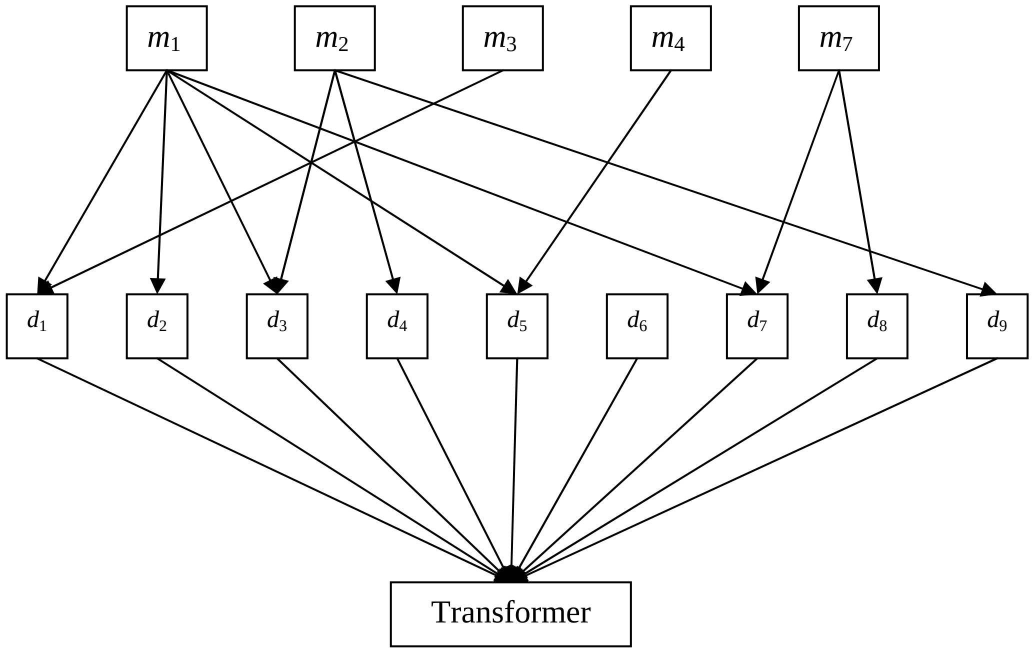

Fig.1 The Bayesian reasoning model 1 of the transformer fault diagnosis

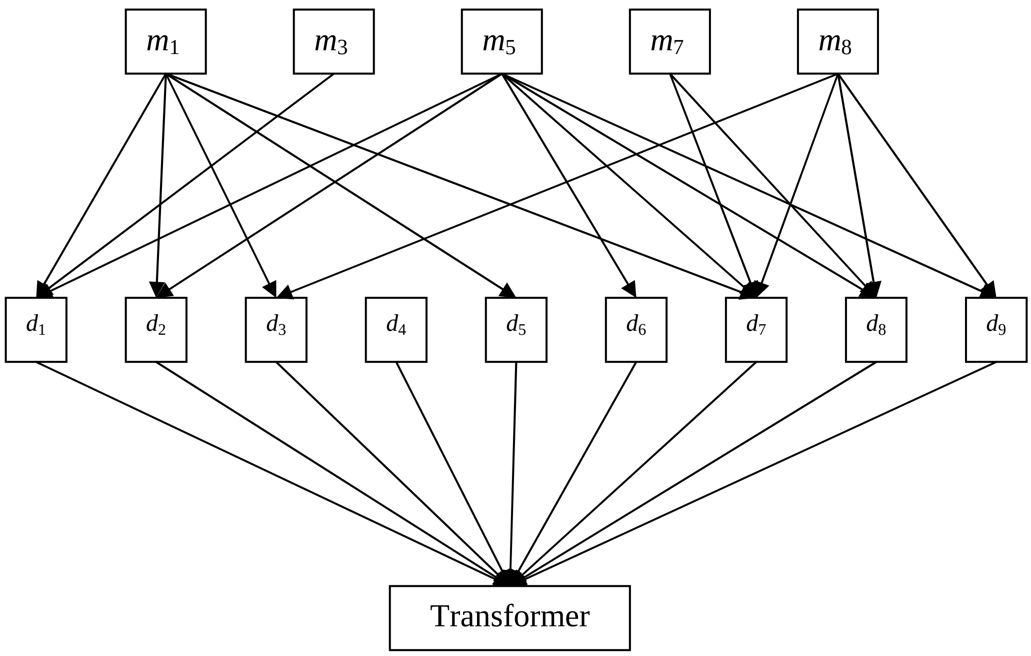

Fig.2 The Bayesian reasoning model 2 of the transformer fault diagnosis

7 Example

The results of the dissolved gas analysis (DGA) in the main transformer are described as shown in Table 7, the earth current of the iron core is 0.1 A, the water capacities in transformer oil and local discharge quantities are in normal scope.

Tab.7 Components of dissolved gas in a transformer

$\times 10^{-6}$

φ(H_2)	φ(CH_4)	φ(C_2H_6)	φ(C_2H_2)	φ(C_2H_4)	φ(CO)	φ(CO_2)
70.4	69.5	28.9	10.4	241.2	704	3350

The rations between each characteristic gas are described by φ(C_2H_2)/ φ(C_2H_4) =0.043, φ(CH_4)/ φ(H_2) =0.99, φ(C_2H_4)/ φ(C_2H_6) =8.35, φ(CO)/φ(CO_2) =0.21, three-ratio-code is for that 002. The earth current of the iron core is normal, and the φ(CO)/φ(CO_2)=0.21 is eligible. According to [19], the fault symptoms of the transformer are worked out by $0.475/m_1+0.083/m_3+0.375/m_4$. Order g=0.2 be a design threshold here. The subjection degree of m_3 is lower than g, it is therefore ignored. Thus we get the symptom set $M=0.475/m_1 +0.375/m_4$. According to Table 3, we have $D=\{d_1, d_2, d_3, d_5, d_7\}$, let the prior probability of each fault be 0.227, 0.063, 0.053, 0.132, 0.120. Then according to Table 3 and the formula (14), applying the diagnostic network model 1 in Fig.1, the probability of each fault occurrence is calculated by

$p(d_1)=0.475\times0.32\times0.227=0.035;$

$p(d_2)=0.475\times0.28\times0.063=0.008;$

$p(d_3)=0.475\times0.08\times0.053=0.002;$

$p(d_5)=[0.475\times0.26+0.375\times0.87]\times0.132=0.059;$

$p(d_7)=0.475\times0.06\times0.12=0.003.$

Since $p(d_5)=\max\{ p(d_1), p(d_2), p(d_3), p(d_5), p(d_7)\}$, the most likely fault is d_5, i.e., tapping switch or down-lead fault. Finally, the field checking proves the correctness of the diagnostic result.

On the other hand, if we applying the diagnostic network 2 in Fig.2, and the prior probability of each fault occurrence is unchangeable all the same, then, we will get the probability of each fault occurrence as follows:

$p(d_1)=0.475\times0.32\times0.227=0.035;$

$p(d_2)=0.475\times0.28\times0.063=0.008;$

$p(d_3)=0.475\times0.08\times0.053=0.002;$

$p(d_5)=0.475\times0.26\times0.132=0.016;$

$p(d_7)=0.475\times0.06\times0.12=0.003.$

Since $p(d_1)=\max\{ p(d_1), p(d_2), p(d_3), p(d_5), p(d_7)\}$, the most likely fault is d_1, i.e., Multi-point earth or local short circuit in iron core. Clearly, the diagnostic result is not consistent with one of the diagnostic networks 1, why?

The reason lies in that the former reduction attribute set $\{m_1, m_2, m_3, m_4, m_7\}$ contains fault information $\{m_1, m_4\}$, while the latter reduction attribute set $\{m_1, m_3, m_5, m_7, m_8\}$ loses fault information m_4, is m_4 redundant information? Clearly, it is not. Hence, if Bayesian reasoning accords with practice, the smallest reduction attribute set should contain all occurred fault symptom information, and otherwise the correct result will be impossibly achieved. In this example, if we select attribute sets $\{m_1, m_2, m_3, m_4, m_7\}$ and $\{m_1, m_2, m_3, m_4, m_8\}$to implement Bayesian reasoning, then we will have correct results, otherwise if we select the other three attribute sets $\{m_1, m_2, m_3, m_5, m_7\}$, $\{m_1, m_2, m_3, m_5, m_8\}$, and $\{m_1, m_3, m_5, m_7, m_8\}$, we will be impossible to get correct answers. Hence, when applying RS to perform attribute reduction, it is noted whether the selected reduction attribute set contains helpful information or not.

In [24], an attribute reduction method is proposed based on alterative precision RS condition entropy, although the constrained condition appreciably loosen, the problem could not resolved well, radically. The reason lies in that the method could not ensured to be almighty in each aspect. Today, there are quite a few examples applying RS to implement attributes reduction, such as [25-26], where the role of RS is same with one mentioned in this paper, that is, to remove redundant information and reduce the size of problem. But when performing probability reasoning, it should be noted that whether the selected reduction attribute set contains usable information or not. Hence, it is required to illustrate that in some combination examples of RS and Bayesian reasoning, if RS is not be applied to tackle the continuous attribute value decision table or the discretized decision table is not used to implement probability reasoning, then it will be another pair of shoes.

In [21] evidence theory (ET) is applied to implemented the transformer fault diagnosis, the results are compared with the proposed method in this text as shown in Table 7.

Tab.7 Comparison of the diagnostic results

Methods	$p(d_1)$	$p(d_2)$	$p(d_3)$	$p(d_5)$	$p(d_7)$
ET	0.035	0.008	0.002	0.056	0.003
Bayesian method	0.035	0.008	0.002	0.059	0.003

Seen from Table 7, the diagnostic results of the two are homogeneous, what is the different is only $p(d_5)$. The probability of $p(d_5)$ using ET is lower than one using the proposed Bayesian method in the text, the reason is that the former considers the residual probability while the latter doesn't. Hence, in theory, the former is more exact than the latter, but more complicated. The latter is more simple and direct, and expediently understanding.

8 Conclusion

According to intelligent complementary thinking of soft computing methodology, rough set is applied to implement the compression of fault characteristics and simplify expert knowledge, and Bayesian optimal classifier based on fuzzy set theory is then applied to perform fault diagnosis of the transformer. On the one hand, the complexities of Bayesian reasoning are reduced. On the other hand, Bayesian optimal classifier with fuzzy information imbedded not only tackles the "bottleneck" difficulty in fuzzy knowledge, but also extends the applying scope of Bayesian optimal classifier. And so, an intelligent fusion method on soft computing methodology is presented related to RS and Bayesian method and fuzzy set in this paper. Fault diagnosis results in the transformer indicate that the method is very effective and quite ubiquitous, and is an effective indeterminate reasoning method.

9 Acknowledgements

This work is partially supported by the 'Qing Lan' Talent Engineering Funds supported by Lanzhou Jiaotong University (QL-06-19A). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

References:

[1] Y. Zhang, A New Artificial Network Approach to Transformer Fault Diagnosis, *IEEE Transactions on Power Delivery*, Vol.11, No.4, 1996, pp.1836-1841.

[2] H. Su, and J. Dang, Composite Neural Networks Applied for Identifying Stream Turbine Vibration Fault Based on Improved PSO, *WSEAS Trans. on Circuits and Systems*, Vol. 5, No.5, 2006, pp.617-624.

[3] T. Bi, Y. Ni, and F. Wu, A Novel Neural Network Approach for Fault Section Estimation, *Proceedings of the CSEE*，Vol.22, No.2, 2002, pp.73-78.

[4]H. Su, and Q. Li, Fuzzy Neural Classifier for Transformer Fault Diagnosis Based on EM Learning, Lecture Notes in Artificial Intelligence, Vol. 4114, 1997, pp. 222–229.

[5] L. Yang, Y. Shang, Y. Zhou, and Z. Yan, Probability Reasoning and Fuzzy Technique Applied for Identifying Power Transformer Malfunction, *Proceedings of the CSEE*, Vol.20, No.7, 2000, pp.19-23.

[6]H. Shu, X. Sun, and D. J. Si, An Approach to Founding and Maintaining ES Knowledge Base for Fault Diagnosis of Power Transformer, *Proceedings of the CSEE*，Vol.22, No.2, 2000, pp. 31-35.

[7] H. Su, Q. Li, Substation Fault Diagnosis Method Based on Rough Set Theory and Neural Network Model, *Power System Technology*, Vol.29, No.16, 2005, pp. 66-77.

[8]H. Su, and Q. Li, Fuzzy Neural Classifier for Fault Diagnosis of Transformer Based on Rough Set Theory, In *Proceedings of the 8th International Conference on Electrical Machines and Systems*, Nanjing, China,2005, pp.2223-2227.

[9]H. Su, and F. Zhao, A novel substation fault diagnosis approach based on RS and ANN and ES, In*Proceeding of 2006 International Conference on Communication, Circuits and Systems Proceedings*, Guilin, China, 2006, pp.2124-2127.

[10] H. Zhao, Z. Mi, and Q. Yong, Substation Fault Diagnosis Based on Petri Nets Embedding, *Automation of Electric Power Systems*, Vol.26, No.4, 2002, pp.32-35.

[11] J. Wang, Y. Ji, Application of Fuzzy Petri Nets Knowledge Repredentation in Electric Power Transformer Fault Diagnosis, *Proceedings of the CSEE*，Vol.23, No.1, 2003, pp. 121-125.

[12] Tsung-Hsi Chiang, Lan-Rong Dung, System Level Verification on High level Synthesis of Data Flow Algorithms Using Petri Net, *WSEAS*

Trans. on Circuits and Systems, Vol.5, No.6, 2006, pp.783-789.
[13] B. Zhang, K. Ma, L. Chen, and M. Zeng, The Fault Diagnosis for Substation Based on Optimizing the Combination of Fault-Masses, *Proceedings of the CSEE*, Vol.24, No.3, 2004, pp.135-139.
[14] W. Xu, D. Wang, Z. Zhou, and et al, Applying of Artificial Neural Network Combined by Genetic Algorithm in Fault Diagnosis of Power Transformers, *Proceedings of the CSEE*, Vol.17, No.2, 1997, pp.109-112.
[15] H. Su, and Q. Li, Cooperative Multi-Immune-Agent Systems for Substation Fault Diagnosis Based on Ontology, *WSEAS Trans. on Power Systems*, Vol.1, No.2, 2006, pp.446-453.
[16] Z. Pawlak, Rough Sets, *International Journal of Information and Computer Science*, Vol.11, No.5, 1982, pp.341-356.
[17]W. Zhang, *Theory and Methodology of Rough Set*, Science Press, Beijing, 2001.
[18] T. Mitchell, *Mchine Learning*, McGraw-Hill, Columnus, 1997.
[19] Y. Zhang, Z. He, J. Zhao, and et al, A Power Network Fault Diagnosis Method Based on Rough Set Theory and Naive Bayesian Networks, *Power System Technology*, Vol.31, No.1, 2007, pp.37-43.
[20]H. Su, Y. Zhang, F. Zhao, and et al, An Ensemble Deterministic Model Based on Rough Set and Fuzzy Set and Bayesian Optimal Classifier, *International Journal of Innovative Computing, Information & Control*, Vol.3, No.4, 2007, pp.977-986.
[21]H. Su, and Q. Li, Transformer Insulation Fault Diagnosis Method Based on Rough Set and Fuzzy Set and Evidence Theory, In *the* 6*th World Congress on Intelligent Control and Automation*, Dalian, China, 2006, pp.5442-5446.
[22]H. Su, A Composite Deterministic Model for Transformer Fault Diagnosis and Maintenance Based on Rough Set and Vague Set and Bayesian Optimal Classifier, *DCDIS Series-A*, Suppl., 2006, pp.1222-1227.
[23]H. Su, Multi-Source Fuzzy Information Fussion Method Based on Bayesian Optimal Classifier, *Acta Automatic Sinica*, 2008, 34(3), pp.282-287.
[24] Li, Song, Z. Hu, and Z. Yang, and C. Cai, Fault Diagnosis Based on Reasoning Integration of Rough Sets and Evidence Theory, *Trans. of Csice*, Vol. 25, No.1, 2007, pp.90-95.
[25] Y. Wang, F. Lü, and M. Li, Synthetic Fault Diagnosis Method of Power Transform Based on Rough Set Theory and Bayesian Network, *Proceedings of CSEE*, Vol. 26, No. 8, 2006, pp.137-141.
[26] Li, Song, Z. Hu, and Z. Yang, and C. Cai, Fault Diagnosis of Diesel Engine by Integrating Rough Sets and Evidence Theory of Chu Spaces, *Trans. of Csice*, Vol. 25, No.3 2007, pp.271-276.