

A New Algorithm of Closeness Degree for Fuzzy Pattern Recognition

Zhigang Li¹, Lingling Li^{1*}, Fenfen Zhu¹, Quanming Zhao²

¹ School of Electrical Engineering, Hebei University of Technology

² School of Information Engineering, Hebei University of Technology

* Corresponding author's Email: lilingling@hebut.edu.cn

Abstract: There are two basic methods of fuzzy pattern recognition: one is direct method based on membership degree, and another is indirect method based on closeness degree. In this paper, different algorithms of closeness degree have been studied and a new method for calculating the closeness degree has been introduced. By combining the concepts of membership degree and closeness degree, the pattern characteristic values derived from both real numbers and fuzzy set expressions can be handled, then soft transition from direct method to indirect method can be realized by the new algorithm; It also could adjust every kind of weights of closeness degree to adapt itself based on the characteristics of classical model and be recognized model. And some case studies would be provided to demonstrate the effectiveness and availability for the new algorithm. The algorithm could also be used in different circumstances to deal with closeness degree other than pattern recognition.

Keywords: fuzzy pattern recognition; fuzzy algorithm; closeness degree; membership degree

1. Introduction

As the emergence of the computer in the 1940s and the rise of artificial intelligence in the 1950s, pattern recognition has become a major branch of artificial intelligence. Pattern recognition is often called pattern classification, which deals with object classifying according to the features of the objects. In recent years, the new pattern recognition methods have been constantly emerging, such as Artificial Neural Networks [1, 2], GA [3], D-S evidence theory [4, 5] and cluster analysis [6]. As for fuzzy pattern recognition, there are two methods, the direct method with the maximum membership principle based on the concept of membership degree and the indirect method using nearest neighbor principle based on the concept of closeness degree [7, 8]. In other fields, closeness degree and membership degree are the most commonly used fuzzy measures, such as fault diagnosis [9], product design [10, 11] and fuzzy comprehensive evaluation

[12] etc.

The current closeness degree algorithms have certain adaptability, and it must be applied to particular occasions. Therefore, in order to avoid error recognition, different algorithms are used in different occasions. The model to be recognized (called model A in the following content) may be a real number set or fuzzy set. A bridge between them is built in this article which is establishing a close degree algorithm to handle two types of data simultaneously.

2. Common algorithms of closeness degree

2.1 Lattice closeness degree

The lattice closeness degree is the closeness degree based on the inner product and outer product.

Assumption: there are two fuzzy sets on the universe X , denoted by $\tilde{A}(x) = u_{\tilde{A}}(x)$, $\tilde{B}(x) = u_{\tilde{B}}(x)$, if their kernel and support set meet: $\ker \tilde{A} \neq \Phi$, $\ker \tilde{B} \neq \Phi$, $\text{supp} \tilde{A} \neq X$, $\text{supp} \tilde{B} \neq X$, then the following two

equations are the lattice closeness degree of \tilde{A}, \tilde{B} .

$$\eta_1(\tilde{A}, \tilde{B}) = \frac{1}{2}[\tilde{A} \circ \tilde{B} + (1 - \tilde{A} \odot \tilde{B})] \quad (1)$$

$$\eta_1(\tilde{A}, \tilde{B}) = (\tilde{A} \circ \tilde{B}) \wedge (1 - \tilde{A} \odot \tilde{B}) \quad (2)$$

Where, $\tilde{A} \circ \tilde{B} = \bigvee_{x \in X} (\tilde{A}(x) \wedge \tilde{B}(x))$, is the inner product of \tilde{A}, \tilde{B} ; $\tilde{A} \odot \tilde{B} = \bigwedge_{x \in X} (\tilde{A}(x) \vee \tilde{B}(x))$, is the outer product of \tilde{A}, \tilde{B} .

2.2 The closeness degree based on integral

If the universe X is continuous, then we have

$$\eta_2(\tilde{A}, \tilde{B}) = \frac{\int_X (\tilde{A}(x) \wedge \tilde{B}(x)) dx}{\int_X (\tilde{A}(x) \vee \tilde{B}(x)) dx} \quad (3)$$

and

$$\eta_2(\tilde{A}, \tilde{B}) = \frac{2 \int_X (\tilde{A}(x) \wedge \tilde{B}(x)) dx}{\int_X \tilde{A}(x) dx + \int_X \tilde{B}(x) dx} \quad (4)$$

where the integral is the Lebesgue integral; if universe $X = R$, the integral is the Riemann integral.

If the universe of is discrete, then we have:

$$\eta_2(\tilde{A}, \tilde{B}) = \frac{\sum_{i=1}^s (\tilde{A}(x_i) \wedge \tilde{B}(x_i))}{\sum_{i=1}^s (\tilde{A}(x_i) \vee \tilde{B}(x_i))} \quad (5)$$

and

$$\eta_2(\tilde{A}, \tilde{B}) = \frac{2 \sum_{i=1}^s (\tilde{A}(x_i) \wedge \tilde{B}(x_i))}{\sum_{i=1}^s (\tilde{A}(x_i) + \sum_{i=1}^s \tilde{B}(x_i))} \quad (6)$$

2.3 The closeness degree based on distance

The closeness degrees based on Hamming distance and Euclidean distance can be formulated as:

$$\eta_3(\tilde{A}, \tilde{B}) = 1 - \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} |\tilde{A}(x) - \tilde{B}(x)| dx \quad (7)$$

and

$$\eta_3(\tilde{A}, \tilde{B}) = 1 - \left(\frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} (\tilde{A}(x) - \tilde{B}(x))^2 dx \right)^{\frac{1}{2}} \quad (8)$$

where the universe X is closed interval $[\alpha, \beta]$ of real number field.

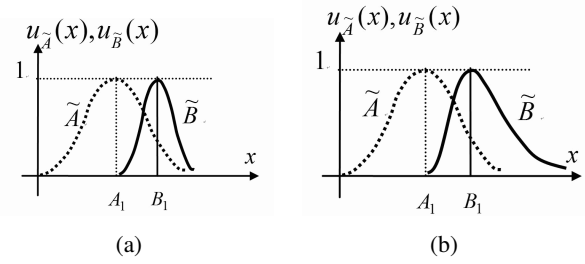


Figure 1 Two fuzzy sets

2.4 Analysis

From Fig. 1, it is shown that two fuzzy sets of \tilde{A}, \tilde{B} , $A_1 = \ker \tilde{A}$, $B_1 = \ker \tilde{B}$; x^* is the intersection of the membership functions between \tilde{A} and \tilde{B} . In Fig1(a), it is obvious that $\eta_1(\tilde{A}, \tilde{B}) \gg \eta_2(\tilde{A}, \tilde{B})$; when moving \tilde{B} gradually towards left to make \tilde{B} coincide with \tilde{A} , then $\eta_1(\tilde{A}, \tilde{B}) = 1$, but $\eta_2(\tilde{A}, \tilde{B})$ is still small. When the right distribution of \tilde{B} gradually moving towards $+\infty$ (See Fig1(b)), $\eta_3(\tilde{A}, \tilde{B})$ always keeps invariant, however, $\eta_2(\tilde{A}, \tilde{B})$ is constantly decreasing, until tending to 0.

As it can be seen, for the same question, totally different conclusions can be obtained by using different algorithms of closeness degree. Assume that \tilde{A} and \tilde{B} in Figure 1 represent 1D fuzzy patterns, using the threshold principle to determine whether it is the same category (assume the threshold value $\lambda = 0.5$). The answer is positive according to $\eta_1(\tilde{A}, \tilde{B})$; whereas it is negative according to $\eta_2(\tilde{A}, \tilde{B})$. The divergence of judgment is because of the essences reflected by the two algorithms of closeness degree are different. The analysis will be done by using terms from Genetic Algorithms (GA) in this paper.

Considering the universe X of \tilde{A}, \tilde{B} as an evolutionary population of GA, and the very element in X as an individual of the population, then $\eta_1(\tilde{A}, \tilde{B})$ just represents the quality of the optimal one while the quality of entire population is ignored. It can be seen from Fig. 1(b) that $\eta_1(\tilde{A}, \tilde{B}) = u_{\tilde{A}}(x^*) = u_{\tilde{B}}(x^*)$ is the membership degree of element x^* to fuzzy set \tilde{A}, \tilde{B} and also has the same meaning with the adaption degree of optimal individual in GA. Here, x^* is the intersection of the membership functions of \tilde{A}, \tilde{B} which is equivalent to the optimal individual in the population. Therefore, $\eta_1(\tilde{A}, \tilde{B})$ describes the similarity of \tilde{A}, \tilde{B} which is solely depending on membership degree. It reflects the concept of membership degree.

In the opposite, $\eta_2(\tilde{A}, \tilde{B})$ only represents the population characteristic without considering the optimal individual characteristic. This difference is the real

reason of divergence between using Eqs. (1)-(2) and Eqs. (3)-(4).

Just like constructing adaption degree function which emphasizes GA, if an algorithm can be constructed of closeness degree which is able to make $\eta_1(\tilde{A}, \tilde{B})$ compatible with $\eta_2(\tilde{A}, \tilde{B})$, then error recognition would be sufficiently reduced.

3. The proposed closeness degree algorithm

A more convenient and simple calculation method is proposed in this paper, and it integrates the advantages of the integral closeness degree and the distance closeness degree:

$$\eta_4(\tilde{A}, \tilde{B}) = 1 - \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} \frac{|u_{\tilde{A}}(x) - u_{\tilde{B}}(x)|}{(u_{\tilde{A}}(x) \vee u_{\tilde{B}}(x))} dx \quad (9)$$

where $u_{\tilde{A}}(x)$ and $u_{\tilde{B}}(x)$ are the membership functions of \tilde{A}, \tilde{B} .

For the different distribution of the support sets $supp\tilde{A}$ and $supp\tilde{B}$ of the fuzzy sets \tilde{A}, \tilde{B} in universe, the following contents discuss the closeness degree algorithm of \tilde{A}, \tilde{B} in different situations .

3.1 Both continuous intervals $supp\tilde{A}$ and $supp\tilde{B}$

Assumption: There are two convex fuzzy sets \tilde{A}, \tilde{B} on the universe $X = R$, whose membership functions are $u_{\tilde{A}}(x)$ and $u_{\tilde{B}}(x)$, and their support sets $supp\tilde{A}$ and $supp\tilde{B}$ are the continuous intervals on universe, then:

$$supp\tilde{A} = \langle A_0^*, A_0^{**} \rangle = \{x | u_{\tilde{A}}(x) > 0, x \in X\}$$

$$supp\tilde{B} = \langle B_0^*, B_0^{**} \rangle = \{x | u_{\tilde{B}}(x) > 0, x \in X\}$$

where $\langle *, ** \rangle$ represents the open interval, closed interval or half open half closed interval, and holds the same meaning as following.

$$\begin{aligned} \lambda_{\tilde{A}, \tilde{B}} &= \frac{|supp\tilde{A}| - |supp\tilde{B}|}{|supp\tilde{A}| + |supp\tilde{B}|} \\ &= \frac{A_0^{**} - A_0^* - B_0^{**} + B_0^*}{A_0^{**} + B_0^{**} - A_0^* - B_0^*} \end{aligned} \quad (10)$$

where $supp\tilde{A}$ and $supp\tilde{B}$ are the 0-strong cuts of \tilde{A}, \tilde{B} , and the support sets of \tilde{A}, \tilde{B} ; $|supp\tilde{A}| = A_0^{**} - A_0^*$, $|supp\tilde{B}| = B_0^{**} - B_0^*$, the geometric meaning is the width of the support set. It is defined that A_0^* and A_0^{**} are not equal to $-\infty$, A_0^{**} and B_0^{**} are not equal to $+\infty$, to ensure $supp\tilde{A}$ and $supp\tilde{B}$ are limited intervals.

Based on the definition above, a closeness degree algorithm is proposed as following:

$$\eta(\tilde{A}, \tilde{B}) = |\lambda_{\tilde{A}, \tilde{B}}| \cdot \eta_1(\tilde{A}, \tilde{B}) + (1 - |\lambda_{\tilde{A}, \tilde{B}}|) \cdot \eta_2(\tilde{A}, \tilde{B}) \quad (11)$$

When the kernels of and are close enough, the closeness degree algorithm is:

$$\begin{aligned} \eta(\tilde{A}, \tilde{B}) &= \frac{\lambda_{\tilde{A}, \tilde{B}} + |\lambda_{\tilde{A}, \tilde{B}}|}{2} \eta_1(\tilde{A}, \tilde{B}) \\ &\quad + (1 - \frac{\lambda_{\tilde{A}, \tilde{B}} + |\lambda_{\tilde{A}, \tilde{B}}|}{2}) \eta_2(\tilde{A}, \tilde{B}) \end{aligned} \quad (12)$$

In the fuzzy pattern recognition, $\eta(\tilde{A}, \tilde{B})$ will be used to calculate the similarity between the model to be recognized (Model A) and typical model (called Model B in the following content). Among them, \tilde{A} and \tilde{B} are the characteristic values.

In Eq. (11), $\lambda_{\tilde{A}, \tilde{B}}$ and $1 - \lambda_{\tilde{A}, \tilde{B}}$ can be regarded as the weights of $\eta_1(\tilde{A}, \tilde{B})$ and $\eta_2(\tilde{A}, \tilde{B})$, so is called the weight parameter.

In Eq. (12), the weight parameter of $\eta_1(\tilde{A}, \tilde{B})$ and $\eta_2(\tilde{A}, \tilde{B})$ are $0.5(\lambda_{\tilde{A}, \tilde{B}} + |\lambda_{\tilde{A}, \tilde{B}}|)$ and $1 - 0.5(\lambda_{\tilde{A}, \tilde{B}} + |\lambda_{\tilde{A}, \tilde{B}}|)$.

Assumption: $\tilde{A}, \tilde{B}, \tilde{C}$ are on the universe X , the following equations can be proved:

- 1) $\eta(\tilde{A}, \tilde{B}) = \eta(\tilde{B}, \tilde{A})$;
- 2) $\eta(\tilde{A}, \tilde{A}) = 1, \eta(\tilde{A}, \Phi) = \eta(\tilde{B}, \Phi) = \eta(\tilde{C}, \Phi) = 0$;
- 3) $C \subseteq B \subseteq A \Rightarrow \eta(\tilde{A}, \tilde{C}) \leq \eta(\tilde{A}, \tilde{B}) \wedge \eta(\tilde{B}, \tilde{C})$;

where η is the closeness degree function on the universe X , and $\eta(\tilde{A}, \tilde{B})$ is the closeness degree of \tilde{A} and \tilde{B} .

There is a more universal closeness degree algorithm:

$$\begin{aligned} \eta(\tilde{A}, \tilde{B}) &= f(|A_{\varepsilon}|, |B_{\varepsilon}|) \cdot \eta'(\tilde{A}, \tilde{B}) \\ &\quad + (1 - f(|A_{\varepsilon}|, |B_{\varepsilon}|)) \cdot \eta''(\tilde{A}, \tilde{B}) \\ &= \lambda_{\tilde{A}, \tilde{B}} \cdot \eta'(\tilde{A}, \tilde{B}) + (1 - \lambda_{\tilde{A}, \tilde{B}}) \cdot \eta''(\tilde{A}, \tilde{B}) \end{aligned} \quad (13)$$

where $\eta'(\tilde{A}, \tilde{B})$ and $\eta''(\tilde{A}, \tilde{B})$ express $\eta_1(\tilde{A}, \tilde{B})$ and $\eta_2(\tilde{A}, \tilde{B})$ respectively, or $\eta_1(\tilde{A}, \tilde{B})$ and $\eta_4(\tilde{A}, \tilde{B})$ in Eqs. (1)-(9).

For the fuzzy sets of normal type, Kosit type and Γ type, the range of support set can not be guaranteed within finite intervals, so $\lambda_{\tilde{A}, \tilde{B}}$ can not be calculate by Eq. (10). This paper presents a general algorithm:

$$\begin{aligned} f(|A_{\varepsilon}|, |B_{\varepsilon}|) &= \frac{|A_{\varepsilon}| - |B_{\varepsilon}|}{|A_{\varepsilon}| + |B_{\varepsilon}|} \\ &= \frac{A_{\varepsilon}^{**} - A_{\varepsilon}^* - B_{\varepsilon}^{**} + B_{\varepsilon}^*}{A_{\varepsilon}^{**} + B_{\varepsilon}^{**} - A_{\varepsilon}^* - B_{\varepsilon}^*} \end{aligned} \quad (14)$$

where $|A_{\varepsilon}|, |B_{\varepsilon}|$ are the ε -cuts of \tilde{A}, \tilde{B} , then the closeness degree can be calculated by Eqs. (11)-(12).

As long as \tilde{A} , \tilde{B} do not belong to the higher semi-trapezoid distribution or the lower semi-trapezoid distribution, choose the appropriate value to ensure $|A_\varepsilon|$ and $|B_\varepsilon|$ as finite intervals, it can use Eqs. (7)-(8) to calculate the closeness degree, which is no longer limited by Eq. (10), requesting A_ε^* and B_ε^* not equal to $-\infty$ as well as A_ε^{**} and B_ε^{**} not equal to $+\infty$.

Lower semi-trapezoid distribution is also known as partial small, which applies to the membership function when x is small (shown in Fig. 2(a)); Higher semi-trapezoid distribution is also known as partial large, which applies to the membership function when is larger (shown in Fig. 2(b)).

But in engineering, the higher semi-trapezoid distribution or the lower semi-trapezoid distribution membership functions do exist; improper handling can cause error recognition. For explaining this method, here, remember: a typical pattern set is $\tilde{B}^{(1)}, \tilde{B}^{(2)}, \dots, \tilde{B}^{(n)}$ (the model features dimension is 1), where $\tilde{B}^{(1)}$ is a lower semi-trapezoid distribution, $\tilde{B}^{(n)}$ is a higher semi-trapezoid distribution; model \tilde{A} is recognized neither as the lower semi-trapezoid distribution nor higher semi-trapezoid distribution.

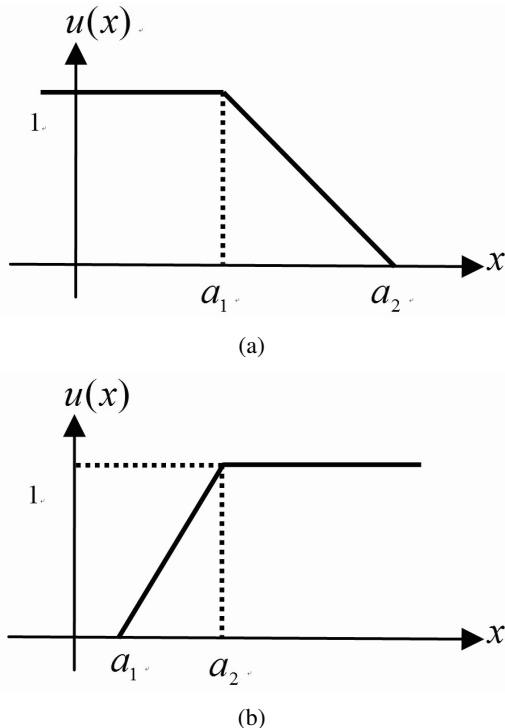


Figure 2 Lower semi-trapezoid and higher semi-trapezoid

The ε -cut sets of $\tilde{B}^{(1)}$ and $\tilde{B}^{(n)}$ are denoted as $B^{(1)}_\varepsilon = [*, \beta^{(1)}]$ and $B^{(n)}_\varepsilon = [\beta^{(n)}, **]$. Here, $*$ is a real number less than $\beta^{(1)}$, depending on the universe of $\tilde{B}^{(1)}$; $**$ is a real number more than $\beta^{(n)}$, depending on the universe of $\tilde{B}^{(n)}$ as well. And the ε -cut set of

$A_\varepsilon = [\alpha', \alpha'']$, where $* \leq \alpha' \leq \alpha'' \leq **$, then the following statements can be obtained:

- A If $\beta^{(1)} \leq \alpha'$, then $\eta(\tilde{A}, \tilde{B}^{(1)}) = 0$; if $\beta^{(1)} > \alpha'$, redefine the universe of $\tilde{B}^{(1)}$ as $[\alpha', \beta^{(1)}]$, $B^{(1)}_\varepsilon = [\alpha', \beta^{(1)}]$ can be got, then calculate $\lambda_{\tilde{A}, \tilde{B}}$ by Eq. (10); when calculating $\eta(\tilde{A}, \tilde{B}^{(1)})$ by Eqs. (11)-(12), $\tilde{B}^{(1)}$ adopts the new universe $[\alpha', \beta^{(1)}]$.
- B If $\beta^{(n)} \geq \alpha''$, then $\eta(\tilde{A}, \tilde{B}^{(n)}) = 0$; if $\beta^{(n)} \leq \alpha''$, redefine the universe of $\tilde{B}^{(n)}$ as $[\beta^{(n)}, \alpha'']$, $B^{(n)}_\varepsilon = [\beta^{(n)}, \alpha'']$ can be got, then calculate $\lambda_{\tilde{A}, \tilde{B}}$ by Eq. (10); when calculating $\eta(\tilde{A}, \tilde{B}^{(n)})$ by Eqs. (11)-(12), $\tilde{B}^{(n)}$ adopts the new universe $[\beta^{(n)}, \alpha'']$. There are three special instructions about the conclusions above:
 - a) If the interval width of support set $supp\tilde{B}^{(1)}$ $|supp\tilde{B}^{(1)}| \neq +\infty$ and $|supp\tilde{A}| \neq +\infty$, set $\varepsilon = 0^+$ to ensure $B_{(1)\varepsilon} = supp\tilde{B}^{(1)}$, $A_\varepsilon = supp\tilde{A}$, then process is like situation A. If this condition can not be met, then an appropriate smaller value for ε should be selected in order to calculate accurate $\eta(\tilde{A}, \tilde{B}^{(1)})$ in the premise that $|A_\varepsilon|$ and $|B_\varepsilon|$ are both finite values.
 - b) If $|supp\tilde{B}^{(n)}| \neq \infty$ and $|supp\tilde{A}| \neq \infty$, then set $\varepsilon = 0^+$, then process is like situation B, but if this condition can not be met, then choose the value ε under same principle of point a).
 - c) Still assume ε -cut set of $A_\varepsilon = [\alpha', \alpha'']$, even if $\tilde{B}^{(1)}$ is not lower semi-trapezoid distribution, but as long as the membership degree $u_{\tilde{B}^{(1)}}(x)|_{x=\alpha'}$ of α' to $\tilde{B}^{(1)}$ is 1, then $\tilde{B}^{(1)}$ can be dealt with as lower semi-trapezoid distribution. As well, even if $\tilde{B}^{(n)}$ is not higher semi-trapezoid distribution, but as long as the membership degree $u_{\tilde{B}^{(n)}}(x)|_{x=\alpha''} = 1$, $\tilde{B}^{(n)}$ can be dealt with as higher semi-trapezoid distribution like situation B. Process like this can make the calculation of $\eta(\tilde{A}, \tilde{B}^{(1)})$ and $(\eta\tilde{A}, \tilde{B}^{(n)})$ more consistent with the objective facts.

3.2 A finite set and an infinite set

Fig. 3 shows this situation. In the figure, \tilde{B} is a triangular fuzzy number, whose support set $supp\tilde{B}$ is continuous interval in real number field R while $supp\tilde{A}$ are some discrete points in closeness degree of \tilde{A} and \tilde{B} , two kinds of methods are adopted to calculate the weight parameter $\lambda_{\tilde{A}, \tilde{B}}$ of Eq. (5).

Method 1: Construct a new fuzzy set \tilde{B}' in the use of universe $X_{\tilde{A}}$ of \tilde{A} and membership function $u_{\tilde{B}}(x)$ of \tilde{B} , and set $\eta(\tilde{A}, \tilde{B}') = \eta(\tilde{A}, \tilde{B})$.

If $X_{\tilde{A}} = \{x_i | i = 1, 2, \dots, s\}$, then define the membership function of \tilde{B}' as below:

$$u_{\tilde{B}'} = \sum_{i=1}^s \frac{u_{\tilde{B}}(x)|_{x=x_i}}{x_i} \quad (15)$$

Then substitute $|A_\varepsilon|$ and $|B_\varepsilon|$ in Eq. (14) with $l_{\tilde{A}}$ and $l_{\tilde{B}}$ in Fig. 3(a), and calculate $\lambda_{\tilde{A}, \tilde{B}}$. Apply Eq. (4) to calculate $\eta_2(\tilde{A}, \tilde{B}')$ while calculating $\eta(\tilde{A}, \tilde{B}')$ by Eqs. (11)-(12).

Method 2: According to the distribution of $u_{\tilde{A}}(x)$, fit it to a fuzzy number of triangle or trapezoidal, construct the fuzzy set \tilde{A}' as Fig. 3(b). Then substitute $supp\tilde{A}$ with $supp\tilde{A}'$, and calculate $\lambda_{\tilde{A}, \tilde{B}}$ by Eq. (10). Even if the distribution of $u_{\tilde{B}}(x)$ is not linear, $u_{\tilde{A}}(x)$ can be fit to the triangular or trapezoidal.

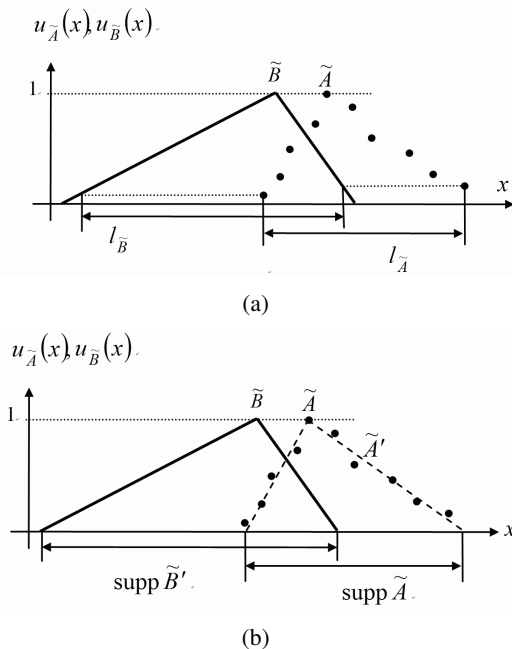


Figure 3 A support set of fuzzy sets \tilde{A} , \tilde{B} is continuous and the other is discrete

3.3 Both finite sets $supp\tilde{A}$ and $supp\tilde{B}$

At first, $u_{\tilde{A}}(x)$ and $u_{\tilde{B}}(x)$ are linearly fitted respectively, constructing two fuzzy sets \tilde{A}' and \tilde{B}' on continuous universe, then calculate $\eta(\tilde{A}', \tilde{B}')$ by using the method proposed in Section 3.1, and set $\eta(\tilde{A}', \tilde{B}') = \eta(\tilde{A}, \tilde{B})$.

4. The features of closeness degree algorithm

At present, the approaches of the fuzzy pattern recognition are divided into direct method and indirect method. The former one, model A is legible, each of whose dimension features are ordinary real numbers; but the latter one, model B is fuzzy, each of whose dimension features are fuzzy numbers. Both kinds of method are using membership degree and closeness degree to recognize objects based on maximum membership principle and the nearest neighbor principle.

In this paper, the closeness degree $\eta(\tilde{A}, \tilde{B})$ of fuzzy sets \tilde{A} and \tilde{B} is synthesized by $\eta_1(\tilde{A}, \tilde{B})$ and $\eta_2(\tilde{A}, \tilde{B})$ (or $\eta_4(\tilde{A}, \tilde{B})$) under some weights. The weight parameter is not a fixed value in the closeness degree algorithm of Eqs. (11)-(13), but can adaptively adjust under the characteristics of eigenvalue from model A and model B. This kind of self-adaption not only makes this algorithm manage every feature of be recognized mode which is expressed by ordinary real numbers and fuzzy numbers simultaneously, but also achieve soft transition between direct and indirect by merging membership and closeness degree, thus it no longer rigidly divides pattern recognition methods into two distinctive types. It has reduced the likelihood of error recognition effectively. Explain as follows:

Weather \tilde{A} and \tilde{B} are the same pattern need to be recognized. Here, suppose \tilde{B} is a m dimensions typical model whose eigenvalues are all fuzzy numbers denoted by m element vectors $\tilde{B} = (\tilde{B}_1, \tilde{B}_2, \dots, \tilde{B}_m)$, and its characteristic factors set is $Z = \{z_j | j = 1, 2, \dots, m\}$; \tilde{A} is m dimensions of recognized mode, whose eigenvalues have both fuzzy numbers and ordinary real numbers.

Suppose the j dimension eigenvalue of \tilde{A} is ordinary real number a_j , and notice that: ordinary real number is the special case of fuzzy number. For ordinary real number a_j , it can be also seen as a fuzzy numbers whose mean value is a_j , with left and right extensions are both zero. So its fuzzy model is shown below:

$$\tilde{A}_j = u_{a_j}(x) = \begin{cases} 0, & x = a_j \\ 1, & x \neq a_j \end{cases} \quad (16)$$

Despite that a_j can be expressed as \tilde{A}_j , it is actually a certain point in real universe R , the support set $supp\tilde{A}_j$ of \tilde{A}_j also is a certain point in the universe R as well, the interval width is:

$$|supp\tilde{A}_j| = 0 \quad (17)$$

Then calculate the weight parameter $\lambda_{\tilde{A}_j, \tilde{B}_j}$ by Eq. (10), closeness degree $\eta(\tilde{A}_j, \tilde{B}_j) = \eta_1(\tilde{A}_j, \tilde{B}_j)$ by Eqs. (11)-

(12), algorithm η_1 reflects the concept of membership degree, so:

- If is just the fuzzy expression of ordinary real number a_j , $\eta(\tilde{A}_j, \tilde{B}_j)$ and $\eta_1(\tilde{A}_j, \tilde{B}_j)$ represents the membership degree of a_j to \tilde{B}_j .
- If \tilde{A}_j is a fuzzy set, $\lambda_{\tilde{A}_j, \tilde{B}_j} \in (0, 1)$, then $\eta(\tilde{A}_j, \tilde{B}_j)$ is synthesized under certain weights by $\eta_1(\tilde{A}_j, \tilde{B}_j)$ and $\eta_2(\tilde{A}_j, \tilde{B}_j)$, so it reflects the concept of membership degree and closeness degree.
- In the progress of the left and right of ordinary real numbers a_j extending from 0 to \tilde{A}_j , before reaching the critical point $|supp\tilde{A}_j| = |supp\tilde{B}_j|$, the weight of $\eta_1(\tilde{A}_j, \tilde{B}_j)$ goes down while the weight of $\eta_2(\tilde{A}_j, \tilde{B}_j)$ rises up, and it largely shows the concept of closeness degree; after reaching the critical point, the situation is just opposite, $\eta(\tilde{A}_j, \tilde{B}_j)$ equals to $\eta_1(\tilde{A}_j, \tilde{B}_j)$ again while the left and right extending from to , represents membership degree again. The change of depends on the weight parameter $\lambda_{\tilde{A}_j, \tilde{B}_j}$. Due to the self-adaption of the membership function distribution of \tilde{A}_j and \tilde{B}_j , it is possible for making smooth transition of $\eta(\tilde{A}_j, \tilde{B}_j)$ between direct and indirect methods.

Fig. 4 shows the two common situations of fuzzy pattern recognition. \tilde{A} and \tilde{B} represent a one-dimensional model A and model B respectively (or a feature value of multi-dimensional model). Because their cores are close, use the Eq. (12) to calculate η . In the engineering practice (such as fault diagnosis), in Fig. 4(a), the model \tilde{A} to a large extent should be identified as \tilde{B} , and the judgment from algorithm η_1 is reasonable when is unreasonable. Now, the algorithm η mainly reflects the judgment of η_1 by the adaptive adjustment of k . To the practical problem expressed in Fig. 4(b), the possibility of \tilde{A} and \tilde{B} belonging to the same model is smaller, and the judgment from algorithm η_2 is reasonable when η_1 is unreasonable. Now, the algorithm η entirely reflects the judgment of η_2 by the adaptive adjustment of k (the weight of η_1 is 0).

Under the situation $|A_\epsilon| \gg |B_\epsilon|$ shown in Fig. 4(b), if the impact of η_1 is added to η , the results of η are usually larger. But it does not match the actual project situation, which will lead to error recognition. In this paper, the integration of η_1 mainly takes the situation (shown in Fig. 4(a)) into account.

Because η not only expresses the closeness degree of two fuzzy sets in the paper, but also expresses the

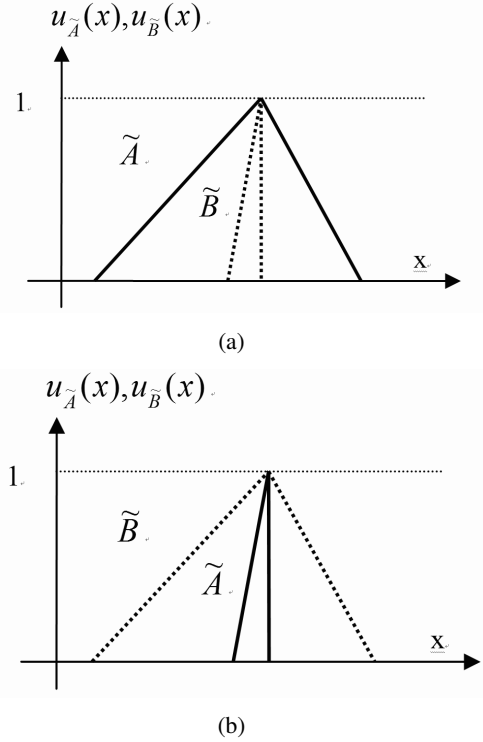


Figure 4 Two common situations of fuzzy pattern recognition

closeness degree of an ordinary real number to a fuzzy set. The η defined in Eq. (12) is a broad closeness degree. From the view of practical engineering, it is a proposed function used in pattern recognition to calculated the similarity degree of a characteristic element between model A and model B. η is not a closeness degree algorithm with strict mathematical sense, because:

- η is not simply expressing the closeness degree between two fuzzy sets in nature, but may be a membership degree of an ordinary real number belonging to a fuzzy set instead;
- $\eta(\tilde{A}, \tilde{B}) \neq \eta(\tilde{B}, \tilde{A})$.

5. Case studies

Based on Eq. (16), it would no longer distinct ordinary real numbers and fuzzy numbers symbolically. m -dimensional vector of be recognized pattern \tilde{A} are denoted as $\tilde{A} = (A_1, A_2, \dots, A_m)$ with the form of fuzzy vector; Characteristics factor set $Z = \{z_j | j = 1, 2, \dots, m\}$; There are typical patterns altogether $\tilde{B}^{(1)}, \tilde{B}^{(2)}, \dots, \tilde{B}^{(m)}$, among $\tilde{B}^{(i)} = (\tilde{B}^{(1)}, \tilde{B}^{(2)}, \dots, \tilde{B}^{(m)})$, $i = 1, 2, \dots, n$. The weight vector of Z is presented as $W = (w_1, w_2, \dots, w_m)$.

The closeness degree weighted between \tilde{A} and $\tilde{B}^{(i)}$ is:

$$\begin{aligned} & \eta(\tilde{A}, \tilde{B}^{(i)}) \\ &= W \circ (\eta(\tilde{A}_1, \tilde{B}_1^{(i)}), \eta(\tilde{A}_2, \tilde{B}_2^{(i)}), \dots, \eta(\tilde{A}_m, \tilde{B}_m^{(i)})) \\ &= \sum_{j=1}^m w_j \cdot \eta(\tilde{A}, \tilde{B}^{(i)}) \end{aligned} \quad (18)$$

Eventually is identified as the typical pattern of

$$\max\{\eta(\tilde{A}, \tilde{B}^{(i)}) | i = 1, 2, \dots, n\}$$

The effectiveness and practicability of the proposed algorithm are demonstrated by two cases.

The first one is the insulation fault diagnosis of power transformer.

Table 1 lists four kinds of eigenvalues of typical failure mode. Characteristic factors Z_1, Z_2 are the volume ratios of the transformer oil melt CH4 and C2H6 (unit 10-6). In Table 1, a three-dimensional array in cell represents a normal fuzzy number.

Table 1 The typical pattern characteristics of power transformer insulation failure

	Z_1	Z_2
Low temperature thermal fault	(161, 90, 90)	(39, 5.0, 5.0)
High temperature thermal breakdown	(320, 30, 30)	(103, 65, 65)
Low energy discharge fault	(73, 20, 20)	(42, 14, 14)
High energy discharge fault	(76, 30, 30)	(14, 4.0, 4.0)

Through measuring of the current fault state repeatedly, the normal fuzzy eigenvalues of model are: $Z_1 = (90, 9, 9)$; $Z_2 = (34, 0.01, 0.01)$.

The weights of characteristic factors Z_1 and Z_2 are both 0.5.

The similarity of the current failure mode and four typical failure modes calculated in the use of η_1, η_2 and η . Three different algorithms are shown in Table 2.

The current failure mode will be judged as low energy discharge by η_1 and η^* ; To this, the judgment of η_2 is high energy discharge. In fact, the current fault really belongs to low energy discharge, therefore, the judgment from η_1 and η^* is right while the judgment from η_2 is wrong.

The other case is product selection.

When designing a relay, the frequency of contact (denoted as Z) requirement is an important technical

Table 2 The similarity between the current failure mode and the typical failure mode

	η_1	η_2	η^*
Low temperature thermal fault	0.4829	0.0926	0.4454
High temperature thermal breakdown	0.1620	0	0.1620
Low energy discharge fault	0.7154	0.1276	0.5744
High energy discharge fault	0.4396	0.2194	0.3380

indicator in choosing the electromagnetic relay (denoted as mode $\tilde{B}^{(1)}$) or solid state relays (denoted as mode $\tilde{B}^{(2)}$). $\tilde{B}^{(1)}$ and $\tilde{B}^{(2)}$ as the eigenvalues of Z can be described as trapezoidal fuzzy numbers $(0, 8, 0, 4)$, $(20, 1000, 20, 0)$ respectively. Users are given designed products (denoted as mode \tilde{A}) of the characteristics of Z values $(8, 100, 2, 20)$. Select the type of product accordingly.

The distribution of $u_{\tilde{Q}}(x)$ for trapezoidal fuzzy set $\tilde{Q}(\alpha_1, \alpha_2, \beta_1, \beta_2)$ is shown in Fig 5. According to the

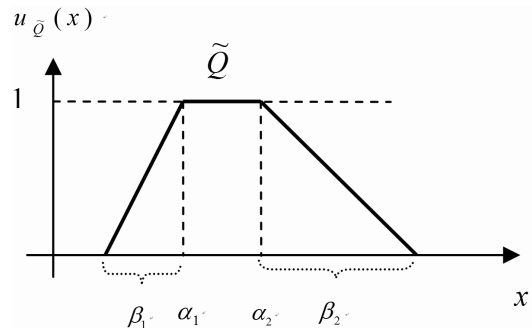


Figure 5 Trapezoidal fuzzy set \tilde{Q}

algorithms above:

$$\begin{aligned} \eta_1(\tilde{A}, \tilde{B}^{(1)}) &= 1; & \eta_1(\tilde{A}, \tilde{B}^{(2)}) &= 1; \\ \eta_1(\tilde{A}, \tilde{B}^{(1)}) &= 0.02727; & \eta_2(\tilde{A}, \tilde{B}^{(2)}) &= 0.09960; \\ \eta(\tilde{A}, \tilde{B}^{(1)}) &= 0.02727; & \eta(\tilde{A}, \tilde{B}^{(2)}) &= 0.87179; \end{aligned}$$

It can be seen that η_1 could not make a choice and fail to the current problem; the conclusion of η_2, η identifying the current product as solid state relay type is right. However, it is shown that $\eta_2(\tilde{A}, \tilde{B}^{(2)}) = 0.09960$, which is just too low to be convincing. While the calculated results of η is persuaded.

Because the kernels of $\tilde{A}, \tilde{B}^{(1)}$ and $\tilde{B}^{(2)}$ are close to each other, use the Eq (12) to calculate $\eta(\tilde{A}, \tilde{B}^{(1)})$, $\eta(\tilde{A}, \tilde{B}^{(2)})$; In the calculation, $\tilde{B}^{(2)}$ is treated as lower semi-trapezoid:

$$\lambda_{\tilde{A}, \tilde{B}} = \frac{(120 - 114)}{(120 + 114)} = 0.02564,$$

$$\eta_1(\tilde{A}, \tilde{B}^{(2)}) = 1; \eta_2(\tilde{A}, \tilde{B}^{(2)}) = 0.86842$$

6. Conclusion

Based on the two mostly common used closeness degree algorithms, a new general closeness degree algorithm is proposed in the paper, which constructs the similarity function between the typical model and the model to be recognized in the fuzzy pattern recognition. This algorithm is a comprehensive reflection of the concept of membership degree and closeness degree. It can not only handle the pattern characteristic values both derived from real numbers and fuzzy set expressions, realizing direct and indirect recognition integration, but also adaptively focus on membership and closeness according to the characteristics of pattern characteristic values, by reducing the occurrences of error recognition. Effectiveness of the algorithm is verified by case studies.

7. Acknowledgments

This work was supported by the National Natural Science Foundation (No. 60771069) and the Hebei Province Natural Science Foundation (No. F2007000 115 and No. 602069).

References

- [1] Kirlangic, M.E. Fractal modeling for EMG pattern recognition via artificial neural networks, In: *Proc. Speech and Signal Processing*, Vol.6, pp.3610-3613, 2000.
- [2] Madani, Kurosh. Artificial neural networks based image processing and pattern recognition: From concepts to real-world applications, In: *1st International Workshops on Image Processing Theory, Tools and Applications*, 2008.
- [3] Suganthan, P.N. Structural pattern recognition using genetic algorithms, In: *Pattern Recognition*, Vol. 35, No. 9, pp. 1883-1893, September 2002.
- [4] Han Li-Qiang, Wang Hua, Gao Ji-Gang, et al. Research of vision recognition on auto rack girders based on improved ART2 neural network and D-S evidence theory, In: *Proc. of the 2008 International Conference on Wavelet Analysis and Pattern Recognition*, Vol. 1, pp. 259-264, 2008.
- [5] Qu Dongcai, Meng Xiangwei, Huang Juan, et al. Research of artificial neural network intelligent recognition technology assisted by dempster-shafer evidence combination theory, In: *Proc. 7th International Conference on Signal Processing*, Vol. 1, pp. 46-49, 2004.
- [6] Seem John E. Pattern recognition algorithm for determining days of the week with similar energy consumption profiles, In: *Energy and Buildings*, Vol. 37, No. 2, pp. 127-139, 2005.
- [7] Li Shiyong. Engineering fuzzy mathematic and its applications. In: *Ha'erBin University of Technology Press*, 2004.
- [8] Li Zhi-gang, Zhao Quan-ming, Zhou Jun-gang, et al. A New Method for the Algorithm of Close Degree in Fuzzy Pattern Recognition. In: *Proc. of the 2nd International Conference on Intelligent Networks and Intelligent Systems*, pp. 338-341. 2009.
- [9] Wang Jiannan. Fault Diagnosis of Electric Apparatus Based on Fuzzy Clustering Analysis. In: *Tianjin, Hebei University of Technology*. 2006.
- [10] Li Ling-ling, Duanmu Yonguang, Li Zhi-gang. The Conceptual Design Method of Mechanical Products. In: *Proc. of IEEE Second International Workshop on Computer Science and Engineering*, pp.468-472. 2009.
- [11] Li Zhi-gang, Li ling-ling, Zhou Jun-gang, et al. A Determination Method of Product Design Parameters with Interval Constraint. In: *Key Engineering Materials*, Vols. 419-420 (2010), Advanced Design and Manufacture II, pp. 245-248.
- [12] Zhigang Li, Lingling Li, Fei Zheng, et al. A Comprehensive Evaluation Approach Based on Fuzzy Theory and Evidential Theory. In: *Proc. of the 6th International Conference on Fuzzy Systems and Knowledge Discovery*, Vol. 3, pp.7-11, 2009.