

Multi-source information fusion model in rule-based Gaussian-shaped fuzzy control inference system incorporating Gaussian density function

Zairan Li^{a,b}, Ting He^c, Luying Cao^c, Tunhua Wu^c, Pamela McCauley^d, Valentina E. Balas^e and Fuqian Shi^{c,*}

^aTianjin Key Laboratory of Process Measurement and Control, School of Electrical Engineering and Automation, Tianjin University, Tianjin, P.R. China

^bWenzhou Vocational & Technical College, Wenzhou, P.R. China

^cCollege of Information & Engineering, Wenzhou Medical University, Wenzhou, P.R. China

^dDepartment of Industrial Engineering and Management Systems, University of Central Florida, Orlando, FL, USA

^eDepartment of Automation and Applied Informatics, Aurel Vlaicu University of Arad, Romania

Abstract. An increasing number of applications require the integration of data from various disciplines, which leads to problems with the fusion of multi-source information. In this paper, a special information structure formalized in terms of three indices (the central presentation, population or scale, and density function) is proposed. Single and mixed Gaussian models are used for single source information and their fusion results, and a parameter estimation method is also introduced. Furthermore, fuzzy similarity computing is developed for solving the fuzzy implications under a Mamdani model and a Gaussian-shaped density function. Finally, an improved rule-based Gaussian-shaped fuzzy control inference system is proposed in combination with a nonlinear conjugate gradient and a Takagi-Sugeno (T-S) model, which demonstrated the effectiveness of the proposed method as compared to other fuzzy inference systems.

Keywords: Gaussian density function, IF-THEN rule, multi-source information fusion, similarity computing, fuzzy control inference system

1. Introduction

The human brain obtains information from different sources; it then merges this information to form concepts and finally outputs natural language (NL), which is powerful and versatile enough to describe the real world. NL can be regarded as the fusion of disparate

information; it is vague, ambiguous, and uncertain. The quantitative calculation and qualitative analysis of NL is the ultimate goal of artificial intelligence. There are two strands of research linking the initial information acquisition with NL: (1) how to simplify the presentation of NL and (2) how to form NL from multi-source information. Usually, humans express emotions of certain objects by using sentences and affective words, but they cannot fully express their intuitive perception of an object simply through separating these terms. Natural Language Processing (NLP) was developed to solve

*Corresponding author. Fuqian Shi, College of Information & Engineering, Wenzhou Medical University, Wenzhou, 325035 P.R. China. Tel.: +86 577 86689913; Fax: +86 577 86689922; E-mail: sfq@wmu.edu.cn.

this problem; however, many difficulties remain in this field. Computing with Words (CW) was also introduced to decrease the complexity related to linguistic variables [16–18]. This has allowed for a more exact expression of the meaning of what a human is thinking about and has provided a feasible direction for NLP under weakened conditions. Zadeh introduced the framework of this phenomenon of uncertainty using Fuzzy Sets (FS) in 2005 [19]. The FS theory was also addressed to describe objects at a coarse-grained level. Herrera and Martínez [5] introduced a 2-tuple fuzzy linguistic representation model for CW without any loss of information. Furthermore, Lawry [13, 14] proposed Label Semantics (LS) for vague concept modeling and reasoning techniques so as to formalize uncertainty in presentation theory. Subsequently, Lawry and Tang [12, 34, 35] proposed a new semantic understanding model: the Prototype Theory (PT). These works discovered the connection between fuzzy presentation technology and high-level semantics. In engineering fields, linguistic representation models combined with affective words have had some applications, such as fuzzy decision making [21, 31] and KANSEI Engineering (KE). Fuzzy inference methodologies have also been shown to be effective in our previous work on Rough Sets [7] and Fuzzy Support Vector Machines (SVMs) [6].

However, it has been regarded as more feasible to focus on multi-source information fusion rather than on NL itself. Moreover, it is important to discover the mechanics of integrating multi-source information in the human brain. Due to the modular and vague appearance of multi-source information, uncertainty reasoning methods and their associated mathematical tools are thought to offer more interpretability and a much stronger generalization capability [24]. Yager developed the theoretical foundation for multi-source information fusion techniques based on set measure and possibility theories [25, 26]. Normally, single-source information consists of steady features that are more easily formalized and parameterized. In previous studies, the sum, product, max/min, and Weighted Arithmetic Mean (WAM) were used to combine single-source information, and each output represented an independent source of information that could be treated separately [15].

Relative to mathematical research and understanding the phenomenon of uncertainty, the integration of information using fuzzy inference techniques pervades many scientific disciplines, such as multivariate and type-2 fuzzy sets; bipolar models [10, 11]; and probability and possibility issues [9, 27]. Information fusion is

the merging of information from disparate sources with differing conceptual, contextual, and typographical representations. It has been successfully applied in data mining and the consolidation of data from unstructured or semi-structured resources, and it has also led to many achievements in various fields [1, 4, 8]. Fusion methods include product fusion (such as the Bayes posterior probability model), linear fusion (SVM classifiers), and nonlinear fusion (super-kernel integration) [23]. Recent developments and applications of fuzzy information fusion can be found in pattern classification, image analysis, decision-making, man-made structures, and medicine [30, 32]. Furthermore, over the past several years, there has been a number of successful applications of fuzzy integrals in decision-making and pattern recognition that have employed multiple information sources [3, 20].

In this paper, we formalize multi-source information as a multivariable group and describe each information structure as a special kind of triple, $I = \langle P, d, \rho \rangle$, where P denotes a typical point of positive examples relative to the information structure I , d is a distance measurement that represents the population of information, and ρ is a Probability Density Function (PDF). The basic idea of this formalized information structure is to assume that the neighborhood radius of each information structure is uncertain, which is limited by PDF- ρ . Thus, we will calculate the value of P relative to an information structure under a given level. An information fusion technique was developed by formalizing this special information structure; furthermore, information fusion employing fuzzy sets was applied in this paper. A Single Gaussian Model (SGM) was applied to single-source information, and a Gaussian Mixed Model (GMM) was applied to the fusion of this information by incorporating probabilistic and statistical methods [28, 36].

The remainder of this paper proceeds as follows. In Section 2, we propose an information structure that incorporates a definition of the information kernel, boundary, and Gaussian PDF. An improved algorithm for parameter estimation is also introduced. Section 3 introduces fuzzy similarity relations and IF-THEN rules for this special information structure. These are helpful for calculating the possibilities in a rule-based fuzzy inference system (FIS). Section 4 develops a rule-based information fusion model using a conjugate gradient and Takagi–Sugeno (T-S) model under a rule-based Gaussian-shaped fuzzy inference system (RGS-FIS). A time-series analysis using natural disaster datasets is also introduced using RGS-FIS, and we

demonstrate the effectiveness of our method in comparison to other methodologies. Finally, in Section 5, we give our conclusions and ideas for future work.

2. Information fusion models by using probability density function

2.1. Definitions

Definitions for our information structure and kernel computing method were established as follows.

Definition 1. Assume object Ω is described by the multi-source information set $I = \{I_k | k = 1, 2, \dots, m\}$ and that measure set $V = \{v_k | k = 1, 2, \dots, m\}$ is a set of information structures corresponding to set I . For $\forall v_k \in V$, we define $v_k = \langle P_k, d_k, \rho_k \rangle$, where P_k is a typical point as the kernel of I_k . Moreover, d_k is a metric of the information structure v_k related to the population or scale of information and will be used for boundary computing. Lastly, ρ_k is a density function on the threshold of v_k .

Definition 2. Let the fusion operator be \oplus so that Ω can be formalized as:

$$I_1 + I_2 + \dots + I_m = v_1 \oplus v_2 \oplus \dots \oplus v_m \quad (1)$$

where \oplus is a minimum operator.

Definition 3. $\forall P_k, Q_k$ in an n -dimensional Euclidean space R^n , $P_k = [P_{k1}, P_{k2}, \dots, P_{kn}]$, and $Q_k = [Q_{k1}, Q_{k2}, \dots, Q_{kn}]$. Moreover, let $d = \|\cdot\|$, and it has the following properties:

- (1) $d(P_k, P_k) = \|P_k\| = \sqrt{(\sum_i P_{ki})}$
- (2) $d(P_k \pm Q_k) = \|P_k \pm Q_k\|, \forall P_k, Q_k \in R^n$
- (3) $\forall \alpha, \beta \in R, P_k, Q_k \in R^n$

We have $d(\alpha P_k \pm \beta Q_k) = \|\alpha P_k \pm \beta Q_k\|$; in addition,

$$d(\alpha P_k + \beta Q_k) \leq |\alpha| d(P_k) + |\beta| d(Q_k).$$

Definition 4. For sample points $\{P_k^l | l = 1, 2, \dots\}$, the statistics-based kernel point computation is calculated as:

$$P_k = \sum_l P_k^l = \left[\sum_l P_{k1}^l, \sum_l P_{k2}^l, \dots, \sum_l P_{kn}^l \right] \quad (2)$$

P_{ki}^l indicates the value of the i -th dimension of the l -th sample point in the k -th information source.

The boundary of v_k gives the scale of the neighborhood of all elements in this special information structure. This is defined below.

Definition 5. For $\forall P_k \in R^n$, there exists a neighborhood,

$$N_{P_k}^\varepsilon = \{X | \|P_k - X\| < \varepsilon, X \in R^n\} \quad (3)$$

Definition 6. For calculating the boundary of I , two sets were defined as:

- The Upper Approximation Boundary (UAB)

$$UP_B = \{P_l | P_l \in N_{P_K}^u\} \quad (4)$$

- The Lower Approximation Boundary (LAB)

$$LP_B = \{P_l | P_l \in N_{P_K}^l\} \quad (5)$$

Therefore, the boundary is $P_B = UP_B \setminus LP_B = B(u, t)$. Thus, we have $P_B = P_K + \lambda(P_B - P_K)$, $\lambda \in [0, 1]$, which exhibits fuzziness attributes at the boundary.

2.2. Probability density function

-Single Gaussian Model for single-source information

The Gaussian distribution is a continuous probability distribution with a bell-shaped PDF in one-dimensional space:

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \quad (6)$$

The parameter μ is the mean or expectation, and σ^2 is the variance. The SGM is applied to induct the density function of the proposed information structure I , and we define:

$$\delta(X, \mu, \Phi) = \frac{1}{\sqrt{(2\pi)^n |\Phi|}} e^{-\frac{1}{2}(X-\mu)^T \Phi^{-1}(X-\mu)} \quad (7)$$

where X is a vector in n -dimensional space, Φ is the covariance matrix, and μ is the mean value of the density function. The density function's properties are determined by (Φ, μ) , so this is a parameter estimation problem [29]. For any point $P_i \in R^n$, its probability density function is $\delta(P_i, \mu, \Phi)$, and if, for any information structure v_k , each P_i in v_k is regarded as an independent event, then the PDF of v_k is:

$$\delta_k = \delta(v_k, \mu, \Phi) = \prod_i^m \delta(P_i, \mu, \Phi) \quad (8)$$

The maximum likelihood estimation can be used to estimate the parameters (Φ, μ) under (8). Taking the logarithm of (8), we have:

$$\begin{aligned}
O(\mu, \Phi) &= \ln\left(\prod_i^m \delta(P_i, \mu, \Phi)\right) \\
&= \sum_i^m \ln(\delta(P_i, \mu, \Phi)) \\
&= \sum_i^m \left[-\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\Phi| \right. \\
&\quad \left. + \frac{1}{2} (P_i - \mu)^T \Phi^{-1} (P_i - \mu)\right] \\
&= -\frac{nm}{2} \ln(2\pi) - \frac{m}{2} \ln |\Phi| \\
&\quad - \frac{m}{2} \sum_i [P_i - \mu]^T \Phi^{-1} (P_i - \mu)
\end{aligned} \quad (9)$$

Taking the partial derivative w.r.t. μ of $O(\mu, \Phi)$ and setting it to 0, we obtain the following:

$$\begin{aligned}
\partial_\mu(O(\mu, \Phi)) &= -\frac{1}{2} \sum_i^m [-2\Phi^{-1}(P_i - \mu)] \\
&= \Delta^{-1} \sum_i^m [(P_i - \mu)] \\
&= \Delta^{-1} \left[\sum_i^m P_i - m\mu\right] \\
&= 0
\end{aligned} \quad (10)$$

This gives $\hat{\mu} = \frac{1}{2} \sum_i P_i$. Similarly, for Φ , we can obtain $\hat{\Phi} = \frac{1}{n-1} \sum_i (P_i - \hat{\mu})(P_i - \hat{\mu})^T$. Thus, if the density of each point in v_k is $\delta(P, \hat{\mu}, \hat{\Phi})$, then our estimation of the parameter μ is:

$$\hat{\mu} = \left(\frac{1}{n} \sum_i e_{1i}, \frac{1}{n} \sum_i e_{2i}, \dots, \frac{1}{n} \sum_i e_{ni} \right) \quad (11)$$

where e_{li} is the coordinate of P_i in R^n .

The covariance $\hat{\Phi}$ is converted to

$$\begin{aligned}
\hat{\Phi} &= \frac{1}{n-1} \sum_i [e_{1i} - \hat{\mu}_1, e_{2i} - \hat{\mu}_2, \dots, e_{ni} - \hat{\mu}_n] \\
&\quad \begin{bmatrix} e_{1i} - \hat{\mu}_1 \\ e_{2i} - \hat{\mu}_2 \\ \dots \\ e_{ni} - \hat{\mu}_n \end{bmatrix} \\
&= \frac{1}{n-1} \sum_{j=1}^n \sum_{i=1}^n (e_{ji} - \hat{\mu}_j)^2
\end{aligned} \quad (12)$$

- Gaussian Mixed Model and parameter estimation

For multi-source information fusion, we need to calculate all of I_k 's density functions as well as calculate the new density function. For m multi-source information structures, let $I_{fusion} = \sum_{i=1}^l \alpha_i \delta(P, \mu_i, \Phi_i)$ for a normalized weight parameter α : i.e., $\sum_i \alpha_i = 1$. To calculate and simplify the covariance matrix Φ , let

$$\Phi = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ 0 & \dots & \dots & 0 \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 \vec{I} \quad (13)$$

From the SGM, we have that

$$\delta(P, \mu, \sigma^2 \vec{I}) = \frac{1}{\sqrt{(2\pi)^n}} \sigma^{-1} e^{-\frac{(P-\mu)^T (P-\mu)}{2\sigma^2}} \quad (14)$$

Calculate:

$$\begin{aligned}
\partial_\mu[\delta(P, \mu, \sigma^2 \vec{I})] &= \frac{1}{\sqrt{(2\pi)^n}} \partial_\mu(\sigma^{-1} e^{-\frac{(P-\mu)^T (P-\mu)}{2\sigma^2}}) \\
&= \frac{1}{\sqrt{(2\pi)^n}} \sigma^{-1} e^{-\frac{(P-\mu)^T (P-\mu)}{2\sigma^2}} \partial_\mu\left(-\frac{(P-\mu)^T (P-\mu)}{2\sigma^2}\right) \\
&= \delta(P, \mu, \sigma^2 \vec{I}) \left(\frac{P-\mu}{\sigma^2}\right)
\end{aligned}$$

and

$$\begin{aligned}
\partial_\Delta[\delta(P, \mu, \sigma^2 \vec{I})] &= \frac{1}{\sqrt{(2\pi)^n}} \left((-1)\sigma^{-2} e^{-\frac{(P-\mu)^T (P-\mu)}{2\sigma^2}} \right) \\
&\quad + \frac{1}{\sqrt{(2\pi)^n}} \sigma^{-1} e^{-\frac{(P-\mu)^T (P-\mu)}{2\sigma^2}} \left[\frac{(P-\mu)^T (P-\mu)}{\sigma^3} \right] \\
&= \delta(P, \mu, \sigma^2 \vec{I}) \left(\frac{(P-\mu)^T (P-\mu)}{\sigma^3} - \frac{1}{\sigma^2} \right)
\end{aligned}$$

Then, for $\Delta = c\vec{I}$, $c \in R$, GMM is defined as $G(P) = \sum_i \alpha_i \delta(P, \mu_i, \sigma_i)$, $i = 1, 2, \dots, lf$. The

number of parameters for estimation is $3l$. If we let $\theta = [\alpha_1, \alpha_2, \dots, \alpha_l, \mu_1, \mu_2, \dots, \mu_l, \sigma_1^2, \sigma_2^2, \dots, \sigma_l^2]$, the object is that:

$$\begin{aligned} L(\theta) &= \ln \left[\prod_i G(P_i) \right] = \sum_i \ln(G P_i) \\ &= \sum_i \ln \left(\sum_{j=1}^l \alpha_j \delta(P_i, \mu_j, \sigma_j^2) \right) \end{aligned} \quad (15)$$

which can be differentiated w.r.t. μ_j and σ_j . Thus, we have that:

$$\partial_{\mu_j}(L(\theta)) = \sum_i \frac{\alpha_j \delta(P_i, \mu_j, \sigma_j^2)}{\sum_{j=1}^l \alpha_j \delta(P_i, \mu_j, \sigma_j^2)} \frac{P_i - \mu_j}{\sigma_j^2} \quad (16)$$

Let $\varphi_j(P_i) = \frac{\alpha_j \delta(P_i, \mu_j, \sigma_j^2)}{\sum_{j=1}^l \alpha_j \delta(P_i, \mu_j, \sigma_j^2)}$, so that:

$$\partial_{\mu_j}(L(\theta)) = \sum_i \varphi_j(P_i) \left(\frac{P_i - \mu_j}{\sigma_j^2} \right) \quad (17)$$

Similarly, we can find:

$$\begin{aligned} \partial_{\sigma_j}(L(\theta)) &= \sum_i \frac{\alpha_j \delta(P_i; \mu_j, \sigma_j^2)}{\sum_{j=1}^l \alpha_j \delta(P_i; \mu_j, \sigma_j^2)} \\ &\quad \left[\frac{(P_i - \mu_j)^T (P_i - \mu_j)^T}{\sigma_j^3} - \frac{1}{\sigma_j^2} \right] \\ &= \sum_i \varphi_j(P_i) \left[\frac{(P_i - \mu_j)^T (P_i - \mu_j)^T}{\sigma_j^3} - \frac{1}{\sigma_j^2} \right] \end{aligned} \quad (18)$$

Setting the above two equations equal to 0, we have

$$\hat{\mu}_j = \frac{\sum_i \varphi_j(P_i) P_i}{\sum_i \varphi_j(P_i)} \quad (19)$$

$$\hat{\sigma}^2 = \frac{1}{3} \frac{\sum_i \varphi_j(P_i) (P_i - \mu_j)^T (P_i - \mu_j)}{\sum_i \varphi_j(P_i)} \quad (20)$$

For α_j , under the constraint $\sum_j \alpha_j = 1$, we use Lagrange multipliers to re-define the object as:

$$\begin{aligned} J &= L(\theta) + \lambda \left(1 - \sum_{i=1}^l \alpha_i \right) \\ &= \sum_i \ln \left(\sum_j \alpha_j \delta(P_i, \mu_j, \sigma_j^2) \right) + \lambda \left(1 - \sum_{i=1}^l \alpha_i \right) \end{aligned} \quad (21)$$

Differentiating this new object w.r.t. α_j , we have that:

$$\begin{aligned} \partial_{\alpha_j} J &= \sum_i \frac{\delta(P_i, \mu_j, \sigma_j^2)}{\sum_{j=1}^l \alpha_j \delta(P_i, \mu_j, \sigma_j^2)} - \lambda \\ &= \frac{1}{\alpha_j} \sum_i \varphi_j(P_i) - \lambda = 0 \end{aligned} \quad (22)$$

$$[\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_l] = \left[\frac{1}{\lambda} \sum_i \varphi_1(P_i), \frac{1}{\lambda} \sum_i \varphi_2(P_i), \dots, \frac{1}{\lambda} \sum_i \varphi_l(P_i) \right] \quad (23)$$

$$\hat{\alpha}_1 + \hat{\alpha}_2 + \dots + \hat{\alpha}_l = \frac{1}{\lambda} \left(\sum_i (\varphi_1(P_i) + \varphi_2(P_i) + \dots + \varphi_l(P_i)) \right) = 1 \quad (24)$$

Furthermore, we know $\lambda = l$, so:

$$[\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_l] = \left[\frac{1}{lf} \sum_i \varphi_1(P_i), \frac{1}{l} \sum_i \varphi_2(P_i), \dots, \frac{1}{l} \sum_i \varphi_l(P_i) \right] \quad (25)$$

where φ is also a function of parameters, and we can resolve this using the following iteration:

Step 1: Let

$$\theta = [\alpha_1, \alpha_2, \dots, \alpha_l, \mu_1, \mu_2, \dots, \mu_l, \sigma_1^2, \sigma_2^2, \dots, \sigma_l^2]$$

Given an initial value and in order to achieve convergence, $\mu_1, \mu_2, \dots, \mu_m$ may be calculated by the cluster method.

Step 2: Calculate $\varphi_j(P_i)$.

Step 3: Calculate $\tilde{\mu}_j = \frac{\sum_i \varphi_j(P_i) P_i}{\sum_i \varphi_j(P_i)}$.

Step 4: Calculate

$$\sigma_j = \frac{1}{lf} \frac{\sum_i \varphi_j(P_i) (P_i - \tilde{\mu}_j)^T (P_i - \tilde{\mu}_j)}{\sum_i \varphi_j(P_i)}.$$

Step 5: Calculate $\alpha_j = \frac{1}{lf} \sum_i \varphi_j(P_i)$.

Step 6: Let

$\hat{\theta} = [\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_l, \hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_l, \hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_l^2]$ If $\|\theta - \hat{\theta}\| < \delta$ for a given threshold δ , then stop the process; otherwise, proceed to Step 2.

In actuality, the density function of information fusion under this special structure is a product of the fusion of SGMs. For all information structures v_k and their SGM densities $\delta(v_k), \delta(I_{fusion}) = \prod_k \delta(v_k)$; therefore, we have that:

$$\begin{aligned} \prod_k \delta(v_k) &= \prod_k \frac{1}{\sqrt{(2\pi)^n} \sigma_k} e^{-\frac{1}{2\sigma_k^2} (P-\mu)^T (P-\mu)} \\ &= \frac{1}{\sqrt{(2\pi)^{nk}} \prod_k \sigma_k} e^{-\sum_k \frac{1}{2\sigma_k^2} (P-\mu)^T (P-\mu)} \\ &= \frac{1}{\sqrt{(2\pi)^{nk}} \prod_k \sigma_k} e^{\alpha P^2 + \beta P + \gamma} \\ &= \frac{1}{\sqrt{(2\pi)^{nk}} \cdot C \cdot \prod_k \sigma_k} e^{-\frac{1}{2\sigma_k^2} (P'-\mu')^T (P'-\mu')} \\ &= C' \delta \end{aligned} \quad (26)$$

where C is an undetermined constant.

In particular, in a one-dimensional space with $\sigma = 1$, we have that:

$$\begin{aligned} \delta(I_{fusion}) &= \frac{1}{\sqrt{(2\pi)^{nk}} \prod_k \sigma_k} e^{\alpha x^2 + \beta x + \gamma} \\ &= \frac{1}{\sqrt{(2\pi)^{nk}}} e^{\alpha x^2 + \beta x + \gamma} \\ &= \frac{1}{\sqrt{(2\pi)^{nk}}} e^{x^2 + \beta' x' + \gamma'} \\ &= C \frac{1}{\sqrt{(2\pi)^n}} e^{(x'-\mu')^2} \end{aligned} \quad (27)$$

This is a linear transformation of the basic Gaussian function. Thus, for any two information structures v_i, v_j , the fusion result is $v_{ij} = \langle \alpha P_{ij}, \beta d_{ij}, \gamma \delta_{ij} \rangle$ where α, β , and γ are undetermined coefficients.

3. Fuzzy implications of information structure under IF-THEN rules

3.1. Fuzzy implications of information structures under IF-THEN rules

In fuzzy sets, the rule “IF x is \bar{A} , THEN y is \bar{B} ” indicates a fuzzy implication between \bar{A} and \bar{B} as denoted by $\bar{A} \rightarrow \bar{B}$. If we let $x, y \in [0, 1]$ be the memberships of \bar{A} and \bar{B} , respectively, we list the Mamdani model for the membership computing as $\forall x, y \in [0, 1], F(x, y) = \text{Min}\{x, y\}$.

We construct a fuzzy membership based on a new fuzzy implication and inference system. We also derive a similarity relationship and apply this to the Gaussian density function-based fuzzy rule inference system. For δ_k in a rule-based IF-THEN inference system, suppose that the rule set is:

IF I_1 is v_1 , then I_o is v_o, ω_1, I_2 is v_2 , then I_o is v_o, ω_2 .

We can integrate these rules as:

IF I_1 is v_1 and I_2 is v_2 , THEN I_o is v_o ,

$$\omega_{ij} = \delta(I_{fusion}) = \delta(v_{ij}) \quad (28)$$

The Mamdani model for δ_k in two-dimensional space (x, y) will be:

$$\text{Min} \left(\frac{1}{\sqrt{2\pi\sigma_{11}\sigma_{12}}} e^{-\frac{1}{2} \left[\frac{(x-\mu_{11})^2 + (y-\mu_{12})^2}{\sigma_{11}^2 + \sigma_{12}^2} \right]}, \frac{1}{\sqrt{2\pi\sigma_{21}\sigma_{22}}} e^{-\frac{1}{2} \left[\frac{(x-\mu_{21})^2 + (y-\mu_{22})^2}{\sigma_{21}^2 + \sigma_{22}^2} \right]} \right) \quad (29)$$

In particular, if $\sigma_{11} = \sigma_{12} = \sigma_{21} = \sigma_{22} = 1$ and $\mu_{11} = \mu_{12} = \mu_{21} = \mu_{22} = 0$, we have that:

$$M(x, y) = \frac{1}{\sqrt{2\pi}} e^{-\left[\frac{x^2 + y^2}{4} \right]} \quad (30)$$

Thus, for any other implication operators, the function of rules will have the form:

$$M(x, y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(Ax^2 + By^2 + Cx + Dy + E)}. \quad (31)$$

4. Applications

4.1. Mamdani model-based fuzzy control inference system using nonlinear conjugate gradient

In the previous section, information was formalized as $v_k = \langle P_k, d_k, \rho_k \rangle$ where P_k is a central point in R^n , $d_k \in R$, and ρ_k is a Gaussian density function. P_k and d_k operate as fuzzy numbers using the fuzzy

logical operation in Section 2. In our fuzzy rule-based inference system if different information (multi-source information) implies the same conclusion, then this information is integrated. Supposing that the multi-source information structure v_k will conclude with a particular assertion at the ρ_k level, we have that:

$$\text{IF } I_1 \text{ is } v_1 \text{ and } I_2 \text{ is } v_2 \text{ and } \dots, \text{ and } I_m \text{ is } v_m \text{ THEN } I_o \text{ is } v_o, \varpi \quad (32)$$

This can be simplified to,

$$\text{IF } v_1 \text{ and } v_2 \text{ and } \dots, \text{ and } v_m \text{ THEN } \varpi(v_1, v_2, \dots, v_m)$$

and

$$\text{IF } P_1 \text{ and } P_2 \text{ and } \dots, \text{ and } P_m \text{ THEN } P_1 \wedge P_2 \wedge \dots \wedge P_m$$

$$\text{IF } d_1 \text{ and } d_2 \text{ and } \dots, \text{ and } d_m \text{ THEN } d_1 \wedge d_2 \wedge \dots \wedge d_m$$

$$\text{IF } \rho_1 \text{ and } \rho_2 \text{ and } \dots \text{ and } \rho_m \text{ THEN } \varpi(\rho_1, \rho_2, \dots, \rho_m).$$

Now, we only discuss the density function δ_k (here ρ_k) in our FIS. Letting $\Theta = [\delta_1, \delta_2, \dots, \delta_n]$, we have that:

$$\text{IF } \Theta \text{ THEN } \varpi(\Theta). \quad (33)$$

From the previous section, we know that $\varpi(\Delta)$ is a Gaussian density function, so the rule is re-labeled as IF X THEN $f(X)$. For this rule set, we have

$$R_i : \text{IF } X \text{ THEN } f_i(X)$$

However, as $f(X)$ is a nonlinear function, it is difficult to find its minimum point under the Mamdani model, so we need to linearize $f(X)$ and use the nonlinear conjugate gradient algorithm to optimize the parameters of $f(X)$.

If we suppose that $f(X) = [Ax - b]^T [Ax - b]$, then the gradient is $\nabla_x f(x) = 2A^T(Ax - b)$, and the objective is to find x subject to $\nabla_x f(x) = 0$. The nonlinear conjugate gradient requires f being twice differentiable, but as f is a Gaussian function, it is infinitely differentiable. Starting from the opposite direction as $\Delta x_0 = -\nabla_x f(x_0)$ with step size α , we have that:

$$\alpha_0 = \arg \min_{\alpha} f(x_0 + \alpha \Delta x_0) \quad (34)$$

$$x_1 = x_0 + \alpha_0 \Delta x_0 \quad (35)$$

This is the first iteration in the direction of Δx_0 , and by setting the initial conjugate direction $s_0 = \Delta x_0$, the following steps will calculate Δx_n :

Step 1: Calculate $\Delta x_n = -\nabla_x f(x_n)$.

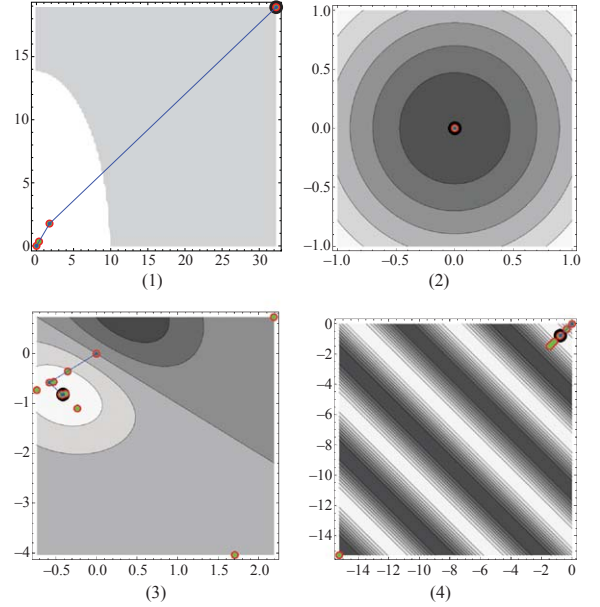


Fig. 1. Mamdani model-based fuzzy inference system using nonlinear conjugate gradient for (1) and (2) as compared with a non-Mamdani model (3)-(4).

Step 2: Calculate $\beta_n : \beta_n = \frac{\Delta x_n^T (\Delta x_n - \Delta x_{n-1})}{\Delta x_{n-1}^T \Delta x_{n-1}}$. (Polak–Ribière)

Step 3: Update the conjugate direction $s_n = \Delta x_n + \beta_n s_{n-1}$.

Step 4: Calculate $\alpha_n = \arg \min_{\alpha} f(x_n + \alpha s_n)$.

Step 5: Update $x_{n+1} = x_n + \alpha_n s_n$.

The algorithm is based on the quadratic function that we use to normalize the Gaussian function $f(x)$ in order to speed up the iterations. Considering a simplified Mamdani model and from formula (20), we know that:

$$M(x, y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(Ax^2 + By^2 + Cx + Dy + E)} \quad (36)$$

Using the nonlinear conjugate gradient, we obtain the results given in Fig. 1 and Table 1 by comparing with other special functions. From Table 1, we know that the Gaussian density function will be approximated in just a few steps by the nonlinear conjugate gradient algorithm, which is the reason we selected the Gaussian distribution as the density function of this special structure. We also compare other forms of density function, which appear to require more steps under the nonlinear conjugate gradient algorithm.

Table 1
Start point, steps, and gradient by nonlinear conjugate gradient

Functions	Initial	(x, y)	Steps	Gradient
(A, B, C, D, E)=(2,1,1,0,-3)	0	(32.2, 18.9)	2	4
(A, B, C, D, E)=(1,1,0,0,3)	1.7	(0, 0)	0	1
$\frac{x+y}{\sqrt{2\pi}} e^{-\frac{(2x^2+y^2-3)}{2}}$	-1.32	(-0.4, -0.8)	7	18
$\text{Sin}(x + y)$	-1	(-0.78, -0.78)	5	13

4.2. Takagi–Sugeno model in RGS-FIS

Takagi and Sugeno [18] proposed a fuzzy IF-THEN rules system as the local input–output relations of a nonlinear system to scale the population of rules under a multi-dimensional fuzzy inference system, known as the T-S model [21]. The normal rules for the T-S model under the special information structure proposed for our information fusion method are:

$$R_{T-S} : \text{IF INPUT - is } I_1, \text{ INPUT - is } I_2, \dots, \text{ INPUT - } n \text{ is } I_n \text{ THEN } I_f = f(I_1, I_2, \dots, I_m).$$

The T-S model outputs a linear, non-constant function that will reduce the population of rules.

From rule set R_{T-S} , we can simplify $I_f = \sum_{i=1}^n a_i d_i + b_i d_i$, in which a_i and b_i are undetermined constants. Let the standard deviation in Equation (6)

$\sigma = \vec{I}, \mu = 0$, and thus, $I_f = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^n a_i e_i^{-\frac{x^2}{2}} + b_i x$. The first part of I_f is a GMM model that can be estimated by Section 2.2-(2), and the second part of I_f is a linear function (see Fig. 2).

Furthermore, for the nonlinear conjugate gradient proposed in Section 4.1, we obtain 100 steps and 301 gradients to find the minimum point (the Mamdani model). As a result, we can simplify this in RGS-FIS under the T-S model to output three linear membership functions. Suppose that the inputs are Gaussian-shaped rules, and the outputs are linear functions. Let the membership function of INPUT 1 and INPUT 2 be a Gaussian function, and the OUTPUT is composed of three linear functions [33]. We have this RGS-FIS system under the T-S model (see Fig. 3).

5. Concluding remarks and future works

This paper proposed a novel information structure applicable to a Gaussian-shaped FIS. We developed the RGS-FIS approach using the nonlinear conjugate gradient algorithm and a T-S model. However, there are two problems with RGS-FIS: one is that new fusion operator

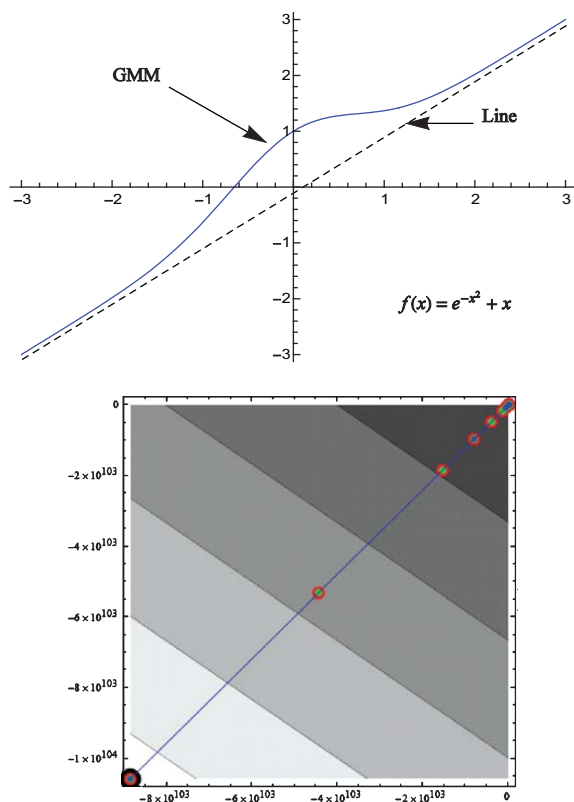


Fig. 2. T-S model's two parts under IF-THEN fusion operators and steps to minimum point using the nonlinear conjugate gradient.

parameters depend on a complex estimation process, and the other is that all data variables are supposed to be independent ($r = 0$). The model selection for similarity computing under rule-based fuzzy implication operations should also be improved.

Future work will focus on the pre-processing of datasets as well as the estimation of model parameters. Pre-processing will tune the parameters of the model to display a simpler mathematical presentation and assure a robust inference process. Furthermore, the fusion operator needs to be improved so that it does not solely depend on fuzzy implications. Although similarity computing is the key factor for calculating the possibility of IF-THEN rules, it is not clear whether

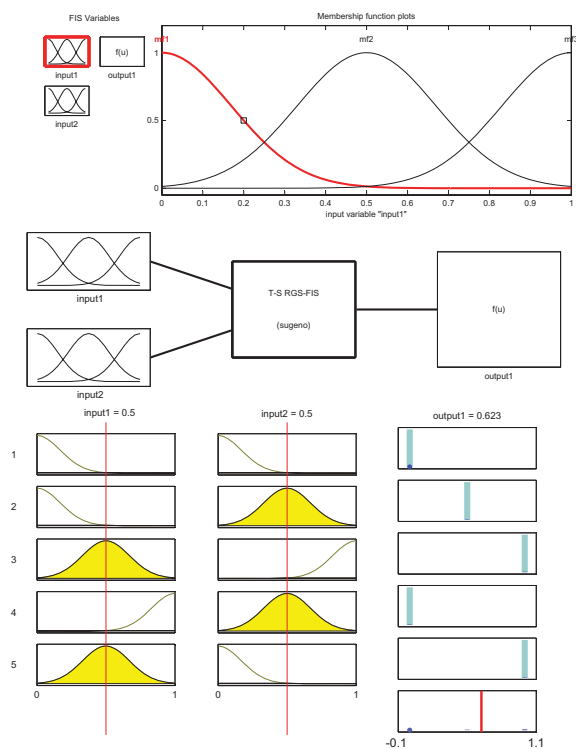


Fig. 3. RGS-FIS with two inputs and three linear output functions.

a feasible algorithm can be developed for this. Hence, the possibility of the IF-THEN rules also needs to be calculated and improved.

Acknowledgments

The authors would like to thank the editors, the anonymous reviewers, and Dr. Thayer El-Dajjani for their most constructive comments and suggestions to improve the quality of this paper. This work is supported by Zhejiang Provincial Natural Science Fund under No. (LY13H180012).

References

- [1] A. Taleb-Ahmed and L. Gautier, On information fusion to improve segmentation of MRI sequences, *Information Fusion* **3** (2002), 103–117.
- [2] A.C. Luo, S.W. Chen and C.Y. Fang, Gaussian successive fuzzy integral for sequential multi-decision making, *International Journal of Fuzzy Systems* **2**(17) (2015), 321–336.
- [3] D.A. Kutsenko and V.G. Sinyuk, Inference methods for systems with many fuzzy inputs, *Journal of Computer and Systems Sciences International* **54**(3) (2015), 375–383.
- [4] E. Zervas, A. Mpimpoudis, C. Anagnostopoulos, O. Sekkas and S. Hadjiefthymiades, Multisensor data fusion for fire detection, *Information Fusion* **12** (2011), 150–159.
- [5] F. Herrera and L. Martínez, A 2-tuple fuzzy linguistic representation model for computing with words, *IEEE Trans Fuzzy Systems* **8**(6) (2000), 746–752.
- [6] F. Shi and J. Xu, Emotional cellular-based multi-class fuzzy support vector machines on product's KANSEI extraction, *Appl Math Inf Sci* **6**(1) (2012), 41–49.
- [7] F. Shi, S. Sun and J. Xu, Employing rough sets and association rule mining in KANSEI knowledge extraction, *Inform Sci* **196** (2012), 118–128.
- [8] G. Pavlin, P. de Oude, M. Maris, J. Nunnink and T. Hood, A multi-agent systems approach to distributed Bayesian information fusion, *Information Fusion* **11** (2010), 267–282.
- [9] I. Couso and L. Sánchez, Upper and lower probabilities induced by a fuzzy random variable, *Fuzzy Sets and Systems* **165** (2011), 1–23.
- [10] J. Aisbett, J.T. Rickard and D. Morgenthaler, Multivariate modeling and type-2 fuzzy sets, *Fuzzy Sets and Systems* **163** (2011), 78–95.
- [11] J. Lawry and Y. Tang, On truth-gaps, bipolar belief and the assertability of vague propositions, *Artificial Intelligence* **191–192** (2012), 20–41.
- [12] J. Lawry and Y. Tang, Uncertainty modelling for vague concepts: A prototype theory approach, *Artificial Intelligence* **173** (2009), 1539–1558.
- [13] J. Lawry, Inexact reasoning, PhD Thesis, University of Manchester, UK, 1994, pp. 66–80.
- [14] J. Lawry, Modelling and Reasoning with Vague Concepts, Studies in Computational Intelligence, Springer-Verlag, Berlin, 2005, pp. 3–20.
- [15] L. Kuncheva, Fuzzy' vs 'Non-fuzzy' in combining classifiers designed by boosting, *IEEE Trans on Fuzzy Systems* **11**(6) (2003), 729–741.
- [16] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, *Inform Sci* **8** (1975), 199–249.
- [17] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning-I, *Inform Sci* **8** (1975), 249–299.
- [18] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning-II, *Inform Sci* **8** (1975), 301–357.
- [19] L.A. Zadeh, Toward a generalized theory of uncertainty (GTU)-an outline, *Inform Sci* **172** (2005), 1–40.
- [20] M. Sugeno, *Theory of fuzzy integrals and its applications*, PhD thesis, Tokyo Institute of Technology, 1974.
- [21] N. Sozhamadevi and S. Sathiyamoorthy, A probabilistic fuzzy inference system for modeling and control of nonlinear process, *Arabian Journal for Science and Engineering* **40**(6) (2015), 1777–1791.
- [22] P. McCauley Bush and H. Wang, Fuzzy linear regression models for assessing risks of cumulative trauma disorders, *Fuzzy Set and Systems* **92** (1997), 317–340.
- [23] R.R. Yager and L. Liu, *Classic Works of the Dempster-Shafer Theory of Belief Functions*, A.P. Dempster and G. Shafer, advisory eds., Springer-Verlag, Heidelberg, Germany, 2008.
- [24] R.R. Yager, Conditional approach to possibility-probability fusion, *IEEE Trans Fuzzy Systems* **20**(1) (2012), 46–56.
- [25] R.R. Yager, Entailment principle for measure-based uncertainty, *IEEE Trans Fuzzy Systems* **20**(3) (2012), 526–535.
- [26] R.R. Yager, Set measure directed multi-source information fusion, *IEEE Trans Fuzzy Systems* **19**(6) (2011), 1031–1039.

- [27] S. Destercke, D. Dubois and E. Chojnacki, Possibilistic information fusion using maximal coherent subsets, *IEEE Trans Fuzzy Systems* **17**(1) (2009), 79–92.
- [28] S. Nefti-Meziani, M. Oussalah and M. Soufian, On the use of inclusion structure in fuzzy clustering algorithm in case of Gaussian membership functions, *Journal of Intelligent and Fuzzy Systems* **28**(4) (2015), 1477–1493.
- [29] T. Denoeux, Maximum likelihood estimation from fuzzy data using the EM algorithm, *Fuzzy Sets and Systems* **183** (2011), 72–91.
- [30] T.D. Pham, Fuzzy posterior-probabilistic fusion, *Pattern Recognition* **44** (2011), 1023–1030.
- [31] T.Z. Ahram, P. McCauley-Bush and W. Karwowski, Estimating intrinsic dimensionality using the multi-criteria decision weighted model and the average standard estimator, *Inform Sci* **180** (2010), 2845–2855.
- [32] W. Dou, S. Ruan, Y. Chen, D. Bloyet and J.-M. Constans, A framework of fuzzy information fusion for the segmentation of brain tumor tissues on MR images, *Image and Vision Computing* **25** (2007), 164–171.
- [33] W.L. Tung and C. Quek, A Mamdani-Takagi-Sugeno based linguistic neural-fuzzy inference system for improved interpretability-accuracy representation, *Proc IEEE-FUZZY*, Korea, 2009, PP. 367–372.
- [34] Y. Tang and J. Lawry, Linguistic modelling and information coarsening based on prototype theory and label semantics, *Int J Approx Reasoning* **50** (2009), 1177–1198.
- [35] Y. Tang, A prototype based rule inference system incorporating linear functions, *Fuzzy Sets and System* **161** (2010), 2831–2853.
- [36] Y.Q. Zhang and H.B. Ji, Gaussian mixture reduction based on fuzzy ART for extended target tracking, *Signal Processing* **97** (2014), 232–241.