Multi-source Fuzzy Information Fusion Method Based on Bayesian Optimal Classifier

SU Hong-Sheng[1]

Abstract To make conventional Bayesian optimal classifier possess the abilities of disposing fuzzy information and realizing the automation of reasoning process, a new Bayesian optimal classifier is proposed with fuzzy information embedded. It can not only dispose fuzzy information effectively, but also retain learning properties of Bayesian optimal classifier. In addition, according to the evolution of fuzzy set theory, vague set is also imbedded into it to generate vague Bayesian optimal classifier. It can simultaneously simulate the twofold characteristics of fuzzy information from the positive and reverse directions. Further, a set pair Bayesian optimal classifier is also proposed considering the threefold characteristics of fuzzy information from the positive, reverse, and indeterminate sides. In the end, a knowledge-based artificial neural network (KBANN) is presented to realize automatic reasoning of Bayesian optimal classifier. It not only reduces the computational cost of Bayesian optimal classifier but also improves its classification learning quality.

Key words Bayesian optimal classifier, fuzzy information, automatic reasoning, neuro-fuzzy

Bayesian optimal classifier makes the likelihood of a new instance to be correctly classified up to maximum by incorporating the posterior probabilities of all assumptions for the same hypothesis space and for the same observed data and for the same prior probabilities of these assumptions[1]. Bayesian optimal classifier may achieve the best classification results from the given data set but the algorithm possesses the two significant deficiencies. One is that it can not deal with fuzzy information effectively, the other is that the computational quantity of the algorithm is very large. To be able to dispose fuzzy information and retain learning properties of Bayesian optimal classifier, we develop a novel Bayesian optimal classifier with fuzzy information embedded. Then according to the evolution of fuzzy set theory, we respectively construct a vague Bayesian optimal classifier and set pair Bayesian optimal classifier, which can be applied to dispose the positive, reverse, and indeterminate fuzzy information, and the desired results may be achieved by weighted average. To reduce the computational cost of Bayesian optimal classifier, an optional suboptimal algorithm called Gibbs algorithm[2] is proposed, whose expected error rate is twice that of Bayesian optimal classifier at most under certain condition[3]. To reduce the computing quantity of Bayesian optimal classifier further and realize the automation of reasoning process, we propose a knowledge-based artificial neural network (KBANN)[1], which can realize automatic reasoning and dispose fuzzy information more effectively, the intelligence, robustness, and ubiquity of Bayesian optimal classifier are therefore dramatically improved.

1 Neuro-fuzzy logic

1.1 Foundation of fuzzy set[4]

Definition 1. Let V be an object space, $\forall x \in V, A \subseteq V$. To study whether x belongs to A or not, a characteristic function $\mu_A(x)$ is defined. Thus x, together with $\mu_A(x)$, constitutes a coupled pair $[x, \mu_A(x)]$. Fuzzy subset A in V may be defined as $A = \{x, \mu_A(x)|x \in V\}$, where $\mu_A(x)$ is called as fuzzy membership function of x to A, and $\mu_A(x) \in [0, 1]$.

Let A and B be two fuzzy subsets in V, and $\mu_A(x)$ and $\mu_B(x)$ be their fuzzy membership functions, respectively. Then the basic fuzzy operations are defined as follows.

$$\mu_{A\cup B}(x) = \max_{x\in V}[\mu_A(x), \mu_B(x)] = \mu_A(x) \vee \mu_B(x) \tag{1}$$

$$\mu_{A\cap B}(x) = \min_{x\in V}[\mu_A(x), \mu_B(x)] = \mu_A(x) \wedge \mu_B(x) \tag{2}$$

$$\mu_{\tilde{A}}(x) = 1 - \mu_A(x), \qquad x \in V \tag{3}$$

1.2 Logic neuron

Logic neuron, proposed in 1993[5], incudes two types of neurons, one is the OR neuron, the other is the AND neuron. They are, respectively, defined below.

Definition 2 (OR neuron). First, each input signal is logically multiplied by its connecting weight, and a logically additive operation is then implemented. Its mathematical model is expressed by

$$y = \mathrm{OR}(\boldsymbol{X}; \boldsymbol{W}) \tag{4}$$

where y is the output of the OR neuron, $\boldsymbol{X}$ is the input of the OR neuron, $\boldsymbol{X} =\{x_1, x_2, \cdots, x_n\}$, and $\boldsymbol{W}$ is the connection weight vector, $\boldsymbol{W}=\{\omega_1, \omega_2, \cdots, \omega_n\}, \omega_i \in [0, 1], i = 0, 1, \cdots, n$. (4) may also be described by

$$y = \bigvee_{i=1}^{n}[x_i \wedge \omega_i] \tag{5}$$

Definition 3 (AND neuron). Firstly, each input signal is logically added by its connecting weight, and a logically multiplicative operation is then implemented. Its mathematical model is expressed by

$$y = \mathrm{AND}(\boldsymbol{X}; \boldsymbol{W}) \tag{6}$$

where y is the output of the AND neuron, $\boldsymbol{X}$ is the input of the AND neuron, $\boldsymbol{X} = \{x_1, x_2, \cdots, x_n\}$, and $\boldsymbol{W}$ is the connection weight vector, $\boldsymbol{W} = \{\omega_1, \omega_2, \cdots, \omega_n\}, \omega_i \in [0, 1], i = 0, 1, \cdots, n$. (6) may also be described by

$$y = \bigwedge_{i=1}^{n}[x_i \vee \omega_i] \tag{7}$$

In addition, a logical processor (LP) composed of the above two types of neurons can realize more complex functions.

Received May 9, 2007; in revised form September 25, 2007
Supported by Gansu Province Natural Science Foundation (3ZS061-A25-045), and the "Qing Lan" Talent Engineering Funds of Lanzhou Jiaotong University (QL-06-19A)
1. School of Automatic and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, P. R. China
DOI: 10.3724/SP.J.1004.2008.00282

2 Bayesian optimal classifier[1]

Definition 4 (Bayesian law). Let $P(h)$ be the prior probability of the hypothesis h, where $h \in H$, and H is the hypothesis space. Let $P(D)$ denote the prior probability of the observed data D, and $P(D/h)$ specify the likelihood that D can be observed when h occurs, and $P(h/D)$ specify the likelihood of h occurrence while D is observed. $P(h/D)$ denotes the posterior probability of h, which reflects the influence of D on h. Thus, the Bayesian law may be described as

$$P(h/D) = P(D/h)P(h)/P(D) \tag{8}$$

Since D is a constant and independent of h, it follows that

$$P(h/D) \propto P(D/h)P(h) \tag{9}$$

Thus, while a new instance D occurs, the most possible classification $h \in H$ to it is called maximum posteriori (MAP) hypothesis. $h_{\rm MAP}$ can be called as MAP hypothesis only when the following formula holds.

$$h_{\rm MAP} \longleftarrow \underset{h\in H}{\arg\max} P(h/D) \tag{10}$$

Up to now, what we discuss only is which one is its most possible hypothesis when a new instance D occurs. Actually, another more interesting problem related to it is which one is the most possible classification when D occurs. For the latter, we may simply apply MAP hypothesis to get possible classification of the new instance, that is,

$$c_{\rm MAP} = \underset{c\in C}{\arg\max} P(C|h_{\rm MAP}) \tag{11}$$

where C is the classification space of the new instance, c is its possible classification, $c \in C$, and $c_{\rm MAP}$ is its most possible classification. But in fact, we still have better algorithm, i.e., Bayesian optimal classifier.

Definition 5 (Bayesian optimal classifier). Let C be the classification space of the new instance D, c be its likely classification, $c \in C$, and $P(c_j/D)$ represent the probability that the new instance D is classified as c_j. Then

$$P(c_j|D) = \sum_{h_i\in H} P(c_j|h_i)P(h_i|D) \tag{12}$$

Then, the optimal classification of the new instance D is c_j because it makes $P(c_j|D)$ up to the maximum, i.e.,

$$\arg\max \sum_{h_i\in H} P(c_j|h_i)P(h_i|D), c_j \in C \tag{13}$$

The classification system generated by (13) is called Bayesian optimal classifier. Under same conditions such as prior probabilities, hypothesis space and observed data, no other method can do better than it. According to (9), Bayesian optimal classifier may be described as

$$\arg\max \sum_{h_i\in H} P(c_j|h_i)P(D|h_i)P(h_i), \quad c_j \in C \tag{14}$$

3 Fuzzy Bayesian optimal classifier

Today, with the evolution of fuzzy set theory, more and more prior information may be expressed as fuzzy subjection function, e.g., transformer faults symptom information[6], etc. Thus, when one fault symptom information D occurs, we may use $u_{hi}(D)$ to express the subjection degree of D to h_i. $u_{hi}(D)$ may be understood as the probability that D belongs to one known symptom type h_i. Clearly, it is fully consistent with $P(h_i|D)$ in terms of numerical value. Hence, we may use $u_{h_i}(D)$ to replace $P(h_i|D)$ in (13), then we have

$$\arg\max \sum_{h_i\in H} P(c_j|h_i)\mu_{h_i}(D), \quad c_j \in C \tag{15}$$

Henceforth, for simplicity, we call Bayesian optimal classifier expressed by (15) as fuzzy Bayesian optimal classifier.

4 Vague Bayesian optimal classifier

Vague set is a natural extension of fuzzy set[7], whose core thinking is to express the two-faced characteristics of the positive and reverse of one element. Hence, vague set can simulate the thinking mode of human beings better, and tackle more complicated problems.

Definition 6 (Vague Bayesian optimal classifier)**.** Let two elements A and B be associated by u defined below.

$$u = a + bi \tag{16}$$

where a expresses the consistency of two sets to same solution, and b expresses the conflict of two sets to same solution, $a, b \in [0, 1], a + b = 1, i = -1$. After the two terms on the right-hand side in (16) are weighted by α and β, we have

$$u = \alpha a + \beta bi \tag{17}$$

where $\alpha + \beta = 1$.

Let w_{ij} express the connecting strength from the new instance i to the hypothesis j, then the possible classification of the new instance is

$$d_j = \alpha\sum_{i=1}^{n} w_{ij}a_i + \beta\sum_{i=1}^{n} w_{ij}b_i i \tag{18}$$

Compared with (15), a_i as well as b_i may be, respectively, interpreted as a probability whether the new instance belongs to the hypothesis h_i or not. Thus, Bayesian optimal classifier finds the maximized d_j in the classification space C, which can be described by

$$\arg\max d_j, \quad d_j \in C \tag{19}$$

5 Set pair Bayesian optimal classifier

On the basis of fuzzy set and vague set, researchers further proposed the set pair analysis (SPA) concept[8]. SPA is considered a new theory to the indeterminate problems solution. The method has gained a rapid advance since 1980.

To get set pair Bayesian optimal classifier, the indeterminate term c requires to be added to the right-side in (16), then we have

$$u = a + cj + bi \tag{20}$$

where $j \in [-1, 1], i = -1, a, b,$ and $c \in [0, 1]$, and $a+b+c = 1$.

After the three terms on the right-hand side in (20) are, respectively, multiplied by α, γ, and β, we get

$$u = \alpha a + \gamma cj + \beta bi \tag{21}$$

where $\alpha + \gamma + \beta = 1$.

According to (18), we then have

$$d_j = \alpha \sum_{i=1}^{n} w_{ij} a_i + \gamma \sum_{i=1}^{n} w_{ij} c_i j + \beta \sum_{i=1}^{n} w_{ij} b_i i \quad (22)$$

Then, according to (19), we find the maximized d_j.

6 KBANN for Bayes optimal classifier

KBANN is a knowledge-based artificial neural network. In KBANN, an initialized network is firstly constructed. For each likely instance, the classification given by the networks equals that done by field theories. The networks apply BP learning algorithm to modify its weights so as to fit in with training samples. It is obvious that if field theories are fully correct, the initialized assumptions will be able to classify all samples without modifying again, otherwise, the initialized assumptions requires modification to improve their fitting accuracy. The difference between conventional BP algorithm and KBANN′s lies in that the former lets the initial weight of the network be very small stochastic quantities whereas the latter be field theories, thus, the learning of KBANN will have a nice beginning, and so its generalized abilities will be stronger. KBANN may also be used to realize automatic reasoning of Bayesian optimal classifier. The advantages of applying it lie in that it can not only dispose fuzzy information but also implement parallel operation. Thus, with the aid of KBANN, Bayesian optimal classifier extends its application scope. Also, it can store field knowledge in the connecting weights of the network and update them in time from new instances learning. Clearly, it reduces the computational cost of Bayesian optimal classifier.

Below we take vague Bayesian optimal classifier as an example. According to (18) and (19), the structure of the KBANN to realize vague Bayesian optimal classifier may be designed as shown in Fig. 1.

Seen from Fig. 1, the upper part of the network may be used to realize the positive classification of all new instances, and the lower part of which may be used to realize the reverse classification of all new instances. The two parts are incorporated with in layer 3. Finally, we get the desired result by logic operation in the output layer. The learning algorithm of the network comprises two parts: one is delta law for LP, and the other is increment learning algorithm of neural networks[9]. Other than conventional BP learning algorithm, the applied error function here is an improved squared error function called TANGPROP algorithm[10], which is expressed by

$$E = \sum_i \left\{ (f(x_i) - f'(x_i))^2 + \mu[(\partial f(x)/\partial(x) - \partial f'(x)/\partial(x))|x = x_i]^2 \right\} \quad (23)$$

In the above equation, x_i is the ith instance, $f(x_i)$ is the aim function value, $f'(x_i)$ is the practical output, and μ is a constant defined by user, which is used to scale the importance between the fitting training data and the differential coefficients of the fitting training data. Note that the first term on the right-side in (23) is the original squared error function, whereas the second term is a newly added squared error, which is used to fit in with the differential coefficients of the training data. Thus, it improves the fitting accuracy of the training data. Also, since all knowledge is stored in connecting weights of the network, the learning process of the network is dynamic.

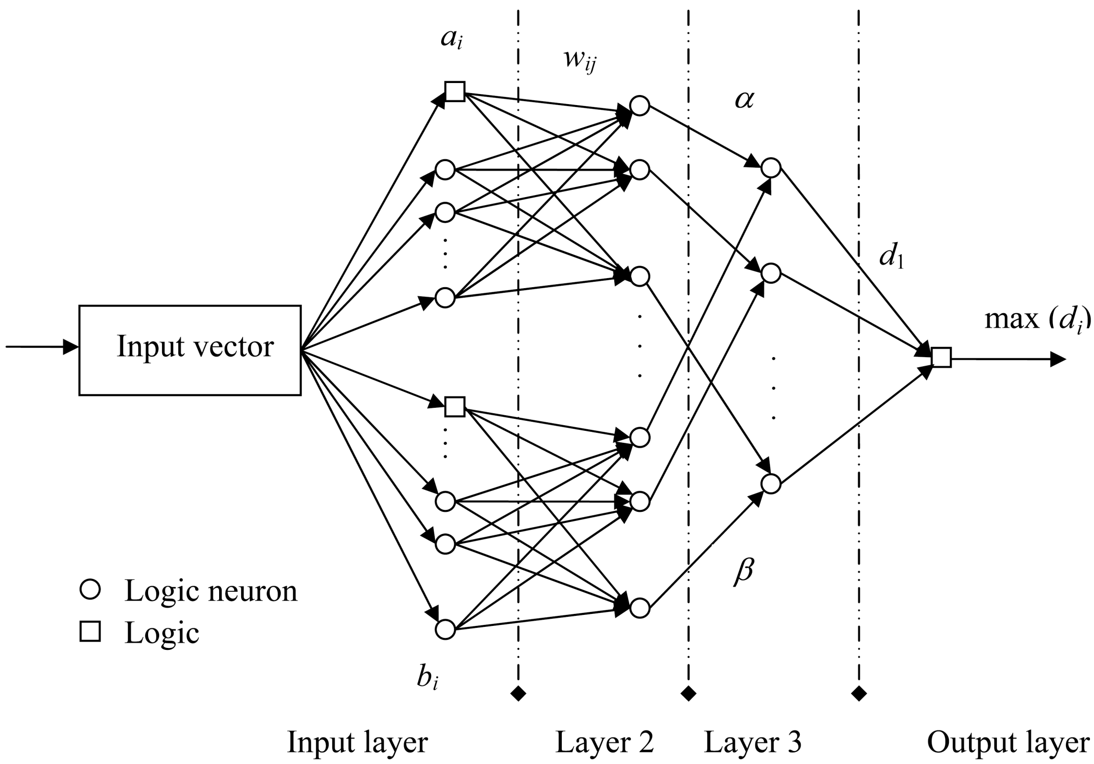


Fig. 1 KBANN based vague Bayesian optimal classifier

Likewise, in light of (19) and (22), we may construct the set pair Bayesian optimal classifier.

7 Examples

There are 296 data samples in all in the diagnostic knowledge base of one transformer, and the fault symptoms and the relevant fuzzy membership functions are described in Table 1[11].

According to the positive and negative fuzzy membership functions in Table 1, a_i and b_i in (18) may be worked out. Thus, for a fuzzy fault symptom information, we may work out its positive and reverse influence coefficients. Fuzzy membership functions here adopt the ascending or the falling semi-trapezoid distribution, which are, respectively, expressed as $\mu^{\uparrow}(a,b,x)$ and $\mu_{\downarrow}(a,b,x)$, and described below.

1) Ascending semi-trapezoid distribution

$$\mu^{\uparrow}(a,b,x)=\begin{cases}0 & \text{if } x\le a\\ (x-a)/(b-a) & \text{if } a<x \text{ and } x\le b\\ 1 & \text{if } x>b\end{cases}$$

2) Falling semi-trapezoid distribution

$$\mu^{\downarrow}(a,b,x)=\begin{cases}1 & \text{if } x\le a\\ (b-x)/(b-a) & \text{if } a<x \text{ and } x\le b\\ 0 & \text{if } x>b\end{cases}$$

Based on expert experience and prior knowledge, the connecting weight w_{ij} from symptom i to fault source j is established in Table 2.

Table 1 Fault symptoms and fuzzy membership functions accordingly

Fault symptom type		Positive fuzzy membership function (M^+)	Negative fuzzy membership function (M^-)
m_1 : Three-ratio-code based heat fault characteristics	C_2H_2/C_2H_4 ($A1$)	$\mu_{\downarrow}(0.08, 0.12, x)$	$\mu_{\uparrow}(0.08, 0.12, x)$
	CH_4/H_2 ($B1$)	$\mu_{\uparrow}(0.8, 1.2, x)$	$\mu_{\downarrow}(0.8, 1.2, x)$
	φ_{H_2}(x10^{-6}) ($C1$)	$\mu_{\uparrow}(120, 180, x)$	$\mu_{\downarrow}(120, 180, x)$
	$\varphi_{C_2H_2}$(x10^{-6}) ($D1$)	$\mu_{\uparrow}(4, 6, x)$	$\mu_{\downarrow}(4, 6, x)$
	$\varphi_{C_1+C_2}$(x10^{-6}) ($E1$)	$\mu_{\uparrow}(120, 180, x)$	$\mu_{\downarrow}(120, 180, x)$
	Generating gas open type	$\mu_{\uparrow}(0.2, 0.3, x)$	$\mu_{\downarrow}(0.2, 0.3, x)$
	Rate($F1$)/(ml/h) close type	$\mu_{\uparrow}(0.4, 0.6, x)$	$\mu_{\downarrow}(0.4, 0.6, x)$
	$A1\cap B1\cap(C1\cup D1\cup E1\cup F1)$		
m_2/(mg/L): Water capacity in transformer oil	110 kV downwards	$\mu_{\uparrow}(28, 42, x)$	$\mu_{\downarrow}(28, 42, x)$
	110 kV upwards	$\mu_{\uparrow}(20, 30, x)$	$\mu_{\downarrow}(20, 30, x)$
m_3: Earth current		$\mu_{\uparrow}(0.196, 0.144, x)$	$\mu_{\downarrow}(0.196, 0.144, x)$
m_4: Three-phase imbalance coefficient	1.6 MVA downwards	$\mu_{\uparrow}(0.032, 0.048, x)$	$\mu_{\downarrow}(0.032, 0.048, x)$
	1.6 MVA upwards	$\mu_{\uparrow}(0.016, 0.024, x)$	$\mu_{\downarrow}(0.016, 0.024, x)$
m_5(pC): Local discharge capacity		$\mu_{\uparrow}(300, 900, x)$	$\mu_{\downarrow}(300, 900, x)$
m_6 : Three-ratio-code based discharge fault characteristics	C_2H_2/C_2H_4 ($A2$)	$\mu_{\uparrow}(0.08, 0.12, x)$	$\mu_{\downarrow}(0.08, 0.12, x)$
	CH_4/H_2 ($B2$)	$\mu_{\downarrow}(0.8, 1.2, x)$	$\mu_{\uparrow}(0.8, 1.2, x)$
	$\varphi_{H_2}(\times10^{-6})$ ($C2$)	$\mu_{\uparrow}(120, 180, x)$	$\mu_{\downarrow}(120, 180, x)$
	$\varphi_{C_2H_2}(\times10^{-6})$ ($D2$)	$\mu_{\uparrow}(4, 6, x)$	$\mu_{\downarrow}(4, 6, x)$
	$\varphi_{C_1+C_2}(\times10^{-6})$ ($E2$)	$\mu_{\uparrow}(120, 180, x)$	$\mu_{\downarrow}(120, 180, x)$
	Generating gas open type	$\mu_{\uparrow}(0.2, 0.3, x)$	$\mu_{\downarrow}(0.2, 0.3, x)$
	Rate($F2$)/(ml/h) close type	$\mu_{\uparrow}(0.4, 0.6, x)$	$\mu_{\downarrow}(0.4, 0.6, x)$
	$A2\cap B2\cap(C2\cup D2\cup E2\cup F2)$		
m_7: Absolute value of the warp of winding transformation ratio	Rated tapping	$\mu_{\uparrow}(0.004, 0.006, x)$	$\mu_{\downarrow}(0.004, 0.006, x)$
$m_8(CO/CO_2)$	A < 0.09	$\mu_{\downarrow}(0.072, 0.018, x)$	$\mu_{\uparrow}(0.072, 0.018, x)$
	B > 0.33	$\mu_{\uparrow}(0.264, 0.396, x)$	$\mu_{\downarrow}(0.264, 0.396, x)$
		$A\cup B$	
m_9	Winding absorption	$\mu_{\downarrow}(1.04, 1.56, x)$	$\mu_{\uparrow}(1.04, 1.56, x)$
	Winding polarization index	$\mu_{\downarrow}(1.2, 1.8, x)$	$\mu_{\uparrow}(1.2, 1.8, x)$

Table 2 Fault types and connection strengths

Symptom type	Fault type	$w_{i,j}$
m_1		0.82
m_3	d_1: Multi-point earth or local short circuit in iron core	0.90
m_5		0.19
m_6		0.30
m_1		0.71
m_5	d_2: Leak magnetism heating or overheat	0.35
m_6		0.29
m_1		0.22
m_2	d_3: Insulating aging	0.27
m_8		0.82
m_2	d_4: Insulating damp	0.72
m_9		0.75
m_1		0.67
m_4	d_5: Tapping switch or down-lead fault	0.87
m_6		0.23
m_5	d_6: Suspend discharge	0.90
m_6		0.86
m_1		0.15
m_5		0.75
m_6	d_7: Winding distortion and circle short	0.68
m_7		0.80
m_8		0.72
m_5		0.90
m_6	d_8: Circle short and insulation damage	0.52
m_7		0.80
m_8		0.68
m_2		0.42
m_5	d_9: Encloser discharge	0.90
m_6		0.88
m_8		0.76

Seen from Tables 1 and 2, the system possesses nine types of fault symptoms and nine types of fault sources. They may, respectively, serve as the inputs and outputs of the KBANN in Fig. 1. Fuzzy membership functions in Table 1 may act as the base functions of fuzzy neurons in the input layer. For fuzzy subjection degrees of m_1 and m_6 in Table 1, we may adopt LP to generate them. Connecting strength $\boldsymbol{W}$ in Table 2 may serve as prior weights from input layer to classification space. The two parameters α and β in the third layer may be modified in the process of learning. Thus, we may use the model in Fig. 1 to realize the fault diagnosis of the transformer. The overall process is described as follows.

1) Applying LP to realize the fuzzy subjection degrees of m_1 and m_6.

To get the subjection degree of m_1, we must calculate $\mu_{A1\cap B1\cap(C1\cup D1\cup E1\cup F1)}$. According to (5) and (7), there are two steps needed to complete it.

First, applying the OR neuron to realize $\mu_{\mathrm{C1\cup D1\cup E1\cup F1}}$. Hence, in (5) let x_i equal $\mu_{C1}, \mu_{D1}, \mu_{E1}$, and μ_{F1}, respectively, $\omega_i = 1$; then

$$\mu_H = \bigvee_{i=1}^{4}[x_i \wedge \omega_i] = \bigvee_{i=1}^{4}[x_i \wedge 1] = \mu_{C1\cup D1\cup E1\cup F1} \tag{24}$$

Second, applying the AND neuron to realize $\mu_{A1\cap B1\cap H}$. Therefore, in (7), let x_i equal μ_{A1}, μ_{B1}, and μ_H, respectively, $\omega_i = 0$. Then

$$\mu_{m1} = \bigwedge_{i=1}^{3}[x_i \vee \omega_i] = \bigwedge_{i=1}^{3}[x_i \vee 0] = \mu_{A1\cap B1\cap H} = \mu_{A1\cap B1\cap(C1\cup D1\cup E1\cup F1)} \tag{25}$$

According to the above two steps, LP of calculating the fuzzy subjection degree of m_1 is shown in Fig. 2. Clearly, for m_6, we also can calculate its fuzzy subjection degree in the same manner.

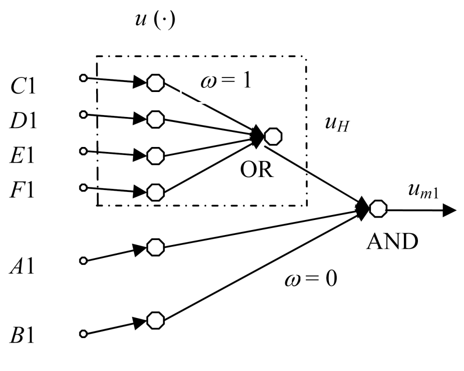


Fig. 2 LP for fuzzy subjection degree of m_1

2) The weight vector $\boldsymbol{W}$

Fig. 1 indicates a fully-connected neural network structure. The field knowledge in Tables 1 and 2 may act as the initial weight values of the network. If prior information in Tables 1 and 2 is fully correct, it will be unnecessary to train the network and will become necessary otherwise. For simplicity, here, we set $u = 0$ in (23).

3) The weights of the positive and reverse instances

In general, the weight of the positive inference is larger than that of the reverse inference. Hence, we set $\alpha = 0.85$, and $\beta = 0.15$ in (18). They can be modified in the process of learning.

Example. The data ($\times 10^{-6}$) of the dissolved gas analysis (DGA) in one transformer are described by $\varphi(H_2) = 70.4, \varphi(CH_4) = 69.5, \varphi(C_2H_6) = 28.9, \varphi(C_2H_2) = 10.4, \varphi(C_2H_4) = 241.2, \varphi(CO) = 704$, and $\varphi(CO_2) = 3350$, the unbalanced coefficient of the winding is described by 0.019, and the earth current of the iron core is described by 0.1A. The ratios between the characteristic gases are calculated by $\varphi(C_2H_2)/\varphi(C_2H_4) = 0.043, \varphi(CH_4)/\varphi(H_2) = 0.99, \varphi(C_2H_4)/\varphi(C_2H_6) = 8.35, \varphi(CO)/\varphi(CO_2) = 0.21$, and the three-ratio-code is, therefore, calculated as 002. However, the code does not exist in the three-ratio-code table. It is, therefore, difficult to identify the fault type of the transformer. However, using our method, through fuzzy operation, we have $M^+ = 0.475/m_1 + 0.083/m_3 + 0.375/m_4$, and $M^- = 0.525/m_6 + 0.917/m_3 + 0.625/m_4 + 1/m_8$. Suppose that the prior knowledge in Tables 1 and 2 is fully correct, according to the networks model in Fig. 1, set $\alpha = 0.85$, and $\beta = 0.15$. The likelihood of each fault occurrence is then calculated by $d_1 = 0.249, d_2 = 0.275, d_3 = 0.064, d_4 = 0, d_5 = 0.495, d_6 = -0.033, d_7 = -0.055, d_8 = -0.159$, and $d_9 = -0.1833$. According to (19), since $d_5 = \max\{d_i, i = 1, 2, \cdots, 9\}$, the most likely diagnostic result is d_5, i.e., tapping switch or down-lead fault. Finally, fielded practical checking proves the correctness of the diagnostic result. This result is fully consistent with that in [11], however, here the applied approach is quite different. And also, the experiment result shows that the proposed method can not only work out the occurrence probabilities of all likely faults but also gives out those of all unlikely faults, and it, therefore, sees problem more fully. The flaw is that the experiment result covers more data, which makes problem analysis become complex.

8 Conclusion

Bayesian optimal classifier based on fuzzy set theory can deal with the fuzziness and indetermination of the observed information, and effectively tackle the "bottle neck" puzzle in fuzzy knowledge acquisition. Meanwhile, it provides a multi-source fuzzy information fusion method based on Bayesian optimal classifier. In theory, fuzzy membership function itself represents a sort of prior information similar to prior probability, hence, the combination of the two is feasible. And also, the paper applies neural networks to simulate Bayesian optimal classifier and verifies the correctness of the proposed model. This will be a new orientation for automatic computing.

References

1 Mitchell T M. *Machine Learning*. Columbus: McGraw-Hill, 1997. 112−140

2 Opper M, Haussler D. Generalization performance of Bayesian optimal classification algorithm for learning a perceptron. *Physical Review Letters*, 1991, **66**(20): 2677−2680

3 Haussler D, Kearns M, Schapire R E. Bounds on the sample complexity of Bayesian learning using information theory and the VC dimension. *Machine Learning*, 1994, **14**(1): 83−113

4 Yang Jun, Feng Zhen-Sheng, Huang Kao-Li, Li Yan, Zhang Yan-Sheng, Jia Hai-Peng. *Intelligent Fault Diagnosis Technology for Equipments*. Beijing: National Defence Press, 2004. 74−118 (in Chinese)

5 Wang Yao-Nan. *Intelligent Information Processing*. Beijing: Higher Education Press, 2003. 192−230 (in Chinese)

6 Su Hong-Sheng. A composite deterministic model for transformer fault diagnosis based on rough set and vague set and Bayesian optimal classifier. *Dynamics of Continuous, Discrete and Impulsive Systems, Series A*, 2006, **13**(Sup.): 1222−1227

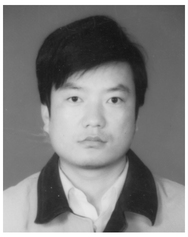
7 Guo W L, Buehrer D J. Vague sets. *IEEE Transactions on System, Man, and Cybernetic*, 1993, **23**(2): 610−614

8 Zhao Ke-Qin. *Set Pair Analysis and Applications*. Hangzhou: Zhejiang Science Press, 2000 (in Chinese)

9 Zeng Huang-Lin. *Intelligent Calculation*. Chongqing: Chongqing University Press, 2004 (in Chinese)

10 Simard P, Victorri B, LeCun Y, Denker J S. Tangent prop-a formalism for specifying selected invariances in an adaptive network. *Advances Neural Information Processing System*. San Mateo, USA: Morgan Kaufmann, 1992. 895−903

11 Su H S, Mi G S. Set pair analysis applied for identifying power transformer faults. In: Proceedings of 2006 International Conference on Machine Learning and Cybernetics. Dalian, China: IEEE, 2006. 1708−1713



SU Hong-Sheng Ph. D. candidate at School of Electrical Engineering, Southwest Jiaotong University. He received his bachelor degree from Lanzhou Jiaotong University in 2001. His research interest covers natural computing and machine learning. E-mail: shsen@163.com