



King Saud University

Saudi Journal of Biological Sciences

www.ksu.edu.sa
www.sciencedirect.com



ORIGINAL ARTICLE

Analytical fuzzy approach to biological data analysis



Weiping Zhang^a, Jingzhi Yang^b, Yanling Fang^c, Huanyu Chen^c, Yihua Mao^{d,*}, Mohit Kumar^{c,*}

^a Department of Electronic Information Engineering, Nanchang University, 330031 Nanchang, China

^b Mprobe Inc., 94303 Palo Alto, USA

^c Binhai Industrial Technology Research Institute of Zhejiang University, 300301 Tianjin, China

^d Zhejiang University College of Civil Engineering and Architecture, 310027 Hangzhou, China

Received 3 November 2016; revised 5 January 2017; accepted 9 January 2017

Available online 25 January 2017

KEYWORDS

Modeling;
Fuzzy membership functions;
Variational optimization

Abstract The assessment of the physiological state of an individual requires an objective evaluation of biological data while taking into account both measurement noise and uncertainties arising from individual factors. We suggest to represent multi-dimensional medical data by means of an optimal fuzzy membership function. A carefully designed data model is introduced in a completely deterministic framework where uncertain variables are characterized by fuzzy membership functions. The study derives the analytical expressions of fuzzy membership functions on variables of the multivariate data model by maximizing the over-uncertainties-averaged-log-membership values of data samples around an initial guess. The analytical solution lends itself to a practical modeling algorithm facilitating the data classification. The experiments performed on the heartbeat interval data of 20 subjects verified that the proposed method is competing alternative to typically used pattern recognition and machine learning algorithms.

© 2017 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Data mining is increasingly motivating area of research due to an abundance of data facilitated by modern era of information technology. Data mining techniques such as classification and clustering play a vital role in the development of medical decision support systems contributing to improved healthcare quality. The medical decision making problems inherently involve complexities and uncertainties and thus the researchers have advocated the integration of fuzzy methodologies in medical data interpretation. The handling of uncertainties by

* Corresponding authors.

E-mail addresses: maoyihua@zjubh.com (Y. Mao), mohit.kumar@zjubh.com (M. Kumar).

Peer review under responsibility of King Saud University.



capturing of knowledge using fuzzy sets and rules together with an interpretability offered by simple linguistic if-then rules are two most important features of fuzzy methodologies. The fuzzy approaches are commonly applied to medical data classification problems (Fan et al., 2011; Gadaras and Mikhailov, 2009; Nguyen et al., 2015; Papageorgiou, 2011; Seera and Lim, 2014). The mathematical analysis of biomedical signals is performed to construct models identifying the mappings between signal features and the patient's state. The mathematical relationship between signal features and the patient's state is affected by uncertainties arising from individual factors (e.g. related to body conditions) that can't be mathematically taken into account. The fuzzy filters have been previously proposed to alleviate the effect of uncertainties on medical data analysis (Kumar et al., 2007, 2008, 2010) wherein robust estimation algorithms have been applied to design a fuzzy model that identifies the functional relation between physiological parameters and subjective rating scores. Also, stochastic fuzzy modeling and analysis techniques have been introduced to take simultaneously the advantages of Bayesian analysis and fuzzy theory for a mathematical handling of the uncertainties in biomedical signal analysis (Kumar et al., 2010, 2012). A recent work (Kumar et al., 2016) introduced in a rigorous manner a stochastic framework for robust fuzzy filtering and analysis of signals. Although Kumar et al. (2016) introduced modeling and analysis framework is general and rests on strong mathematical foundations, it considers only the signal and thus can't be directly applied to nonsignal multivariate data samples. There remains the need of automated design methods to fully exploit the uncertain handling capabilities of fuzzy systems. The typically used approaches to design the fuzzy sets and systems include evolutionary algorithms (Alcala et al., 2009; Antonelli et al., 2012; Cococcioni et al., 2011; Gacto et al., 2010; Pulkkinen and Koivisto, 2010; Robles et al., 2009), data clustering (Celikyilmaz and Turksen, 2008; Chen and Chen, 2007; Liao et al., 2003; Oh et al., 2003), adaptive filtering (Aliasghary and Arghavani, 2012; Kumar et al., 2006, 2009; Mottaghi-Kashtiban et al., 2008; Simon, 2005), and information theoretic concepts (Aliasghary and Arghavani, 2012; Au et al., 2006; Makrehchi et al., 2003). The determination of fuzzy membership functions remains a challenge as membership functions, due to the non-linearity of the problem, can't be optimized analytically. Thus, most design methods of fuzzy membership functions lack in mathematical theory and are based on numerical algorithms which might be slow and inexact. Recently, (Kumar et al., 2016) introduced an analytical approach for the determination of fuzzy membership functions using the variational optimization method. The proposed analytical approach of (Kumar et al., 2016) allows to mathematically incorporate the given modeling scenario in fuzzy membership functions' design problem and thus can be potentially extended to medical data modeling scenario. The authors observe that the application of fuzzy paradigm in medicine, despite being an extensively studied area, doesn't provide a rigorous analytically derived methodology or approach to interpret medical data while taking mathematically into account the measurement noise as well as the individuality.

The medical data are multi-dimensional whose good representation by means of fuzzy membership functions is the aim of the mathematical theory presented in this study. This text introduces a data model that takes into account both

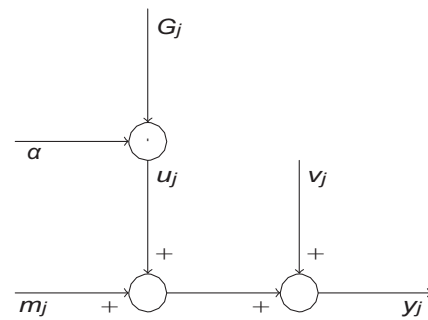


Figure 1 An uncertain signal model for a scalar y_j .

measurement noise and uncertainties arising from individuality related factors. A multivariate data sample, represented as $y = [y_1 \cdots y_P]^T \in RP$, is assumed to be generated by an uncertain signal model displayed in Fig. 1. It is assumed an uncertain signal model for a scalar y_j . Here, y_j is the observed value of an unknown scalar m_j being affected by measurement noise v_j and uncertainty u_j . The uncertainty u_j (equal to the dot product of $G_j \in RK$ and $\alpha \in RK$) is being generated by a linear combination of K different sources: $(\alpha_1, \cdots, \alpha_K)$ that the j th element of y is generated as

$$y_j = m_j + u_j + v_j$$

where v_j is the measurement noise, u_j is the uncertainty affecting the model, and m_j is an unobserved scalar variable. The uncertainties are assumed to be generated by linearly transforming a K -dimensional ($K \leq P$) vector $\alpha = \alpha = [\alpha_1 \cdots \alpha_K]^T \in \mathbb{R}^K$ as follows:

$$\begin{bmatrix} u_1 \\ \vdots \\ u_P \end{bmatrix} = \begin{bmatrix} G_{11} & \cdots & G_{1K} \\ \vdots & & \vdots \\ G_{P1} & \cdots & G_{PK} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_K \end{bmatrix}.$$

Defining $G_j = [G_{j1} \cdots G_{jK}]^T \in \mathbb{R}^K$, u_j can be expressed as the dot product of G_j and α , i.e.,

$$u_j = (G_j)^T \alpha.$$

Our approach is of

1. treating all the variables (appearing in the uncertain signal model of Fig.1) as uncertain being characterized by fuzzy membership functions.
2. assuming that medical data, under the given status of a patient, is generated by a finite mixture of uncertain signal models of the type that of Fig. 1.
3. determining the fuzzy membership functions on variables with the help of experimentally measured data samples in an analytical manner using variational optimization (Kumar et al., 2016).

The approach results in a tractable solution to model the multivariate data samples by means of fuzzy membership functions and thus medical decision support systems can be built up on the top of the data models.

The modeling of data using a finite mixture of signal models of the type of Fig. 1 is typically considered in a stochastic setting assuming variables as random (i.e. characterized by

probability distribution functions) and Bayesian framework is commonly used for the inference of posterior distributions. The originality of this study lies in solving the modeling problem in a completely deterministic framework where fuzzy membership functions are defined over variables to characterize uncertainties about their values. The optimal shapes of fuzzy membership functions are determined via analytically maximizing the “over uncertainties averaged log membership” values of data samples around an initial guess. The maximization problem is analytically solved using variational optimization as suggested initially in Kumar et al. (2016). The contribution of this study is to derive the analytical expressions of fuzzy membership functions on variables of the multivariate data model leading to the development of a classification algorithm. It is demonstrated through experimental data that our approach is competing alternative to typically used classification algorithms including “k-nearest neighbors”, “support vector machines”, “decision tree”, “random forest”, “AdaBoost”, “Gaussian naive Bayes”, “linear discriminant analysis”, and “quadratic discriminant analysis”. The better classification performance of our approach is attributed to the efficient modeling of the data distribution in multi-parametric space. The significance of this work is that the analytically derived expressions for fuzzy membership functions for representing uncertainties associated with medical data would facilitate a system theoretic approach to mathematically design the medical expert systems. This would provide researchers, unlike typically used ad-hoc numerical algorithms, a mathematical theory on fuzzy membership functions’ applications in medicine.

This text is organized into sections. Section 2 introduces an uncertain model of multivariate data and an analytical solution for optimizing the data model is provided in Section 3. A practical algorithm, based on the derived analytical solution, is stated in Section 4 for the modeling of multivariate data samples. Section 5 applies the proposed approach on the experimental heartbeat interval data of 20 subjects followed by concluding remarks in Section 6.

2. An uncertain model of multivariate data

By an uncertain model, it is meant that system variables are characterized by fuzzy membership functions. Despite the availability of a wide range of fuzzy membership function types, only following two types of fuzzy membership functions are chosen to model the variables for keeping the analysis in its most basic form:

Definition 1. (Gaussian’s membership function (Kumar et al., 2016)). The Gaussian membership function on a vector $x \in \mathbb{R}^n$, with mean equal to m_x and precision equal to Λ_x , is defined as

$$\mu(x; m_x, \Lambda_x) = \exp\left(-\frac{1}{2}(x - m_x)^T \Lambda_x (x - m_x)\right), \quad m_x \in \mathbb{R}^n, \Lambda_x^{-1} > 0.$$

Definition 2 (Gamma membership function (Kumar et al., 2016)). The Gamma membership function on a non-negative scalar z can be defined as

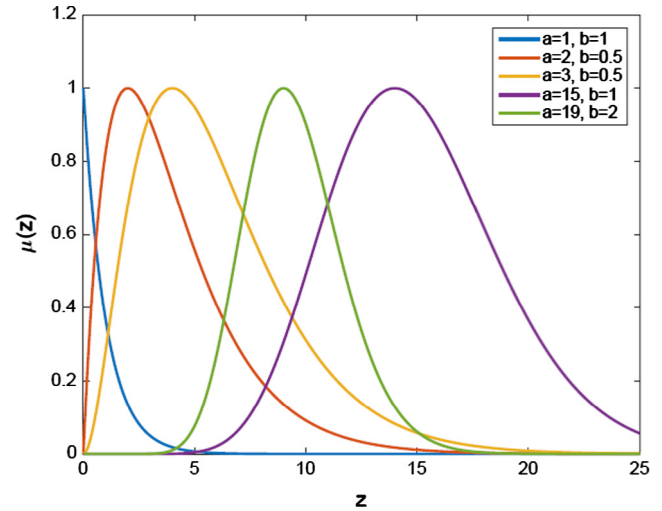


Figure 2 A few examples of Gamma membership functions (Kumar et al., 2016).

$$\mu(z; a, b) = \left(\frac{b}{a-1}\right)^{a-1} \exp(a-1)(z)^{a-1} \exp(-bz), \quad a \geq 1, b > 0.$$

A few examples of this type of membership functions for different values of a and b are provided in Fig. 2. The parameter a is referred to as the shape parameter and b is referred to as the rate parameter (i.e. the reciprocal of the scale parameter). The peak of the membership function is given at $(a-1)/b$. The skewness of the membership function is inversely proportional to the value of a . The Gamma membership function can alternatively be represented as

$$\mu(z; r, s) = (s)^r \exp(r)(z)^r \exp(-srz), \quad r \geq 0, s > 0$$

The relations between the parameters of two forms of Gamma membership functions are as follows:

$$r = a - 1, \quad s = b/(a - 1).$$

All of the variables, appearing in Fig. 1, are assigned carefully either of Gaussian or Gamma membership function in Definition 3, 4, 5, 6, 7, and 8.

Definition 3 (Fuzzy membership function on v_j). The fuzzy membership function on $v_j \in \mathbb{R}$ is defined as zero-mean Gaussian with scaled precisions as

$$\mu(j; \lambda_y, z_{y_j}) = \exp\left(-\frac{\lambda_y z_{y_j}^2}{2j}\right) \quad (1)$$

where $\lambda_y > 0$ is the precision scaled by $z_{y_j} > 0$. The uncertainties of λ_y and z_{y_j} are characterized by the following Gamma membership functions:

$$\mu(\lambda_y; a_{\lambda_y}, b_{\lambda_y}) = \left(\frac{b_{\lambda_y}}{a_{\lambda_y} - 1}\right)^{a_{\lambda_y} - 1} \exp(a_{\lambda_y} - 1)(\lambda_y)^{a_{\lambda_y} - 1} \exp(-b_{\lambda_y} \lambda_y),$$

$$a_{\lambda_y} \geq 1, \quad b_{\lambda_y} > 0 >$$

$$\mu(z_{y_j}; r_y, s_y) = (s_y)^{r_y} \exp(r_y)(z_{y_j})^{r_y} \exp(-r_y s_y z_{y_j}), \quad r_y \geq 0, \quad s_y > 0.$$

Here, $r_y > 0$, and $s_y > 0$ are uncertain as well as characterized by the following Gamma membership functions:

$$\mu(r_y; a_{r_y}, b_{r_y}) = \left(\frac{b_{r_y}}{a_{r_y} - 1}\right)^{a_{r_y}-1} \exp(a_{r_y} - 1)(r_y)^{a_{r_y}-1} \exp(-b_{r_y}r_y),$$

$$a_{r_y} \geq 1, b_{r_y} > 0$$

$$\mu(s_y; a_{s_y}, b_{s_y}) = \left(\frac{b_{s_y}}{a_{s_y} - 1}\right)^{a_{s_y}-1} \exp(a_{s_y} - 1)(s_y)^{a_{s_y}-1} \exp(-b_{s_y}s_y),$$

$$a_{s_y} \geq 1, b_{s_y} > 0$$

Definition 4 (Fuzzy membership function on y_j). The fuzzy membership function on $y_j \in \mathbb{R}$, for a given $(m_j, G_j, \alpha, \lambda_y, z_{y_j})$, is defined as

$$\mu(y_j; m_j, G_j, \alpha, \lambda_y, z_{y_j}) = \exp\left(-\frac{\lambda_y z_{y_j}}{2}(y_j - m_j - (G_j)^T \alpha)^2\right).$$

The membership function on y_j is derived by replacing j in (1) by $y_j - m_j - (G_j)^T \alpha$.

Definition 5 (Fuzzy membership function on y). The multivariate fuzzy membership function on $y \in \mathbb{R}^P$, for a given $(\{m_j\}_{j=1}^P, \{G_j\}_{j=1}^P, \alpha, \lambda_y, \{z_{y_j}\}_{j=1}^P)$, is defined as the product of its individual elements' membership functions as

$$\begin{aligned} \mu(y; \{m_j\}_{j=1}^P, \{G_j\}_{j=1}^P, \alpha, \lambda_y, \{z_{y_j}\}_{j=1}^P) &= \prod_{j=1}^P \mu(y_j; m_j, G_j, \alpha, \lambda_y, z_{y_j}) \\ &= \exp\left(-\frac{\lambda_y}{2} \sum_{j=1}^P z_{y_j} (y_j - m_j - (G_j)^T \alpha)^2\right) \end{aligned}$$

Definition 6 (Fuzzy membership function on m). The multivariate fuzzy membership function on $m = [m_1 \cdots m_P]^T \in \mathbb{R}^P$ is defined as Gaussian as

$$\mu(m; m_o, \Lambda_o) = \exp\left(-\frac{1}{2}(m - m_o)^T \Lambda_o (m - m_o)\right),$$

$$m_o \in \mathbb{R}^P, \Lambda_o > 0.$$

Definition 7 (Fuzzy membership function on α). The multivariate fuzzy membership function on $\alpha \in \mathbb{R}^K$ is defined as zero-mean Gaussian with precision equal to unity matrix as

$$\mu(\alpha) = \exp\left(-\frac{1}{2}(\alpha)^T \alpha\right).$$

Definition 8 (Fuzzy membership function on G_j). The multivariate fuzzy membership function on $G_j = [G_{j1} \cdots G_{jk}]^T \in \mathbb{R}^K$ is defined as zero-mean Gaussian as

$$\mu(G_j; \{\phi_k\}_{k=1}^K) = \exp\left(-\frac{1}{2} \sum_{k=1}^K (G_{jk})^2 \phi_k\right)$$

where $\phi_k > 0$ is the precision of k th element of G_j and is uncertain characterized by the following Gamma membership function:

$$\mu(\phi_k; a_\phi, b_\phi) = \left(\frac{b_\phi}{a_\phi - 1}\right)^{a_\phi-1} \exp(a_\phi - 1)(\phi_k)^{a_\phi-1} \exp(-b_\phi \phi_k),$$

$$a_\phi \geq 1, b_\phi > 0.$$

To model the multivariate data sample distributed arbitrarily in P -dimensional data space, a mixture of finite number of uncertain signal models is considered in Definition 9.

Definition 9 (Fuzzy membership of y as a finite mixture of uncertain signal models). The fuzzy membership function on $y = [y_1 \cdots y_P]^T \in \mathbb{R}^P$, for a given $(\{\pi_i\}_{i=1}^C, \Omega)$, is defined as a mixture of C different uncertain signal models as

$$\begin{aligned} \mu(y; \{\pi_i\}_{i=1}^C, \Omega) &= \exp\left(-\frac{\pi_1}{2} \lambda_y^1 \sum_{j=1}^P z_{y_j}^1 (y_j - m_j^1 - (G_j)^T \alpha)^2 \cdots \right. \\ &\quad \left. - \frac{\pi_C}{2} \lambda_y^C \sum_{j=1}^P z_{y_j}^C (y_j - m_j^C - (G_j)^T \alpha)^2\right) \end{aligned}$$

where $\pi_i \in [0, 1]$ is the mixing proportion of the i th uncertain signal model with $\sum_{i=1}^C \pi_i = 1$, and Ω is a set of parameters defined as

$$\Omega = \{\{\alpha^i\}_{i=1}^C, \{G_j\}_{j=1}^P, \{\phi_k\}_{k=1}^K, \{m^i\}_{i=1}^C, \{z_{y_j}^i\}_{j=1}^P\}_{i=1}^C, r_y, s_y, \{\lambda_y^i\}_{i=1}^C\}$$

where $\alpha^i \in \mathbb{R}^K$ ($K \leq P$) is uncertain characterized by the following Gaussian membership function

$$\mu(\alpha^i) = \exp\left(-\frac{1}{2}(\alpha^i)^T \alpha^i\right);$$

$G_j = [G_{j1} \cdots G_{jk}]^T \in \mathbb{R}^K$ is uncertain characterized by the following Gaussian membership function

$$\mu(G_j; \{\theta_k\}_{k=1}^K) = \exp\left(-\frac{1}{2} \sum_{k=1}^K (G_{jk})^2 \theta_k\right), \quad \theta_k > 0$$

$\theta_k > 0$ is uncertain characterized by the following Gamma membership function:

$$\mu(\theta_k; \alpha_\theta, b_\theta) = \left(\frac{b_\theta}{\alpha_\theta - 1}\right)^{\alpha_\theta-1} \exp(\alpha_\theta - 1)(\theta_k)^{\alpha_\theta-1} \exp(-b_\theta \theta_k),$$

$$\alpha_\theta \geq 1, b_\theta > 0;$$

$m^i = [m_1^i \cdots m_P^i]^T \in \mathbb{R}^P$ is uncertain characterized by the following Gaussian membership function:

$$\mu(m^i; m_o^i, \Lambda_o^i) = \exp\left(-\frac{1}{2}(m^i - m_o^i)^T \Lambda_o^i (m^i - m_o^i)\right),$$

$$m_o^i \in \mathbb{R}^K, \Lambda_o^i > 0;$$

$z_{y_j}^i > 0$ is uncertain scalar characterized by the following Gamma membership function:

$$\mu(z_{y_j}^i; r_y, s_y) = (s_y)^{r_y} \exp(r_y)(z_{y_j}^i)^{r_y} \exp(-r_y s_y z_{y_j}^i), \quad r_y \geq 1, s_y > 0;$$

r_y is uncertain characterized by the following Gamma membership function:

$$\mu(r_y; a_{r_y}, b_{r_y}) = \left(\frac{b_{r_y}}{a_{r_y} - 1}\right)^{a_{r_y}-1} \exp(a_{r_y} - 1)(r_y)^{a_{r_y}-1} \exp(-b_{r_y}r_y),$$

$$a_{r_y} \geq 1, b_{r_y} > 0;$$

s_y is uncertain characterized by the following Gamma membership function:

$$\mu(s_y; a_{s_y}, b_{s_y}) = \left(\frac{b_{s_y}}{a_{s_y}} \right)^{a_{s_y}-1} \exp(a_{s_y} - 1) (S_y)^{a_{s_y}-1} \exp(-b_{s_y} S_y), \quad a_{s_y} \geq 1, b_{s_y} > 0;$$

$\lambda_y^i > 0$ is uncertain scalar characterized by the following Gamma membership function:

$$\mu(\lambda_y^i; a_{\lambda_y}, b_{\lambda_y}) = \left(\frac{b_{\lambda_y}}{a_{\lambda_y}} \right)^{a_{\lambda_y}-1} \exp(a_{\lambda_y} - 1) (\lambda_y^i)^{a_{\lambda_y}-1} \exp(-b_{\lambda_y} \lambda_y^i), \quad a_{\lambda_y} \geq 1, b_{\lambda_y} > 0.$$

3. Analytical optimization of mixture of uncertain signal models

Given N data samples, $\{y^n\}_{n=1}^N$, the aim is to define the multivariate fuzzy membership function on y in an “optimal” manner. The approach is to optimize the fuzzy membership function (defined on y by Definition 1) with respect to $\{\pi_i\}_{i=1}^C$ while taking into account the uncertainties of the parameters represented z_{yj}^i by set Ω . To take into account the uncertainties of the parameters represented by the set Ω , the “optimal” membership functions on the parameters must be first determined. For this, assume that $q(\alpha^i)$, $q(G_j)$, $q(\emptyset_k)$, $q(m^i)$, $q(z_{yj}^i)$, $q(r_y)$, $q(s_y)$, and $q(\lambda_y^i)$ are arbitrary fuzzy membership functions on α^i , G_j , \emptyset_k , m^i , r_y , s_y and λ_y^i respectively. Define a function, $q(\Omega)$, as follows

$$q(\Omega) = \left\{ \prod_{i=1}^C q(\alpha^i) \right\} \left\{ \prod_{j=1}^P q(G_j) \right\} \left\{ \prod_{k=1}^K q(\emptyset_k) \right\} \left\{ \prod_{i=1}^C q(m^i) \right\} \times \left\{ \prod_{i=1}^C \prod_{j=1}^P q(z_{yj}^i) \right\} q(r_y) q(s_y) \left\{ \prod_{i=1}^C q(\lambda_y^i) \right\}$$

Define a differential functional, $\partial\Omega$, as follows

$$\partial\Omega = \left\{ \prod_{i=1}^C \partial\alpha^i \right\} \left\{ \prod_{j=1}^P \partial G_j \right\} \left\{ \prod_{k=1}^K \partial\emptyset_k \right\} \left\{ \prod_{i=1}^C \partial m^i \right\} \left\{ \prod_{i=1}^C \prod_{j=1}^P \partial z_{yj}^i \right\} \partial(r_y) \partial(s_y) \times \left\{ \prod_{i=1}^C \partial\lambda_y^i \right\}$$

Define a differential functional, $\mu(\Omega)$, as follows

$$\mu(\Omega) = \left\{ \prod_{i=1}^C \mu(\alpha^i) \right\} \left\{ \prod_{j=1}^P \mu(G_j; \{\emptyset_k\}_{k=1}^K) \right\} \left\{ \prod_{k=1}^K \mu(\emptyset_k; a_{\emptyset}, b_{\emptyset}) \right\} \times \left\{ \prod_{i=1}^C \mu(m^i; m_o, \Lambda_o) \right\} \times \left\{ \prod_{i=1}^C \prod_{j=1}^P \mu(z_{yj}^i; r_y, s_y) \right\} \mu(r_y; a_{r_y}, b_{r_y}) \times \mu(s_y; a_{s_y}, b_{s_y}) \left\{ \prod_{i=1}^C \mu(\lambda_y^i; a_{\lambda_y}, b_{\lambda_y}) \right\}$$

The optimization process maximizes an objective functional, \mathcal{F} , defined as

$$\begin{aligned} \mathcal{F} & \left(\left\{ \pi_i^n \right\}_{i=1}^C \right), q(\Omega) \\ & = \frac{1}{\int \partial\Omega q(\Omega)} \int \partial\Omega q(\Omega) \frac{\sum_{n=1}^N \log(\mu(y^n; \{\pi_i^n\}_{i=1}^C, \Omega))}{N} \\ & \quad - \frac{1}{\int \partial\Omega q(\Omega)} \int \partial\Omega q(\Omega) \log \left(\frac{q(\Omega)}{\mu(\Omega)} \right) - \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^C \pi_i^n \log \left(\frac{\pi_i^n}{\pi_i^o} \right) \end{aligned} \quad (2)$$

\mathcal{F} is maximized with respect to $q(\alpha^i)$, $q(G_j)$, $q(\emptyset_k)$, $q(m^i)$, $q(z_{yj}^i)$, $q(r_y)$, $q(s_y)$, and $q(\lambda_y^i)$ and $\{\pi_i^n\}_{i=1}^C$ under the following constraints:

1. *Fixed Integral Constraints on Membership Functions:*

$$\begin{aligned} \int \partial\alpha^i q(\alpha^i) & = k_{\alpha^i} > 0, \\ \int \partial G_j q(G_j) & = k_{G_j} > 0, \quad \int \partial\emptyset_k q(\emptyset_k) = k_{\emptyset_k} > 0, \\ \int \partial m^i q(m^i) & = k_{m^i} > 0, \\ \int \partial z_{yj}^i q(z_{yj}^i) & = k_{z_{yj}^i} > 0, \quad \int \partial r_y q(r_y) = k_{r_y} > 0, \\ \int \partial s_y q(s_y) & = k_{s_y} > 0, \quad \int \partial\lambda_y^i q(\lambda_y^i) = k_{\lambda_y^i} > 0. \end{aligned}$$

2. *Unity Maximum Value Constraints on Membership Functions:* The values of $k_{\alpha^i}, k_{G_j}, k_{\emptyset_k}, k_{m^i}, k_{z_{yj}^i}, k_{r_y}, k_{s_y}$, and $k_{\lambda_y^i}$ are so chosen such that maximum value of $q(\alpha^i)$, $q(G_j)$, $q(\emptyset_k)$, $q(m^i)$, $q(z_{yj}^i)$, $q(r_y)$, $q(s_y)$, and $q(\lambda_y^i)$ is equal to one.

3. *Unity Sum Constraint on Mixing Proportions:*

$$\sum_{i=1}^C \pi_i^n = 1, \pi_i^n \in [0, 1].$$

The first term of \mathcal{F} computes the averaged log-membership value of data samples when the average is taken over uncertain parameters Ω being modeled by membership function $q(\Omega)$. The second term of \mathcal{F} regularizes the maximization problem toward initial guess $\mu(\Omega)$. The third term of \mathcal{F} regularizes the estimation of π_i^n toward initial guess π_i^o .

Result 1. The analytical expressions for variational membership functions, that maximize \mathcal{F} under Fixed Integral and Unity Maximum Value Constrains, are

$$q^*(\alpha^i) = \exp \left(-\frac{1}{2} (\alpha^i - \hat{m}_{\alpha^i})^T \hat{\Lambda}_{\alpha^i} (\alpha^i - \hat{m}_{\alpha^i}) \right), \quad (3)$$

$$\hat{\Lambda}_{\alpha^i} = I + \sum_{n=1}^N \sum_{j=1}^P \frac{\pi_j^n}{N} \frac{\hat{a}_{\alpha^i}}{b_{\alpha^i}} \frac{\hat{a}_{z_{yj}^i}}{b_{z_{yj}^i}} (\hat{m}_{G_j} (\hat{m}_{G_j})^T + (\hat{\Lambda}_{G_j})^{-1})$$

$$\hat{m}_{\alpha^i} = (\hat{\Lambda}_{\alpha^i})^{-1} \left\{ \sum_{n=1}^N \sum_{j=1}^P \frac{\pi_j^n}{N} \frac{\hat{a}_{\alpha^i}}{b_{\alpha^i}} \frac{\hat{a}_{z_{yj}^i}}{b_{z_{yj}^i}} (y_j^n - I_j^P \hat{m}_{m^i}) \hat{m}_{G_j} \right\} \quad (4)$$

$$q^*(G_j) = \exp \left(-\frac{1}{2} (G_j - \hat{m}_{G_j})^T \hat{\Lambda}_{G_j} (G_j - \hat{m}_{G_j}) \right),$$

$$\hat{\Lambda}_{G_j} = \begin{bmatrix} \hat{a}_{\phi_1} \\ \hat{b}_{\phi_1} \\ \hat{a}_{\phi_2} \\ \hat{b}_{\phi_2} \\ \vdots \\ \hat{a}_{\phi_K} \\ \hat{b}_{\phi_K} \end{bmatrix} \prod_{n=1}^N \prod_{i=1}^C \frac{\hat{\pi}_i^n \hat{a}_{\lambda_y^i} \hat{a}_{z_{y_j}^i}}{\hat{b}_{\lambda_y^i} \hat{b}_{z_{y_j}^i}} (\hat{m}_{\alpha^i} (\hat{m}_{\alpha^i})^T + (\hat{\Lambda}_{\alpha^i})^{-1})$$

$$\hat{m}_{G_j} = (\hat{\Lambda}_{G_j})^{-1} \left\{ \prod_{n=1}^N \prod_{i=1}^C \frac{\hat{\pi}_i^n \hat{a}_{\lambda_y^i} \hat{a}_{z_{y_j}^i}}{\hat{b}_{\lambda_y^i} \hat{b}_{z_{y_j}^i}} (y_j^n - I_j^P \hat{m}_{m^i}) \prod_{\alpha^i} \right\}$$

(5), (6)

$$q^*(\phi_k) = \left(\frac{\hat{b}_{\phi_k}}{\hat{a}_{\phi_k} - 1} \right)^{\hat{a}_{\phi_k} - 1} \exp(\hat{a}_{\phi_k} - 1) (\phi_k)^{\hat{a}_{\phi_k} - 1} \exp(-\hat{b}_{\phi_k} \phi_k),$$

$$\hat{a}_{\phi_k} = a_\phi$$

(7)

$$q^*(z_{y_j}^i) = \left(\frac{\hat{b}_{z_{y_j}^i}}{\hat{a}_{z_{y_j}^i} - 1} \right)^{\hat{a}_{z_{y_j}^i} - 1} \exp(\hat{a}_{z_{y_j}^i} - 1) (z_{y_j}^i)^{\hat{a}_{z_{y_j}^i} - 1} \exp(-\hat{b}_{z_{y_j}^i} z_{y_j}^i). \quad (13)$$

$$\hat{a}_{z_{y_j}^i} = \frac{\hat{a}_{r_y}}{\hat{b}_{r_y}} + 1$$

$$\hat{b}_{\phi_k} = b_\phi + \frac{1}{2} \sum_{j=1}^P \{ (I_k^K \hat{m}_{G_j})^2 + \text{Tr}((\hat{\Lambda}_{G_j})^{-1} (I_k^K)^T I_k^K) \}$$

(8)

$$\hat{b}_{z_{y_j}^i} = \frac{\hat{a}_{r_y}}{\hat{b}_{r_y}} \frac{\hat{a}_{s_y}}{\hat{b}_{s_y}} + \frac{1}{2} \sum_{n=1}^N \frac{\hat{\pi}_i^n \hat{a}_{\lambda_y^i}}{\hat{b}_{\lambda_y^i}} \left((y_j^n - I_j^P \hat{m}_{m^i} - (\hat{m}_{G_j})^T \hat{m}_{x^i})^2 \right. \\ \left. + \text{Tr}((\hat{\Lambda}_{m^i})^{-1} (I_j^P)^T I_j^P) + (\hat{m}_{x^i})^T (\hat{\Lambda}_{G_j})^{-1} \hat{m}_{x^i} + (\hat{m}_{G_j})^T (\hat{\Lambda}_{x^i})^{-1} \hat{m}_{G_j} \right. \\ \left. + \text{Tr}((\hat{\Lambda}_{G_j})^{-1} (\hat{\Lambda}_{x^i})^{-1}) \right) \quad (14)$$

$$q^*(m^i) = \exp\left(-\frac{1}{2} (m^i - \hat{m}_{m^i})^T \hat{\Lambda}_{m^i} (m^i - \hat{m}_{m^i})\right),$$

$$\hat{\Lambda}_{m^i} = \Lambda_o^i + \sum_{n=1}^N \prod_{i=1}^C \frac{\hat{\pi}_i^n \hat{a}_{\lambda_y^i}}{\hat{b}_{\lambda_y^i}} \begin{bmatrix} \hat{a}_{z_{y_1}^i} \\ \hat{b}_{z_{y_1}^i} \\ \vdots \\ \hat{a}_{z_{y_P}^i} \\ \hat{b}_{z_{y_P}^i} \end{bmatrix}$$

(9)

$$\hat{m}_{m^i} = (\hat{\Lambda}_{m^i})^{-1} \left\{ \prod_{o=1}^P m_o^i \right. \\ \left. + \sum_{n=1}^N \prod_{i=1}^C \frac{\hat{\pi}_i^n \hat{a}_{\lambda_y^i}}{\hat{b}_{\lambda_y^i}} \begin{bmatrix} \hat{a}_{z_{y_1}^i} \\ \hat{b}_{z_{y_1}^i} \\ \vdots \\ \hat{a}_{z_{y_P}^i} \\ \hat{b}_{z_{y_P}^i} \end{bmatrix} \left(y_j^n - [(\hat{m}_{G_1})^T \hat{m}_{\alpha^i} \cdots (\hat{m}_{G_P})^T \hat{m}_{\alpha^i}]^T (1) \right) \right\}$$

$$q^*(\lambda_y^i) = \left(\frac{\hat{b}_{\lambda_y^i}}{\hat{a}_{\lambda_y^i} - 1} \right)^{\hat{a}_{\lambda_y^i} - 1} \exp(\hat{a}_{\lambda_y^i} - 1) (\lambda_y^i)^{\hat{a}_{\lambda_y^i} - 1} \exp(-\hat{b}_{\lambda_y^i} \lambda_y^i). \quad (11)$$

$$\hat{a}_{\lambda_y^i} = a_{\lambda_y}$$

$$\hat{b}_{\lambda_y^i} = b_{\lambda_y} + \frac{1}{2} \sum_{n=1}^N \sum_{j=1}^P \frac{\hat{\pi}_i^n \hat{a}_{z_{y_j}^i}}{\hat{b}_{z_{y_j}^i}} \left((y_j^n - I_j^P \hat{m}_{m^i} - (\hat{m}_{G_j})^T \hat{m}_{x^i})^2 \right. \\ \left. + \text{Tr}((\hat{\Lambda}_{m^i})^{-1} (I_j^P)^T I_j^P) + (\hat{m}_{x^i})^T (\hat{\Lambda}_{G_j})^{-1} \hat{m}_{x^i} \right. \\ \left. + (\hat{m}_{G_j})^T (\hat{\Lambda}_{x^i})^{-1} \hat{m}_{G_j} + \text{Tr}((\hat{\Lambda}_{G_j})^{-1} (\hat{\Lambda}_{x^i})^{-1}) \right) \quad (12)$$

$$f_i^n = -\frac{1}{2} \sum_{j=1}^P \frac{\hat{a}_{\lambda_y^i}}{\hat{b}_{\lambda_y^i}} \frac{\hat{a}_{z_{y_j}^i}}{\hat{b}_{z_{y_j}^i}} \left((y_j^n - I_j^P \hat{m}_{m^i} - (\hat{m}_{G_j})^T \hat{m}_{x^i})^2 \right. \\ \left. + \text{Tr}((\hat{\Lambda}_{m^i})^{-1} (I_j^P)^T I_j^P) + (\hat{m}_{x^i})^T (\hat{\Lambda}_{G_j})^{-1} \hat{m}_{x^i} \right. \\ \left. + (\hat{m}_{G_j})^T (\hat{\Lambda}_{x^i})^{-1} \hat{m}_{G_j} + \text{Tr}((\hat{\Lambda}_{G_j})^{-1} (\hat{\Lambda}_{x^i})^{-1}) \right) \quad (15)$$

$$\hat{\pi}_i^n = \frac{\pi_i^o \exp(f_i^n)}{\sum_{i=1}^C \pi_i^o \exp(f_i^n)}. \quad (16)$$

$$q^*(r_y) = \left(\frac{\hat{b}_{r_y}}{\hat{a}_{r_y} - 1} \right)^{\hat{a}_{r_y} - 1} \exp(\hat{a}_{r_y} - 1)(r_y)^{\hat{a}_{r_y} - 1} \exp(-\hat{b}_{r_y} r_y), \quad (17)$$

$$\hat{a}_{r_y} = a_{r_y}$$

$$\hat{b}_{r_y} = b_{r_y} + \frac{\hat{a}_{s_y}}{\hat{b}_{s_y}} \sum_{i=1}^C \sum_{j=1}^P \frac{\hat{a}_{z_{ij}}}{\hat{b}_{z_{ij}}} - CP\{\psi(\hat{a}_{s_y}) - \log(\hat{b}_{s_y})\} - CP$$

$$- \sum_{i=1}^C \sum_{j=1}^P \{\psi(\hat{a}_{z_{ij}}) - \log(\hat{b}_{z_{ij}})\} \quad (18)$$

$$q^*(s_y) = \left(\frac{\hat{b}_{s_y}}{\hat{a}_{s_y} - 1} \right)^{\hat{a}_{s_y} - 1} \exp(\hat{a}_{s_y} - 1)(s_y)^{\hat{a}_{s_y} - 1} \exp(-\hat{b}_{s_y} s_y), \quad (19)$$

$$\hat{a}_{s_y} = a_{s_y} + CP \frac{\hat{a}_{r_y}}{\hat{b}_{r_y}}$$

$$\hat{b}_{s_y} = b_{s_y} + \frac{\hat{a}_{r_y}}{\hat{b}_{r_y}} \sum_{i=1}^C \sum_{j=1}^P \frac{\hat{a}_{z_{ij}}}{\hat{b}_{z_{ij}}} \quad (20)$$

Once the membership functions representing the uncertainties on the parameters have been optimally determined, the optimal multivariate fuzzy membership function on $y = [y_1 \cdots y_P]^T \in \mathbb{R}^P$ is defined by averaging over the uncertainties such that

$$\mu^*(y) \propto \exp \langle \log(\mu(y; \{\pi_i\}_{i=1}^C, \Omega)) \rangle_{q^*(\Omega)}$$

where

$$\pi_i = \frac{\pi_i^o \exp(f_i)}{\sum_{i=1}^C \pi_i^o \exp(f_i)}$$

$$f_i = -\frac{1}{2} \sum_{j=1}^P \frac{\hat{a}_{z_{ij}}}{\hat{b}_{z_{ij}}} \frac{\hat{a}_{z_{ij}}}{\hat{b}_{z_{ij}}} \left((y_j - I_j^p \hat{m}_{m^i} - (\hat{m}_{G_j})^T \hat{m}_{z^i})^2 \right.$$

$$+ \text{Tr}((\hat{\Lambda}_{m^i})^{-1} (I_j^p)^T I_j^p) + (\hat{m}_{z^i})^T (\hat{\Lambda}_{G_j})^{-1} \hat{m}_{z^i}$$

$$\left. + (\hat{m}_{G_j})^T (\hat{\Lambda}_{z^i})^{-1} \hat{m}_{G_j} + \text{Tr}((\hat{\Lambda}_{G_j})^{-1} (\hat{\Lambda}_{z^i})^{-1}) \right).$$

After evaluating the integral, $\langle \log(\mu(y; \{\pi_i\}_{i=1}^C, \Omega)) \rangle_{q^*(\Omega)}$, the expression of the optimal membership function on y is as follows:

$$\log(\mu^*(y)) \propto -\frac{1}{2} \sum_{i=1}^C \sum_{j=1}^P \pi_i \frac{\hat{a}_{z_{ij}}}{\hat{b}_{z_{ij}}} \frac{\hat{a}_{z_{ij}}}{\hat{b}_{z_{ij}}} \left((y_j - I_j^p \hat{m}_{m^i} - (\hat{m}_{G_j})^T \hat{m}_{z^i})^2 \right.$$

$$+ \text{Tr}((\hat{\Lambda}_{m^i})^{-1} (I_j^p)^T I_j^p) + (\hat{m}_{z^i})^T (\hat{\Lambda}_{G_j})^{-1} \hat{m}_{z^i}$$

$$\left. + (\hat{m}_{G_j})^T (\hat{\Lambda}_{z^i})^{-1} \hat{m}_{G_j} + \text{Tr}((\hat{\Lambda}_{G_j})^{-1} (\hat{\Lambda}_{z^i})^{-1}) \right).$$

Finally, the constant of proportionality is chosen equal to one resulting in

$$\mu^*(y) = \exp \left(-\frac{1}{2} \sum_{i=1}^C \sum_{j=1}^P \pi_i \frac{\hat{a}_{z_{ij}}}{\hat{b}_{z_{ij}}} \frac{\hat{a}_{z_{ij}}}{\hat{b}_{z_{ij}}} \left\{ (y_j - I_j^p \hat{m}_{m^i} - (\hat{m}_{G_j})^T \hat{m}_{z^i})^2 \right. \right.$$

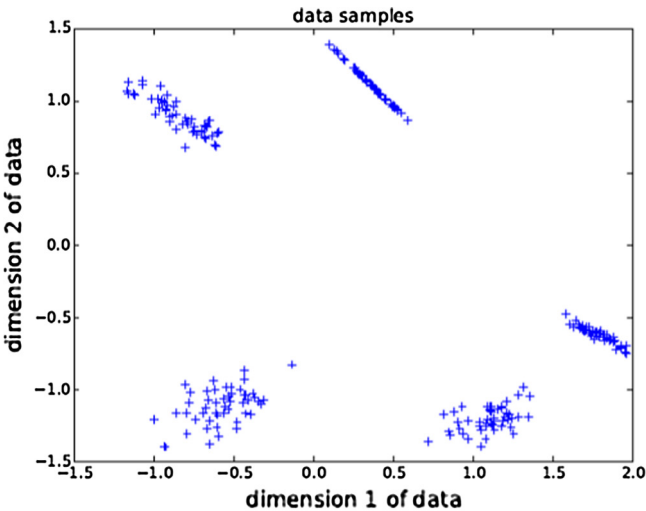
$$+ \text{Tr}((\hat{\Lambda}_{m^i})^{-1} (I_j^p)^T I_j^p) + (\hat{m}_{z^i})^T (\hat{\Lambda}_{G_j})^{-1} \hat{m}_{z^i}$$

$$\left. \left. + (\hat{m}_{G_j})^T (\hat{\Lambda}_{z^i})^{-1} \hat{m}_{G_j} + \text{Tr}((\hat{\Lambda}_{G_j})^{-1} (\hat{\Lambda}_{z^i})^{-1}) \right\} \right). \quad (21)$$

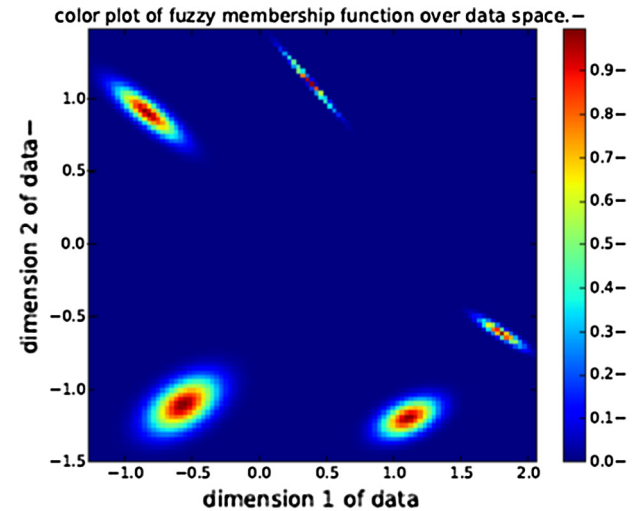
4. An Algorithm for multivariate data modeling

4.1. Algorithm

The analytical solution to mixture of uncertain signal models, derived in section (3), lends itself to Algorithm 1 for the modeling of multivariate data samples by determining membership functions on all of the variables and parameters. Algorithm 1 suggests to choose initial values of parameters based on k-means clustering and eigenvalue decomposition of sample covariance matrix.



(a) 2-dimensional data samples



(b) $\mu(y; \mathcal{M})$ (Algorithm 1 resulted in $C_{opt} = 5$.)

Figure 3 An example of the model learned from 2-dimensional data samples using Algorithm 1 (with $\beta = 0.5$).

Algorithm 1 Multivariate data modeling algorithm

Require: Data samples $\{y^n\}_{n=1}^N$.

- 1: Choose C_{min} (the minimum value of C), C_{max} (the maximum value of C), K_{min} (the minimum value of K), and K_{max} (the maximum value of K) as

$$K_{min} = 1, K_{max} = P, C_{min} = 1, C_{max} = \text{round}(N^\beta)$$

where $\beta \in [0, 0.5]$ is a constant controlling the maximum possible number of signal models in the mixture and $\text{round}(\cdot)$ rounds the argument to nearest integer. Choose,

$$a_\phi = \frac{N}{2}, b_\phi = 1, a_{\lambda_y} = \frac{N}{2}, b_{\lambda_y} = 1, a_{r_y} = \frac{N}{2}, b_{r_y} = 1, a_{s_y} = b_{s_y} = 1.$$

- 2: **for** $C = C_{min}$ to C_{max} **do**

- 3: Use k -means clustering to partition $\{y^n\}_{n=1}^N$ into C different clusters for choosing m_\circ^i equal to the centroid of i -th cluster and Λ_\circ^i equal to the inverse covariance of the data samples belonging to i -th cluster. The scalar $\hat{\pi}_i^n$ is initialized as equal to 1, if y^n belongs to i -th cluster, equal to 0, otherwise.

- 4: The vector \hat{m}_{m^i} is initialized as equal to m_\circ^i and matrix $\hat{\Lambda}_{m^i}$ is initialized as equal to Λ_\circ^i .

- 5: **for** $K = K_{min}$ to K_{max} **do**

- 6: Let $\{e_k\}_{k=1}^K$ denote the K largest eigenvalues of sample covariance of $\{y^n\}_{n=1}^N$ and $\{V_k\}_{k=1}^K$ (where $V_k = [V_{1k} \cdots V_{Pk}]^T \in \mathbb{R}^P$) denote the corresponding eigenvectors. The vector \hat{m}_{G_j} is initialized as

$$\hat{m}_{G_j} = [\sqrt{e_1}V_{j1} \cdots \sqrt{e_K}V_{jK}]^T.$$

The matrix $\hat{\Lambda}_{G_j}$ is initialized as equal to identity matrix of size K .

- 7: The parameters $(\hat{a}_{\lambda_y^i}, \hat{b}_{\lambda_y^i}, \hat{a}_{z_{y_j}^i}, \hat{b}_{z_{y_j}^i}, \hat{a}_{\phi_k}, \hat{b}_{\phi_k}, \hat{a}_{r_y}, \hat{b}_{r_y}, \hat{a}_{s_y}, \hat{b}_{s_y})$ are all initialized as equal to 1.

- 8: Initialize $\{(\hat{m}_{\alpha^i}, \hat{\Lambda}_{\alpha^i})\}_{i=1}^C$ by using (3-4).

9: Re-initialize $\{(\hat{m}_{G_j}, \hat{\Lambda}_{G_j})\}_{j=1}^P$ by using (5-6).

- 10: Re-initialize $\{(\hat{a}_{\phi_k}, \hat{b}_{\phi_k})\}_{k=1}^K$ using (7-8).

- 11: Re-initialize $\{(\hat{m}_{m^i}, \hat{\Lambda}_{m^i})\}_{i=1}^C$ using (9-10).

- 12: Re-initialize $\{(\hat{a}_{\lambda_y^i}, \hat{b}_{\lambda_y^i})\}_{i=1}^C$ using (11-12).

- 13: Re-initialize $\{(\hat{a}_{z_{y_j}^i}, \hat{b}_{z_{y_j}^i})\}_{j=1}^P\}_{i=1}^C$ using (13-14).

- 14: Re-initialize $\{(\hat{\pi}_i^n)\}_{n=1}^N\}_{i=1}^C$ using (15-16).

- 15: Re-initialize $(\hat{a}_{r_y}, \hat{b}_{r_y})$ using (17-18).

- 16: Re-initialize $(\hat{a}_{s_y}, \hat{b}_{s_y})$ using (19-20).

- 17: Update all parameters finally by running an iteration of equations (3-20).

- 18: Compute average fuzzy membership value of data samples as

$$\text{avg_memb_val}(C, K) = \frac{1}{N} \sum_{n=1}^N \mu^*(y^n)$$

where $\mu^*(y)$ is given by (21).

- 19: **end for**

- 20: **end for**

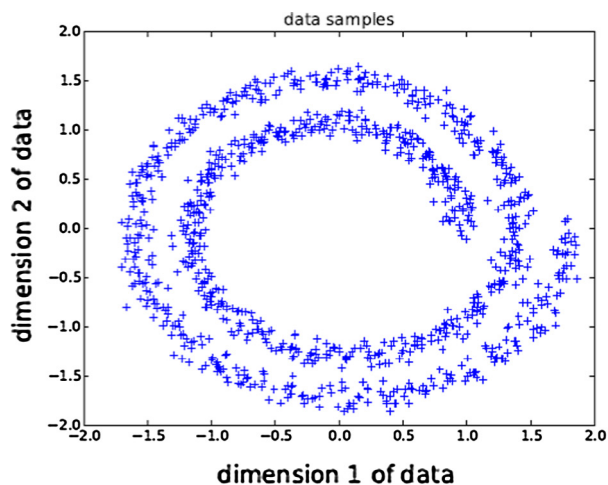
- 21: The values of C and K , maximizing avg_memb_val , are the optimal ones. That is,

$$(C_{opt}, K_{opt}) = \arg \max_{(C, K)} \text{avg_memb_val}(C, K).$$

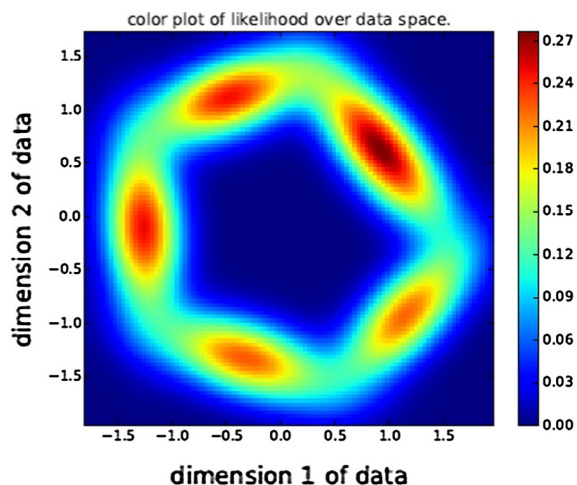
- 22: **return** $\{(\hat{m}_{m^i}, \hat{\Lambda}_{m^i})\}_{i=1}^{C_{opt}}$ corresponding to the case of $K = K_{opt}$.
-

Remark 1 (Complexity and Iterations) Algorithm 1 is based on the invoking of parameters updating rules (3–20). The time complexity of the algorithm, as a result of computing the inverse of a $P \times P$ sized matrix in update rule (10), is $O(P^3)$. Algorithm 1, after initializing the parameters, invokes a single

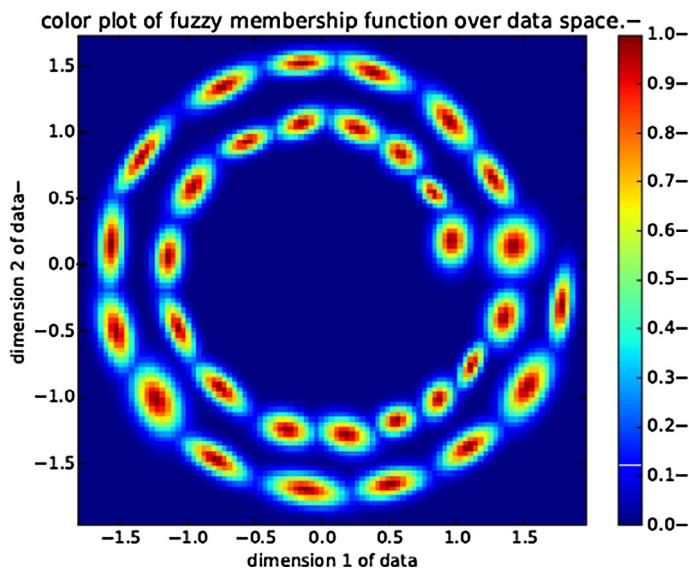
iteration of parameters updating rules. Thanks to the analytically derived solution due to which a single iteration is sufficient for parameters to nearly converge after initializing the parameters carefully. However, the optimal values of C and K are determined by maximizing the average fuzzy member-



(a) data samples



(b) GMM (BIC selected a model with 5 components out of all models with number of components varying from 1 to 32.)



(c) $\mu(y; \mathcal{M})$ (Algorithm 1, $C_{opt} = 32$.)

Figure 4 An example of the comparison between the Gaussian mixture models and Algorithm 1 (with $\beta = 0.5$).

ship value of the data samples through repeated application of update rules.

Remark 2 (Free parameter β in Algorithm 1) Algorithm 1 has only single free parameter, $\beta \in [0, 0.5]$, to be chosen by the user. The maximum possible number of signal models in the mixture, C_{max} , depends on the value of β . It will be demonstrated through experiments that algorithm's performance is not highly sensitive to the choice of β .

4.2. Data distribution modeling

The application of Algorithm 1 on given data samples $\{y^n\}_{n=1}^N$ results in the determination of C_{opt} different fuzzy member-

ship functions on unobserved variable m which (membership functions) are defined as

$$\mu^i(m; \hat{m}_{m^i}, \hat{\Lambda}_{m^i}) = \exp\left(-\frac{1}{2}(m - \hat{m}_{m^i})^T \hat{\Lambda}_{m^i} (m - \hat{m}_{m^i})\right),$$

$$\forall i \in \{1, \dots, C_{opt}\}.$$

Let \mathcal{M} be the set of parameters returned by Algorithm, i.e., $\mathcal{M} = \{(\hat{m}_{m^i}, \hat{\Lambda}_{m^i})\}_{i=1}^{C_{opt}}$. Finally, a data model, constructed from $\{y^n\}_{n=1}^N$ using Algorithm, is represented by a fuzzy membership function defined as

Table 1 A comparison of different classification algorithms with the proposed method in term of classification accuracy on testing data.

Method	Dataset 1	Dataset 2	Dataset 3
Nearest neighbors	100%	100%	75%
Linear SVM	91%	46%	51%
RBF SVM	90%	100%	59%
Decision tree	98%	100%	80%
Random forest	98%	100%	73%
AdaBoost	93%	97%	80%
Naive Bayes	92%	97%	57%
LDA	90%	29%	52%
QDA	90%	96%	57%
Analytical fuzzy ($\beta = 0.5$)	100%	100%	82%

$$\mu(y; \mathcal{M}) = \max_{1 \leq i \leq C_{opt}} \left\{ \exp \left(-\frac{1}{2} (y - \hat{m}_{m^i})^T \hat{\Lambda}_{m^i} (y - \hat{m}_{m^i}) \right) \right\}. \quad (22)$$

4.3. Classification

The data modeling capability of functional $\mu(m; \mathcal{M})$ can be exploited for the classification purpose. If $\mathcal{M}^1, \dots, \mathcal{M}^S$ are S different sets returned by Algorithm corresponding to the data samples of S different classes, then the class-label associated to a vector y could be predicted as

$$pred.label(y) = \arg \max_{1 \leq s \leq S} \mu(y; \mathcal{M}^s) \quad (23)$$

4.4. Demonstrations on Toy data sets

Fig.3 shows an example of the 2-dimensional data samples and a display of the fuzzy membership function $\mu(y; \mathcal{M})$ (calculated using (22)) over the data space. As depicted in Fig.3, the distribution of the samples $\{y^n\}_{n=1}^N$ in P-dimensional space is modeled by the fuzzy membership function $\mu(y; \mathcal{M})$. Stochastic mixture models have been extensively studied in the literature and are typically used to learn data distributions. The most commonly used Gaussian mixture models(GMM) fit

the given data samples by assuming that each data sample has been generated by a stochastic mixture of a finite number of the Gaussian distributions. ‘‘Expectation Maximization’’ algorithm is typically used for the learning of the Gaussian mixture models from data samples where the number of components in the mixture can be efficiently selected using the Bayesian information criterion (BIC). There may arise the situations when GMM don’t give favorable results. Fig. 4(a) is an example of data samples where better performance of Algorithm 1 than GMM (together with BIC) is observed. A comparison between color plots of GMM based likelihood (displayed in Fig.4(b)) and Algorithm 1 based fuzzy membership function (displayed in Fig. 4(c)) demonstrates the effectiveness of Algorithm 1 in modeling the distribution of data samples.

5. Heartbeat intervals classification

The section applies the proposed methodology on the experimentally recorded heartbeat intervals (referred to as the R-R intervals) of 20 different subjects while they were performing two different types of tasks in a chemical laboratory of Zhejiang University. One task involved manual pipetting of the chemical solutions while the other task involved working with the computer. The aim is to classify heartbeat intervals of a subject between the two tasks. The P-dimensional data samples were created from the sequence of R-R intervals as(see Table 1)

$$Y^i = [RR_{i-p+1} \dots RR_i]^T$$

where RR_i is i th heartbeat interval. The R-R intervals corresponding to the first half of the task duration serve as the training data and that of second half as testing data. Table 2 lists the median of classification accuracy over 20 subjects, obtained on testing data by different classification methods, for different values of data dimension P. The better classification accuracy of the analytical fuzzy approach in Table 2 supports the arguments that proposed approach could be an effective tool for modeling and analysis of biomedical data.

6. Concluding remarks

The theoretical contribution of this work is to propose an analytical fuzzy approach that provides a principled basis for

Table 2 A The median accuracy (in %) of different algorithms in classifying the testing heartbeat intervals between two tasks performed by subjects.

Method	Median of % accuracy ($P = 2$) % accuracy ($P = 2$)	Median of % accuracy ($P = 4$)	Median of % accuracy ($P = 6$)	Median of % accuracy ($P = 8$)
Nearest neighbors	87.11	90.33	91.08	92.65
Linear SVM	87.11	89.24	90.64	91.58
RBF SVM	84.07	84.17	86.99	90.11
Decision tree	84.95	87.22	88.83	89.57
Random forest	86.75	88.93	90.84	92.51
AdaBoost	88.36	90.72	91.87	92.60
Naive Bayes	87.40	89.27	91.05	92.18
LDA	88.67	90.70	91.59	92.99
QDA	88.04	88.46	90.08	90.97
Analytical fuzzy ($\beta = 0$)	88.75	91.16	92.14	93.14

determining the fuzzy membership functions to handle uncertainties in a modeling problem. The theoretical results form the basis for designing an algorithm that results in an efficient modeling of the data distribution in multi-parametric space. The analytically derived expressions for fuzzy membership functions for representing uncertainties associated with biomedical data should facilitate a system theoretic approach to mathematically design the medical expert systems.

References

- Alcala, R., Ducange, P., Herrera, F., Lazzarini, B., Marcelloni, F., 2009. A multiobjective evolutionary approach to concurrently learn rule and data bases of linguistic fuzzy-rule- based systems. *IEEE Trans. Fuzzy Syst.* 17 (5), 1106–1122.
- Aliasghary, M., Arghavani, N., 2012. H_∞ estimation for optimization of rational-powered membership functions. In: 2012 IEEE 13th International Symposium on Computational Intelligence and Informatics (CINTI), pp. 251–256.
- Antonelli, M., Ducange, P., Marcelloni, F., 2012. Genetic training instance selection in multiobjective evolutionary fuzzy systems: a coevolutionary approach. *IEEE Trans. Fuzzy Syst.* 20 (2), 276–290.
- Au, W.H., Chan, K., Wong, A.K., 2006. A fuzzy approach to partitioning continuous attributes for classification. *IEEE Trans. Knowl. Data Eng.* 18 (5), 715–719.
- Celikyilmaz, A., Turksen, I., 2008. Enhanced fuzzy system models with improved fuzzy clustering algorithm. *IEEE Trans. Fuzzy Syst.* 16 (3), 779–794.
- Chen, L., Chen, C., 2007. Pre-shaped fuzzy c-means algorithm (pfcm) for transparent membership function generation. In: IEEE International Conference on Systems, Man and Cybernetics. ISIC, pp. 789–794.
- Cococcioni, M., Lazzarini, B., Marcelloni, F., 2011. On reducing computational overhead in multi-objective genetic takagi-sugeno fuzzy systems. *Appl. Soft Comput.* 11 (1), 675–688.
- Fan, C.Y., Chang, P.C., Lin, J.J., Hsieh, J., 2011. A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification. *Appl. Soft Comput.* 11 (1), 632–644.
- Gacto, M., Alcala, R., Herrera, F., 2010. Integration of an index to preserve the semantic interpretability in the multiobjective evolutionary rule selection and tuning of linguistic fuzzy systems. *IEEE Trans. Fuzzy Syst.* 18 (3), 515–531.
- Gadaras, I., Mikhailov, L., 2009. An interpretable fuzzy rule-based classification methodology for medical diagnosis. *Artif. Intell. Med.* 47 (1), 25–41.
- Kumar, M., Stoll, R., Stoll, N., 2006. Deterministic approach to robust adaptive learning of fuzzy models. *IEEE Trans. Syst. Man Cybern. B Cybern.* 36 (4), 767–780.
- Kumar, M., Stoll, N., Kaber, D., Thurow, K., Stoll, R., 2007a. Fuzzy filtering for an intelligent interpretation of medical data. In: Proc. IEEE International Conference on Automation Science and Engineering (CASE 2007), Scottsdale, Arizona USA, pp. 225–230.
- Kumar, M., Weippert, M., Vilbrandt, R., Kreuzfeld, S., Stoll, R., 2007b. Fuzzy evaluation of heart rate signals for mental stress assessment. *IEEE Trans. Fuzzy Syst.* 15 (5), 791–808.
- Kumar, M., Arndt, D., Kreuzfeld, S., Thurow, K., Stoll, N., Stoll, R., 2008. Fuzzy techniques for subjective workload score modelling under uncertainties. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* 38 (6), 1449–1464.
- Kumar, M., Stoll, N., Stoll, R., 2009a. Adaptive fuzzy filtering in a deterministic setting. *IEEE Trans. Fuzzy Syst.* 17 (4), 763–776.
- Kumar, M., Stoll, N., Stoll, R., 2009b. On the estimation of parameters of takagi-sugeno fuzzy filters. *IEEE Trans. Fuzzy Syst.* 17 (1), 150–166.
- Kumar, M., Weippert, M., Arndt, D., Kreuzfeld, S., Thurow, K., Stoll, N., Stoll, R., 2010a. Fuzzy filtering for physiological signal analysis. *IEEE Trans. Fuzzy Syst.* 18 (1), 208–216.
- Kumar, M., Weippert, M., Stoll, N., Stoll, R., 2010b. A mixture of fuzzy filters applied to the analysis of heartbeat intervals. *Fuzzy Optim. Decis. Making* 9 (4), 383–412.
- Kumar, M., Neubert, S., Behrendt, S., Rieger, A., Weippert, M., Stoll, N., Thurow, K., Stoll, R., 2012a. Stress monitoring based on stochastic fuzzy analysis of heartbeat intervals. *IEEE Trans. Fuzzy Syst.* 20 (4), 746–759.
- Kumar, M., Stoll, N., Thurow, K., Stoll, R., 2012b. Physiological signals to individual assessment for application in wireless health systems. In: Proc. 9th International Multi-Conference on Systems, Signals and Devices (SSD), pp. 1–6.
- Kumar, M., Stoll, N., Stoll, R., Thurow, K., 2016a. A stochastic framework for robust fuzzy filtering and analysis of signals-part i. *IEEE Trans. Cybern.* 46 (5), 1118–1131.
- Kumar, M., Stoll, N., Stoll, R., Thurow, K., 2016b. Variational optimization of fuzzy membership functions. *Artif. Intell. Under-Rev.*
- Liao, T.W., Celmins, A.K., Hammell, R.J., 2003. A fuzzy c-means variant for the generation of fuzzy term sets. *Fuzzy Sets Syst.* 135 (2), 241–257.
- Makrehchi, M., Basir, O., Kamel, M., 2003. Generation of fuzzy membership function using information theory measures and genetic algorithm. In: Bilgiç, T., De Baets, B., Kaynak, O. (Eds.). In: *Fuzzy Sets and Systems – IFSA 2003, Lecture Notes in Computer Science*, vol. 2715. Springer, Berlin Heidelberg, pp. 603–610.
- Mottaghi-Kashtiban, M., Khoei, A., Hadidi, K., 2008. Optimization of rational-powered membership functions using extended kalman filter. *Fuzzy Sets Syst.* 159 (23), 3232–3244.
- Nguyen, T., Khosravi, A., Creighton, D., Nahavandi, S., 2015. Medical data classification using interval type-2 fuzzy logic system and wavelets. *Appl. Soft Comput.* 30, 812–822.
- Oh, S.K., Pedrycz, W., Park, H.S., 2003. Hybrid identification in fuzzy-neural networks. *Fuzzy Sets Syst.* 138 (2), 399–426.
- Papageorgiou, E.I., 2011. A new methodology for decisions in medical informatics using fuzzy cognitive maps based on fuzzy rule-extraction techniques. *Appl. Soft Comput.* 11 (1), 500–513.
- Pulkkinen, P., Koivisto, H., 2010. A dynamically constrained multi-objective genetic fuzzy system for regression problems. *IEEE Trans. Fuzzy Syst.* 18 (1), 161–177.
- Robles, I., Alcalá, R., Benítez, J.M., Herrera, F., 2009. Evolutionary parallel and gradually distributed lateral tuning of fuzzy rule-based systems. *Evol. Intell.* 2 (1–2), 5–19.
- Seera, M., Lim, C.P., 2014. A hybrid intelligent system for medical data classification. *Expert Syst. Appl.* 41 (5), 2239–2249.
- Simon, D., 2005. H_∞ estimation for fuzzy membership function optimization. *Int. J. Approximate Reasoning* 40 (3), 224–242.