This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2017.2688342, IEEE Transactions on Fuzzy Systems

Analysis of Data Generated from Multidimensional Type-1 and Type-2 Fuzzy Membership Functions

Desh Raj, Aditya Gupta, Bhuvnesh Garg, Kenil Tanna, and Frank Chung-Hoon Rhee, *Member, IEEE*

***Abstract*—Due to the numerous applications that utilize different types of fuzzy membership functions (MFs), it may sometimes be difficult to choose an appropriate MF for a particular application. In this paper, we establish preliminary guidelines to direct this selection by proposing a three-stage method. In the "forward" stage, different MFs such as crisp MFs, type-1 (T1) fuzzy MFs, and type-2 (T2) fuzzy MFs are generated from multidimensional data sets. Next, in the "reverse" stage, data is generated back from these MFs by considering different bin sizes. In doing so, various data sets may be generated for different applications which require fuzzy data. Finally, for the "similarity analysis" stage, we propose an iterative algorithm that makes use of the results of Wilcoxon Signed Rank (WSR) and Wilcoxon Rank Sum (WRS) tests to compare the original data and the generated data. From the results of these tests, recommendations concerning the suitability of MFs for a specific application may be suggested by observing the accuracy of representation and the requirements of the application. With this analysis, the objective is to gain insight on when T2 fuzzy sets may be considered to outperform T1 fuzzy sets, and vice versa. Several examples are provided using synthetic and real data to validate the iterative algorithm for data sets in various dimensions.**

***Index Terms*—Fuzzy sets, membership functions, type-1 fuzzy, type-2 fuzzy, Wilcoxon Signed Rank, Wilcoxon Rank Sum.**

I. Introduction

FUZZY sets were introduced as a method for imitating the human intuition of classifying objects into groups based on a fixed set of rules, by providing a smooth membership function (MF) to represent the output using interpolation processes [1]. Once fuzzy logic systems were identified as rigorous representatives of set memberships, fuzzy rule-based systems were developed using type-1 (T1) fuzzy MFs [2], and further extended to type-2 (T2) [3]. T2 fuzzy MFs assign uncertainty to the T1 fuzzy MFs as secondary memberships [4], [5], which may be constant-valued, such as in interval type-2 (IT2) fuzzy MFs [6], or parametrized functions, such as in general type-2 (GT2) fuzzy MFs. These parametrized functions themselves may take on different shapes, thus generating a multitude of fuzzy MFs.

D. Raj and K. Tanna are with the Department of Computer Science and Engineering, Indian Institute of Technology Guwahati (IITG), India.
(email: {r.desh, kenil}@iitg.ernet.in)
A. Gupta and B. Garg are with the Department of Mathematics, Indian Institute of Technology Guwahati (IITG), India.
(email: {aditya.gupta, g.umang}@iitg.ernet.in)
F. C.-H. Rhee is with Hanyang University, Korea.
(email: frhee@fuzzy.hanyang.ac.kr)
The project was conceptualized and completed at the Computational Vision and Fuzzy Systems Lab, Hanyang University.

Applications of fuzzy logic systems have been extended to fields such as classification of coded video streams [7], control of mobile robots [8], equalization of nonlinear fading channels [9], interval-representation of sets [10], forecasting of time-series [11], diagnosis of diseases [12], pre-processing radiographic images [13], medical applications [14], and numerous others. Interval-valued sets, in particular, have been used to improve various clustering algorithms such as fuzzy C-means (FCM) [15] and kernel possibilistic C-means (KPCM) [16].

With the availability of different types of MFs, there may be a predicament about which MF would best represent a data set for a particular application. Recently, this issue has been frequently mentioned [17]–[19] due to increasing focus on T2 fuzzy sets. In John and Coupland's [17] article on the challenges and misconceptions about the fuzzy model, the authors argue that while the present T2 fuzzy logic systems (FLSs) are employed as extensions to T1 FLSs, there may be future research to identify areas, such as linguistics, where high levels of uncertainty necessitate the use of T2 FLSs. Further, studies to reduce computational complexity for T2 FLSs may prove important for shifting fuzzy models from T1 to T2 or higher levels of fuzziness. However, as described by Mendel [18] and Wu [19], T2 FLSs do not always outperform T1 FLSs, and it may be advantageous to identify applications where T1 FLSs suffice, given the uncertainty involved.

For this reason, it may be necessary to outline a general method for determining the suitability of a MF to represent a data set, in accordance with the requirements of the application. For example, pathological applications of fuzzy sets may require high precision irrespective of an increase in computational complexity. On the other hand, robot navigation applications may demand a more flexible representation of trained data to avoid overfitting.

To make this analysis possible, we propose a three-phase procedure as shown in Fig. 1. In the first phase known as "forward stage," different MFs, namely crisp MF, T1 fuzzy MF, and T2 fuzzy MF, are obtained from initial data referred to as the "original" data set. The second phase, called "reverse stage" involves generating data back from the MFs obtained in the previous stage. In the final "similarity analysis" stage, these data sets are compared with the original data to determine maximum similarity. For this analysis, we propose an iterative algorithm which employs the Wilcoxon Signed Rank (WSR) and Wilcoxon Rank Sum (WRS) tests on non-parametric data, and returns the smallest bin size (i.e., highest precision level) for which two data sets may be considered similar.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2017.2688342, IEEE Transactions on Fuzzy Systems

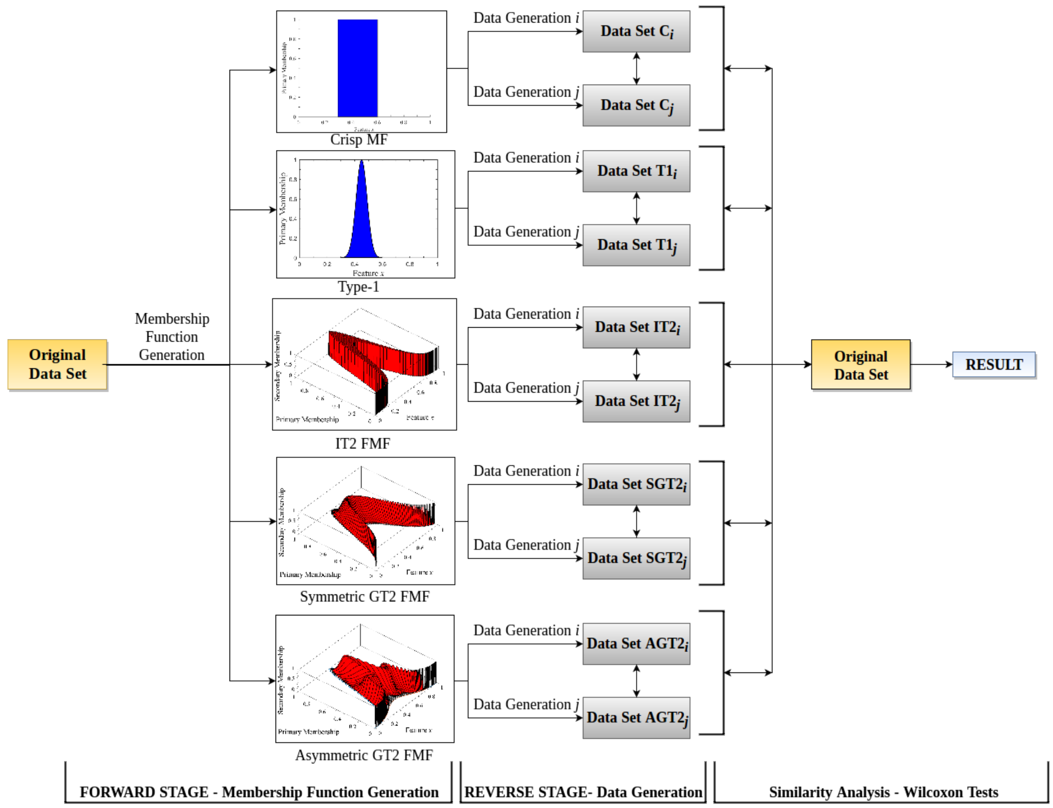


Fig. 1: Flow diagram showing different stages of the data analysis.

At this point, it may be necessary to examine the need for choosing a fuzzy model over a probabilistic model for dealing with ambiguous data sets. In the discussion on the relationship between fuzzy logic and probability theory [20], Zadeh enumerates several situations where a probabilistic model may fall short of providing a comprehensive understanding of ambiguity. Some of these include an oversight of linguistic terms for fuzziness such as *several*, *few*, and *likely*, and failure to quantify a deterministic fuzzy event. The latter is also sketched out effectively in the famous "swamp water example" proposed by Bezdek [21]. There are various comprehensive discussions [22] regarding the advantages of fuzzy logic over probability theory and vice-versa, and it may be cited as consensus that fuzzy models are more useful in situations where the type of ambiguity, i.e., uncertainty or randomness, is inherently unknown. Although we have illustrated data samples that are randomly distributed, for all practical purposes, the randomness may be considered to be a substitute for fuzziness. This is because probabilistic distributions may be modeled as fuzzy sets, but the opposite is not always true.

Our major contributions in this paper may be listed as follows.

- We propose a novel iterative method to perform similarity analysis between multidimensional fuzzy sets.
- We suggest some preliminary guidelines regarding the choice of MFs by performing experiments on synthetic and real data sets.
- We attempt to establish a framework under which future research on the subject may be performed.

The remainder of this paper is organized as follows. Section II provides an overview of related studies on similarity analysis of fuzzy MFs. In Section III, we briefly discuss the theory of fuzzy sets. It is followed by an extensive description of our proposed method, i.e., the forward, reverse, and similarity analysis stages, in Section IV. Section V illustrates the algorithm on several data sets. We conclude by suggesting guidelines for the choice of MFs which may suit the requirements of the application concerned.

II. Related Studies

There has been extensive study on the similarity analysis of fuzzy sets [23]. Among the early works, Krumhansl [24] explored the interrelation between spatial distance and similarity by outlining a distance-density geometric model. The model was compared to the set-theoretic approach [25] in which objects were represented as collections of features and similarity was described as a feature-matching process.

The set-theoretic approach was explored further in a fuzzy context [26], where the authors argued that the popularity of case-based reasoning [27] in artifical intelligence had put forth the importance of similarity in reasoning. Distance metrics were again proposed as similarity measures, with applications such as clustering and the definition of fuzzy numbers. Murthy et al. [28] elucidated the need for quanitfying the "correlation" between fuzzy MFs, and defined the correlation coefficient in terms of the normalized covariance between data samples. Bonissone [29] represented fuzzy sets as a lingustic model

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2017.2688342, IEEE Transactions on Fuzzy Systems

and used semantic similarity as the measure. In [30], different orders of polynomials for fitting MFs used in fuzzy controllers were evaluated.

Although these techniques are convenient for determining the measure of similarity, recommendations on the use of particular fuzzy sets in applications based on such a similarity analysis are not provided.

III. Background

Fuzzy sets can be completely characterized by their MFs [2], which represent the degree of certainty (or uncertainty) as an extension of valuation. MFs were originally defined as a quantitative estimation of fuzziness as follows [1].

If we consider $\mathbf{X}$ to be a set of elements such that any generic element of $\mathbf{X}$ is denoted by x, then $\mathbf{X} = \{x\}$. Any fuzzy set $\mathbf{A}$ of $\mathbf{X}$ is then characterized by a MF $\mu_{\mathbf{A}}(x)$ which associates a real number in the interval [0,1] with each element in $\mathbf{X}$, such that the values of $\mu_{\mathbf{A}}(x)$ represent the grade of membership of x in $\mathbf{A}$. At its extremities, a value of 0 denotes no membership and that of 1 denotes full membership. In general, a higher value of $\mu_{\mathbf{A}}(x)$ is associated with higher grade of membership of x in $\mathbf{A}$.

In this section we define crisp, T1 fuzzy, and T2 fuzzy sets in n dimensions. Equivalent definitions for one and two-dimensional sets may be obtained by substituting n with 1 and 2, respectively.

1) A crisp set $\mathbf{A}$ defined on a universe of discourse $\mathbf{X}^n$ can be defined by listing all the n-tuples $(x_1, x_2, \ldots, x_n)$ which belong to the set $\mathbf{A}$. This can be represented as

$$\mu_{\mathbf{A}}(x_1, x_2, \ldots, x_n) = \begin{cases} 1 & \text{if } (x_1, x_2, \ldots, x_n) \in \mathbf{A} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

2) A type-1 fuzzy set $\mathbf{A}$ defined on a universe of discourse $\mathbf{X}^n$ is characterized by a MF $\mu_{\mathbf{A}}(x_1, x_2, \ldots, x_n)$, i.e.,

$$\mathbf{A} = \{((x_1, x_2, \ldots, x_n), \mu_{\mathbf{A}}(x_1, x_2, \ldots, x_n)) \mid x_i \in \mathbf{X}\}. \quad (2)$$

3) A type-2 fuzzy set, denoted by $\widetilde{\mathbf{A}}$, is characterized by MF $\mu_{\widetilde{\mathbf{A}}}((x_1,\ldots,x_n),u)$, where $x_i \in \mathbf{X}$ and $u \in J_{(x_1,\ldots,x_n)} \subseteq$ [0,1], i.e.,

$$\widetilde{\mathbf{A}} = \{((x_1, \ldots, x_n), u), \mu_{\widetilde{\mathbf{A}}}((x_1, \ldots, x_n), u) \mid \forall x_i \in \mathbf{X} \; and \; u \in J_{(x_1,\ldots,x_n)} \subseteq [0,1]\}, \quad (3)$$

where $0 \le \mu_{\widetilde{\mathbf{A}}}((x_1, \ldots, x_n), u) \le 1$. T2 fuzzy MF $\widetilde{\mathbf{A}}$ can also be expressed as the integral of all primary membership values as described for 1-dimensional (1-D) data. Further, the primary membership for n dimensions may be any parametrized hyper-geometric functions, such as a hyper-pyramid or a hyper-Gaussian.

IV. Proposed Method

In this section, we describe our proposed method for selecting an appropriate MF according to the data set and application requirements. As represented in Fig. 1, the method comprises of three stages:

1) *Forward stage*: Obtaining crisp and fuzzy MFs from the "original" data set.
2) *Reverse stage*: Generation of data back from the obtained MFs.
3) *Similarity analysis*: Comparing the similarity between the original data and the generated data using an iterative algorithm based on Wilcoxon's tests.

We now discuss each of the above stages, and provide illustrations for one and two-dimensional data sets.

A. Forward stage

Data samples are quantified mathematically by a vector composed of the values of some measurable features. Depending on their relative impact on algorithmic performance (e.g., in case of pattern recognition), computational overhead, and mutual correlation, certain features may not be appropriate for representing the data. Consequently, it may be preferable to determine a general set of derivations to obtain the MFs from data sets in different dimensions. We now describe the method for MF generation from data in the "forward stage."

Crisp MF generation:

To generate a 1-D crisp MF, we consider an arbitrary feature which is represented by a labeled data set $\mathbf{X} = x_1, x_2, \ldots, x_n$ and $0 \le x_i \le p$ for some $p \in \mathbb{R}$. The entire range [0, p] can be divided into m bins (denoted by B_j, where $j \in \{1,\ldots,m\}$) of equal width such that each bin has width of $w = p/m$. Then, the MF for bin B_j can be represented as

$$\mu_{B_j}(x_i) = \begin{cases} 1 & \text{if } \frac{p}{m}(j-1) \le x_i \le \frac{p}{m} j \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

Fig. 2(a) shows a data set consisting of 25 samples and the corresponding crisp membership is represented in Fig. 2(b). The range [0,1] (i.e., p = 1) has been divided into m = 25 bins resulting in width w = 0.04.

Further, let us consider n samples consisting of two features, where sample $\mathbf{x}_i = \{(x_{i1}, x_{i2})\}$ s.t. $0 < x_{i1} \le p$ and $0 < x_{i2} \le q$ for some $p, q \in \mathbb{R}$. The range [0, p] and [0, q] on the x_1 and x_2 axes is then divided into bins of equal width w. There are $m_1 = p/w$ and $m_2 = q/w$ bins along the x_1 and x_2 axis, respectively, and consequently, there are pq/w^2 bins in total. Hence, the crisp MF for this data set can be defined by

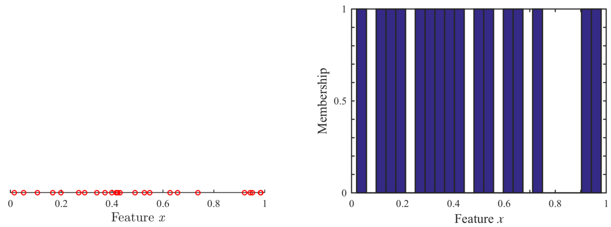
$$\mu_{\mathbf{A}}(x_{i1}, x_{i2}) = \begin{cases} 1 & \begin{array}{l} \text{if } \dfrac{p}{m_1}(i-1) \le x_{i1} \le \dfrac{p}{m_1} i \\ \text{and } \dfrac{q}{m_2}(i-1) \le x_{i2} \le \dfrac{q}{m_2} i \end{array} \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

As an illustration, a 2-D data set consisting of 100 samples and its corresponding crisp MF (assuming w = 0.1) are shown in Figs. 3(a) and 3(b), respectively. Using a similar approach, we can generate a crisp MF for n-dimensional data as well.

Type-1 fuzzy MF generation:

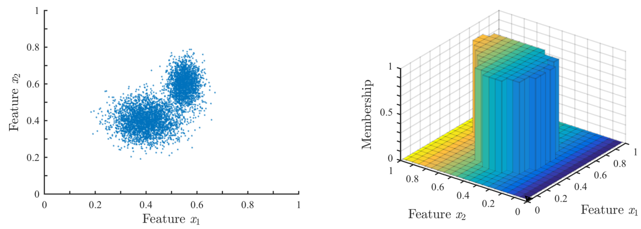
To construct the T1 fuzzy MF, the histogram of the given sample for each labeled class is first obtained, smoothed using a triangular window, and then normalized [31], [32]. Depending on the distribution of data, different types of MFs such triangles, trapezoids, Gaussian functions, and S-functions, can be generated [31], [33], [34]. To illustrate the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2017.2688342, IEEE Transactions on Fuzzy Systems



(a) (b)

Fig. 2: Illustration of a 1-D data set containing 25 samples: (a) scatter plot and (b) the corresponding crisp MF.



(a) (b)

Fig. 3: Illustration of a 2-D data set containing 100 samples: (a) scatter plot and (b) the corresponding crisp MF.

method, we first describe the generation of a Gaussian MF from 1-D data.

Suppose the smoothed histogram in Fig. 4(a) represents a sample 1-D data. Now, consider a multidimensional Gaussian function given by

$$G(\mathbf{x}) = a \cdot \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right), \quad (6)$$

where a is the height, $\boldsymbol{\mu}$ is mean vector, and $\sum$ is the covariance matrix. If the histogram has N significant peaks, we can model it as the sum of Gaussian functions as

$$\widetilde{G}(\mathbf{x}) = \Sigma_{i=1}^{N} G_i(\mathbf{x}). \quad (7)$$

To approximate the smoothed histogram as Gaussian functions,

$$J(\mathbf{p}) = \frac{1}{2}\left(\widetilde{G}(\mathbf{x}) - H(\mathbf{x})\right)^2 \quad (8)$$

can be considered as the objective function to be minimized, where parameter vector $\mathbf{p}_i = (a_i, \boldsymbol{\mu}_i, \sum_i)$ is for the Gaussian function $\widetilde{G}_i(\mathbf{x})$ and $H(\mathbf{x})$ is the smoothed histogram of the input data.

The objective function is used for fitting Gaussians to the polynomial-fitted histograms. This problem of converging to a local minima rather than the global minima is generic to any gradient descent approach. In our experience, to converge at the global minima with gradient descent, it is essential to heuristically estimate the initial parameters properly using one of several methods, such as those involving clustering, annealing, or a Monte Carlo approach [35], [36]. In the case of a Gaussian, we found that considering the histogram-fitted polynomial's significant maxima as the initial peaks of the Gaussians, and the polynomial's minima locations as the initial variance resulted in fast and better convergence with gradient descent. We apply a heuristic of our own to meet convergence. In general, any optimization techniques may be used for this purpose, regardless of the characteristics of the loss function. For instance, the parameters of a Gaussian function may be estimated using a number of known techniques such as maximum likelihood estimation (MLE), expectation maximization (EM), among others. The EM method is often employed for this purpose due to its simplicity and fast convergence. On obtaining the Gaussian densities, they can be further normalized for generating a valid fuzzy MF.

We now propose the following heuristic approach that consists of four steps to obtain the initial parameters [31].

Step 1: The smoothed histograms for the given sample data are generated.

Step 2: By least squares approximation, a polynomial function (PF) of the lowest possible degree (i.e., to avoid over fitting) is fitted such that each fit has a reasonably small error.

Step 3: The extrema (maxima and minima) values for the PF in Step 2 is calculated and the ratio of number of Gaussians to the number of positive valued maxima is determined, ignoring the ones that have small peaks.

Step 4: The heights of the Gaussians are initialized by the maxima values (peak values) and the mean values are initialized as the locations of these peaks. The standard deviation of each Gaussian is initialized as the shortest value among the distances between the mean of the Gaussian and the nearest minima or roots of the PF.

The gradient descent method can be used to estimate the parameter vector $\mathbf{p}_i$ which minimizes (8) by the update rule

$$\mathbf{p}_i^{new} = \mathbf{p}_i^{old} - \rho\frac{\partial J}{\partial \mathbf{p}_i}, \quad (9)$$

where ρ is a positive learning constant.

For Gaussian functions, the partial derivatives of J with respect to each component of $\mathbf{p}_i$ are

$$\frac{\partial J}{\partial a_i} = \left(\Sigma_{j=1}^{N} G_j(\mathbf{x}) - H(\mathbf{x})\right) \cdot \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right), \quad (10)$$

$$\frac{\partial J}{\partial \boldsymbol{\mu}_i} = \frac{1}{2}\left(\Sigma_{j=1}^{N} G_j(\mathbf{x}) - H(\mathbf{x})\right) \cdot G_i(\mathbf{x}) \cdot (\mathbf{x}-\boldsymbol{\mu}_i)^T \cdot \left(\boldsymbol{\Sigma}_i^{-T} + \boldsymbol{\Sigma}_i^{-1}\right), \quad (11)$$

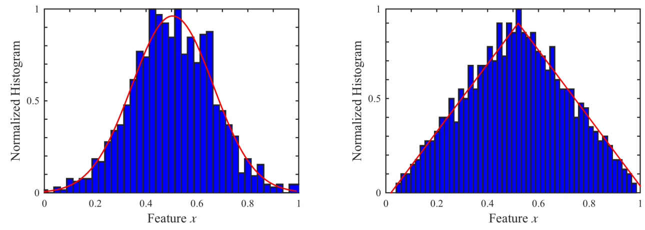
and

$$\frac{\partial J}{\partial \boldsymbol{\Sigma}_i} = \frac{1}{2}\left(\Sigma_{j=1}^{N} G_j(\mathbf{x}) - H(\mathbf{x})\right) \cdot G_i(\mathbf{x}) \cdot \left(\boldsymbol{\Sigma}_i^{-T}(\mathbf{x}-\boldsymbol{\mu}_i) \cdot (\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-T}\right). \quad (12)$$

The resulting MF obtained by 1-D Gaussian function fitting is shown in Fig. 4(a). Apart from using a Gaussian to represent a T1 fuzzy MF, a triangular function can also be used when the distribution of data is linearly increasing or decreasing. Fig. 4(b) illustrates the triangular MF fitted on the histograms obtained from another 1-D data set.

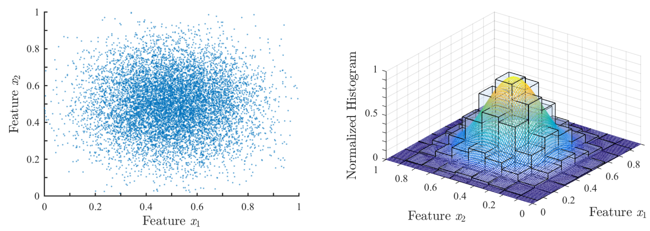
Further, let us consider Fig. 5(a) which shows the scatter plot of a sample 2-D data set. Using the same method

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2017.2688342, IEEE Transactions on Fuzzy Systems



(a) (b)

Fig. 4: Examples of MF fitting on data: (a) Gaussian fitting and (b) triangular fitting.



(a) (b)

Fig. 5: Illustration of a Gaussian 2-D data set: (a) scatter plot, (b) the corresponding smoothed histogram and Gaussian T1 fuzzy MF.

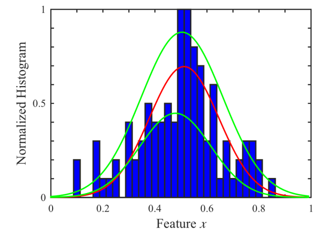


Fig. 6: Gaussian fitting on a histogram for obtaining the UMF and LMF.

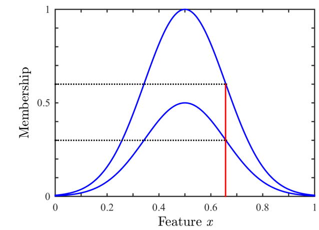
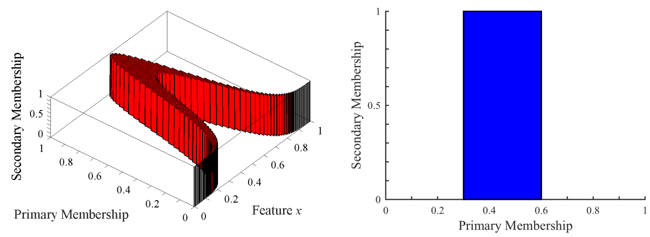


Fig. 7: *Vertical slice* on UMF and LMF of 1-D data.



(a) (b)

Fig. 8: Illustration of an IT2 fuzzy MF generated from a 1-D data set: (a) the fuzzy MF and (b) cross-section of the curve intersected by plane $\mathbf{P}$ at $x = x_0$.

described above, we fit a Gaussian surface over the smoothed histograms obtained from the data and the corresponding T1 fuzzy MF is given in Fig. 5(b). In general, for any n-dimensional data set, a hyper-Gaussian or an appropriate hypergeometric surface can be fitted over the data using a similar approach.

Type-2 fuzzy MF generation:

A *type*-2 fuzzy set [4], [37], denoted by $\widetilde{\mathbf{A}}$, models the uncertainty associated with the primary MF, and this uncertainty or fuzziness is also known as the secondary membership.

For a 1-D data set, a general type-2 (GT2) fuzzy MF is described as the axes being feature x, primary membership u, and secondary membership $\mu_{\widetilde{\mathbf{A}}}(x,u)$. The 2-D support of $\mu_{\widetilde{\mathbf{A}}}(x,u)$ (i.e., the area between upper membership function (UMF) and lower membership function (LMF)) is called the *footprint of uncertainty* (FOU) of $\widetilde{\mathbf{A}}$, i.e.,

$$\text{FOU}(\widetilde{\mathbf{A}}) = \left\{(x,u) \in \mathbf{X} \times [0,1] \,|\, \mu_{\widetilde{\mathbf{A}}}(x,u) > 0\right\}, \qquad (13)$$

i.e., the FOU consists of the region comprising points (x,u) for which the secondary membership takes some positive value.

If the primary membership at all samples is a continuous, closed interval, and the secondary membership assumes a uniformly constant value, i.e., $\mu_{\widetilde{\mathbf{A}}}(x,u) = 1$, through the entire FOU, it is called an interval type-2 (IT2) fuzzy MF [38].

We will now describe the generation of three T2 fuzzy MFs, namely IT2, GT2 with symmetric Gaussians as secondary memberships (SGT2), and GT2 with asymmetric Gaussians as secondary memberships (AGT2) [39]. The corresponding primary membership for each of these is assumed to be either a symmetric or asymmetric Gaussian distribution, so the UMF and LMF of the FOU are obtained by again fitting Gaussian functions to the smoothed histogram values that are above and below the fitted GF obtained previously. Fig. 6 illustrates how these Gaussians are fitted to obtain the FOU.

We consider a *vertical slice* [4], [40] (shown in Fig. 7) at $x = x_0$, intersecting the UMF and LMF at $\bar{u}(x_0)$ and $\underline{u}(x_0)$, respectively. The secondary membership can be expressed as

$$\mu_{\widetilde{\mathbf{A}}}(x) = \begin{cases} 1 & \forall x \in \mathbf{X}, \forall u \in J_x \subseteq [0,1] \\ 0 & \text{otherwise.} \end{cases} \qquad (14)$$

This function is represented in Fig. 8(a). If we draw the plane $\mathbf{P}$ at $x = x_0$, it will intersect the $x-u$ plane along the dotted vertical slice in Fig. 7. The cross-section of the graph cut along the plane $\mathbf{P}$ is shown in Fig. 8(b).

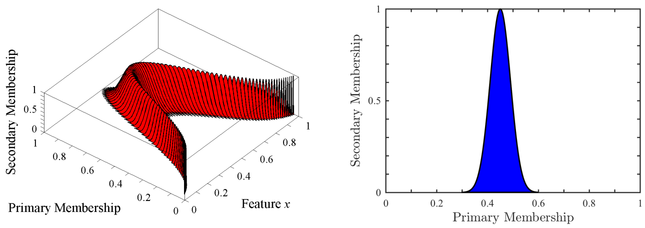
For an SGT2 fuzzy MF, the cross-section is a uniform Gaussian distribution. Corresponding to x_0, the position of the mean $u = b_0$ and standard deviation c_0 (note, $3c_0$ is the 99% confidence interval for Gaussian distribution) is defined, either discretely or by a parametrized function of x. For this case, the T1 fuzzy MF lies between the UMF and LMF and can be considered as the mean. The Gaussian secondary membership for the SGT2 fuzzy MF is given by

$$\mu_{\widetilde{\mathbf{A}}}(x,u) = a \cdot \exp\left(-\frac{(u-b_0)^2}{2c_0^2}\right), \qquad (15)$$

where $a = 1$, $b_0 = \mu_{\mathbf{A}}(x)$ (primary membership), and $c_0 = (\bar{u}(x_0) - \underline{u}(x_0))/3$. This is graphically represented in Fig. 9(a), and the cross-section of the graph intersected by the plane $\mathbf{P}$ at $x = x_0$ is shown in Fig. 9(b).

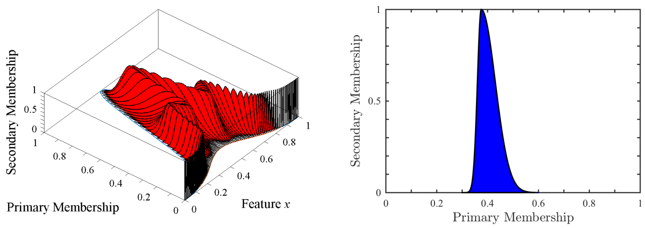
An AGT2 fuzzy MF for a 1-D data set can be visualized by "snaky" surfaces. To obtain this, multiple histograms are generated by varying the bin sizes and Gaussians are fitted on each of these such that each lies inside the FOU. On taking a vertical slice at $x = x_0$, a Gaussian is obtained as the secondary membership. Since the mean of the distribution can

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2017.2688342, IEEE Transactions on Fuzzy Systems



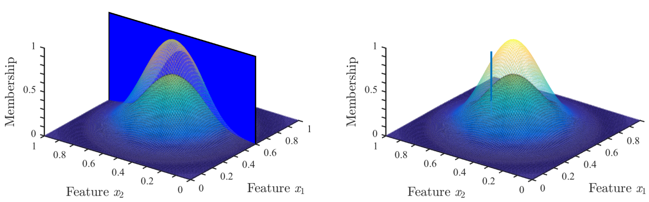
(a) (b)

Fig. 9: Illustration of a SGT2 fuzzy MF generated from a 1-D data set: (a) the fuzzy MF and (b) cross-section of the curve intersected by plane **P** at $x = x_0$.



(a) (b)

Fig. 10: Illustration of a AGT2 "snaky" fuzzy MF generated from a 1-D data set: (a) the fuzzy MF and (b) cross-section of the curve intersected by plane **P** at $x = x_0$.



(a) (b)

Fig. 11: Representations of secondary membership for 2-D data: (a) vertical slice and (b) primary membership drill.

be located anywhere in the FOU, different values of variance needs to be taken on either side of the mean. For mean $u = b_0$ ($\underline{u}(x_0) \leq b_0 \leq \bar{u}(x_0)$), the upper and lower variances are given as

$$c_1 = \frac{\overline{u}\,(x_0) - b_0}{6} \text{ and } c_2 = \frac{b_0 - \underline{u}\,(x_0)}{6}, \tag{16}$$

respectively. Hence the secondary membership becomes

$$\mu_{\tilde{\mathbf{A}}}\,(x,u) = \begin{cases} \exp\left(-\frac{(u-b_0)^2}{2c_1^2}\right) & u \leq b_0 \\ \exp\left(-\frac{(u-b_0)^2}{2c_2^2}\right) & u > b_0. \end{cases} \tag{17}$$

This is graphically represented by the "snaky" surface plot in Fig. 10(a), and the cross-section of the graph intersected by the plane $x = x_0$ is shown in Fig. 10(b).

For 2-D data sets, since visually representing the secondary membership for the entire region inside the FOU is not possible, we may consider a *vertical slice*, for example, at $x_1 = x_0$ (see Fig. 11(a)), which reduces the problem into to a 1-D data set by fixing one of the dimensions to a constant value. However, considering a vertical slice may present two complications for this case:

1) A *vertical slice* provides membership information for all samples lying on the intersection of the slice with the 2-D x_1-x_2 plane, which may not be required.
2) Infinite planes (or slices) may be drawn passing through the point $x_1 = x_0$, and only one of these planes, the one parallel to the x_2 axis, makes the problem analogous to 1-D data.

For this reason, to represent the secondary memberships appropriately, it is essential to represent it at a particular point, and this necessitates the use of a *vertical drill* over a vertical slice, as shown in Fig. 11(b). Furthermore, for data sets in higher dimensions, the concept of verticality becomes meaningless, and for such cases, the vertical drill may be renamed as "primary membership drill." The same approach may be generalized to an n-dimensional data set as well.

B. Reverse stage

We now describe the method for data generation from MFs in the "reverse stage." For sake of brevity, we limit our discussion to 2-D data sets, and argue that the same method can be extended to higher dimensions.

Data generation from crisp MFs:

Let us consider a crisp MF similar to Fig. 3(b). Suppose the entire data set is distributed over a square region of area M^2, where $M \in \mathbb{R}$. The entire square region is then divided into a grid of size $n \times n$ ($n \in \mathbb{N}$), such that each cell is a square of side $a = M/n$. Inserting a vertical drill through the center of each of the n^2 cells of the grid returns a binary matrix of order $n \times n$.

Suppose the maximum number of points inside each cell in the grid is taken as N_0. To generate data, we scatter N_0 points randomly in the cells with a vertical drill value of 1, and those with a value of 0 receive no points in the data. When this process is repeated for each cell, the x-coordinates (y-coordinates) of all the samples in the scatter plot are grouped and labeled as Feature x_1 (x_2).

In theory, infinite number of points can be scattered in any region. However, in practice, while generating scatter plots in a finite area, there may exist overlapping points. We now determine the probability of obtaining repeated data in a labeled set of features while generating the data set. Suppose the coordinates of generated points have a precision of p decimal points, so that there can be 10^p distinct points in a range of 1 unit. Since the length of side of each square cell is $a = M/n$, the maximum possible number of distinct data points in a grid cell is $10^p M/n$.

From probability theory, we can distribute $\binom{10^p \frac{M}{n}}{N_0}$ distinct points in the region, as opposed to $\binom{10^p \frac{M}{n}+N_0-1}{10^p \frac{M}{n}-1}$ number of ways without such a constraint of uniqueness. From these values, we can obtain the probability for uniqueness of all data points in one feature as $\binom{10^p \frac{M}{n}}{N_0}/\binom{10^p \frac{M}{n}+N_0-1}{10^p \frac{M}{n}-1}$. Hence, the probability that there is at least one redundant data pair in either of the two features can be given as $1-\left(\binom{10^p \frac{M}{n}}{N_0}/\binom{10^p \frac{M}{n}+N_0-1}{10^p \frac{M}{n}-1}\right)^2$.

As an illustration, if p = 3, M = 100, n = 10, and N_0 = 10, this probability can be calculated as 0.018, which is infinitesimally small. Hence, this data can be considered sufficiently distinct.

Data generation from T1 fuzzy MFs:

Let us consider a T1 fuzzy MF in the form of a Gaussian distribution as shown in Fig. 5(b). Suppose a vertical drill is driven through the center of any generic grid cell C and it intersects the surface at point $T(x, y, z)$. The T1 fuzzy membership of points belonging to the cell in which (x, y) lies is z, where $z \in$ [0,1]. If we assume that the maximum number of points that can lie in any grid cell is N_0, then the number of points which need to be scattered in cell C is $N_0 z$.

Data generation from T2 fuzzy MFs:

For a T2 fuzzy MF, the representation consists of two surfaces, denoting the UMF and the LMF, respectively. We consider a vertical drill through the point (x, y), lying in cell C, which intersects the UMF and LMF at heights z_U and z_L, respectively.

(i) For an IT2 fuzzy MF, $f((x, y), u) = 1 \; \forall z_L \leq u \leq z_U$. Hence, to obtain a specific number of points that need to be present in this grid cell of the data set, we select a real number z uniformly at random from the range $[z_L, z_U]$. Then, similar to previous explanation, the number of points to be scattered may be determined as $N_0 z$. To select a number uniformly at random from a finite range, the Acceptance-Rejection [41] or Inverse Transform Sampling [42] methods may be used. In practice, the former is usually preferred since it does not require computation of the cumulative distribution function or its inverse.

(ii) For a GT2 fuzzy MF, the secondary membership may be in the form of a uniform or skewed Gaussian distribution. In either case, since the maximum value possible is 1, we have a normal distribution function over the range $[z_L, z_U]$. To generate a random number from this distribution, the Marsaglia-Bray method [43], or the Box-Muller method [44] can be used to the same effect.

Once we have selected the random real number z which defines the fuzzy membership of the cell C, the number of points to be scattered may be given as $N_0 z$ (as described). After generating the scatter plot, the *x* and *y* coordinates of the plot points are separated into labeled features x_1 and x_2 to obtain the 2-D data.

C. Similarity analysis

Wilcoxon "Signed-Rank" [45], [46] and "Rank-Sum" tests [47], [48] are used to compare the non-parametric data sets generated from the MFs, with the original data [49], [50].

Wilcoxon Signed-Rank (WSR) Test:

The test is applied over null hypothesis, i.e., if the difference of all pairs of data are computed and arranged in ascending order, the distribution must have an expected value of 0. Let us consider two sets of data, namely A = $\{a_1, a_2, \ldots, a_n\}$ and B = $\{b_1, b_2, \ldots, b_n\}$. After computing paired differences and ranking them, the Z-value (measures of standard deviation) of WSR test is determined as

$$Z = \frac{\left(W - \frac{n(n+1)}{4}\right)}{\sqrt{\frac{n(n+1)(2n+1)}{24}\left(\pi^+ + \pi^- - \left(\pi^+ - \pi^-\right)^2\right)}}, \qquad (18)$$

where $W = \Sigma_{i=1}^{n} \left[sgn\left(x_i - y_i\right) \cdot R_i\right]$, R_i = rank of the pair, and $\pi^+ = P\left(a_i > b_i\right)$ and $\pi^- = P\left(a_i < b_i\right)$ are sampling probabilities. Depending on the Z-value and an input α which determines the significance level of the decision of the hypothesis test, the test returns two statistics, namely p and h. The p-value of the test, returned as a real value between 0 to 1, is the probability of observing a test statistic as or more extreme than the observed value under the null hypothesis. Statistic h denotes the result of the hypothesis test. If h = 1, it indicates the rejection of the null hypothesis at the $100\alpha\%$ significance level and if h = 0, it indicates failure.

Wilcoxon Rank-Sum (WRS) Test:

This is also a non-parametric statistical hypothesis test similar to WSR, with the exception that the data sets are allowed to be of different sizes. The computation of test statistics differs from the WSR test in that the statistic U in the WRS test depends on the number of times sample b precedes sample a in an ordered arrangement of the elements in the two independent sets A and B. If the size of the samples A and B are n_{A} and n_{B}, respectively, such that $n_{\mathrm{A}} < n_{\mathrm{B}}$ then

$$U_{\mathrm{A}} = W - \frac{n_{\mathrm{A}}\left(n_{\mathrm{A}} + 1\right)}{2} \text{ and } U_{\mathrm{B}} = W - \frac{n_{\mathrm{B}}\left(n_{\mathrm{B}} + 1\right)}{2}. \qquad (19)$$

After computing the test statistic U, the Z-value of the WRS test can be determined as

$$Z = \frac{W - \mu_W}{\sigma_W} = \frac{W - \frac{n_{\mathrm{A}}(N+1)}{2}}{\sqrt{\frac{n_{\mathrm{A}} n_{\mathrm{B}}(N+1)}{12}}}, \qquad (20)$$

where N = n_{A} + n_{B}, μ_W = approximate mean of normal distribution, and σ_W = approximate standard deviation of normal distribution. Similar to WSR, the WRS test also returns the statistics p and h, depending upon the Z-value and significance level input α.

Suppose we have functions SIGNEDRANK(A,B) and RANKSUM(A,B) which return integral h values 0 or 1 according to the criteria described earlier. We will now describe an algorithm which takes two 1-D data sets, namely X_1 and X_2, as input and returns a minimal bin size such that the data sets cannot be proved dissimilar by Wilcoxon tests. If the sizes returned for two different pairs of data are a_i and b_i, respectively, such that $a_i < b_i$, it can be deduced that the similarity between the first pair is greater than the second pair. If the data is multidimensional, the same algorithm can be extended by creating the arrays A and B such that they are a linear representation of the data. This is analogous to representing multidimensional arrays by a simple array in programming logic. The algorithm is described below.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2017.2688342, IEEE Transactions on Fuzzy Systems

Algorithm 1 Iterative Algorithm for Determining Minimum Bin Size for Similarity

procedure WILCOXONMINIMALBINSIZE(X_1,X_2)
  $n \leftarrow$ size(X_1)
  $maxValue \leftarrow$ MAX{MAX(X_1),MAX(X_2)}
  $binSize \leftarrow maxValue$
  $dataIsSimilar \leftarrow true$
  while $binSize > 0$ AND $dataIsSimilar == true$ **do**
    $m \leftarrow maxValue/binSize$
    A, B $\leftarrow arraysofsizeminitializedto0$
    for $0 \leq i \leq n$ **do**
      A$[x_i/binSize] \leftarrow$ A$[x_i/binSize] + 1$
      B$[x_i/binSize] \leftarrow$ B$[x_i/binSize] + 1$
    end for
    if $SIGNEDRANK$(A, B) $== 0$ AND
     $RANKSUM$(A, B) $== 0$ **then**
      $binSize \leftarrow binSize/2$
    else
      $dataIsSimilar \leftarrow false$
    end if
  end while
  return $binSize$
end procedure

V. EXPERIMENTAL RESULTS

A. Synthetic data sets

Uniformly distributed rectangular data:

Let us consider a 2-D data set, namely Sample A, consisting of a total of 1500 data samples distributed uniformly across two clusters of different sizes, located at a uniform distance of 0.01 units. The first cluster is between (0.2,0.3) and (0.5,0.5), while the second ranges from (0.5,0.5) to (0.8,0.8).

On generating the crisp MF from Sample A, we obtain the surfaces represented by

$$\mu_{\mathbf{A}}(x,y) = \begin{cases} 1 & \text{if } x \in (0.2, 0.5], y \in (0.3, 0.5] \\ & \text{or } x \in (0.5, 0.8], y \in (0.5, 0.8] \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

Now, let us consider a T1 fuzzy MF generated from the same data set. On fitting two Gaussians G_1 and G_2 on the data samples and obtaining the required parameters: amplitude of the distribution a, mean $\boldsymbol{\mu}$, and covariance $\sum$, the T1 fuzzy MF can be written as the sum of G_1 and G_2, i.e.,

$$f(x,y) = \exp\left(-\frac{1}{2(1-\rho_1^2)}\left[\frac{(x-\mu_{1,X})}{\sigma_{1,X}^2} + \frac{(y-\mu_{1,Y})}{\sigma_{1,Y}^2} - \frac{2\rho_1(x-\mu_{1,X})(y-\mu_{1,Y})}{\sigma_{1,X}\sigma_{1,Y}}\right]\right) + \exp\left(-\frac{1}{2(1-\rho_2^2)}\left[\frac{(x-\mu_{2,X})}{\sigma_{2,X}^2} + \frac{(y-\mu_{2,Y})}{\sigma_{2,Y}^2} - \frac{2\rho_2(x-\mu_{2,X})(y-\mu_{2,Y})}{\sigma_{2,X}\sigma_{2,Y}}\right]\right), \tag{22}$$

where $\mu_{1,X}$ = 0.3553, $\mu_{1,Y}$ = 0.4053, $\mu_{2,X}$ = 0.6546, $\mu_{2,Y}$ = 0.6547, $\sigma_{1,X}$ = 0.0872, $\sigma_{1,Y}$ = 0.0582, $\sigma_{2,X}$ = 0.0871, $\sigma_{2,Y}$ = 0.0870, ρ_1 = 0.0106, and ρ_2 = 0.0120.

Since the normalized histograms generated from Sample A is precisely uniform, we can fit only one Gaussian each on either of the two clusters, and as such, obtaining an FOU for a T2 fuzzy MF is not possible. Hence, generating a crisp and a T1 fuzzy MF suffices in the case of Sample A. We now generate data sets back from the two kinds of MFs (crisp and T1 fuzzy) and compare their similarity to the original data using Algorithm 1.

TABLE I:
WILCOXON TEST STATISTICS FOR COMPARISON OF SAMPLE A WITH DATA SETS C2-A AND T12-A.

Bin Size (n)	Test Name	Test Statistics	Sample A vs C2-a data	Sample A vs T12-a data
0.2	WSR	p	1	0.4565
		h	0	0
	WRS	p	1	0.8753
		h	0	0
0.1	WSR	p	1	0.6600
		h	0	0
	WRS	p	1	0.3313
		h	0	0
0.05	WSR	p	0.7380	0.9035
		h	0	0
	WRS	p	0.8477	0.0452
		h	0	1
0.02	WSR	p	0.6148	0.9512
		h	0	0
	WRS	p	0.7211	0.0391
		h	0	1

The Wilcoxon's test results for the comparison between Sample A and data sets C2-a and T12-a, generated from the crisp and T1 fuzzy MF, respectively, are shown in Table I. We find that Wilcoxon's tests can disprove the null hypotheses for bin sizes 0.01 and 0.02, respectively. Hence, the data set generated using crisp MF is considered more similar to the original data than that generated from T1 fuzzy MF.

Normal distribution across two clusters in 1-D:

We now consider a 1-D data set, namely Sample B1 consisting of 500 data samples distributed uniformly across two Gaussian-shaped clusters of different sizes. The crisp, T1, IT2, SGT2, and AGT2 fuzzy MFs for Sample B1 are shown in Fig. 12(a), 12(b), 12(c), and 12(d), respectively.

The results of comparison of generated data with Sample B1 are shown in Fig. 13. It can be observed that although the different T2 fuzzy MFs provide similar accuracy of representation, they perform better than the crisp and T1 fuzzy MFs. This is expected due to the number of samples constituting the Gaussian may not be sufficient for a well-defined distribution, thereby leading to an uncertain Gaussian boundary.

Normal distribution across two clusters in 2-D:

Let us consider a 2-D data namely Sample B2 that consists of 5000 samples distributed uniformly across two Gaussian-

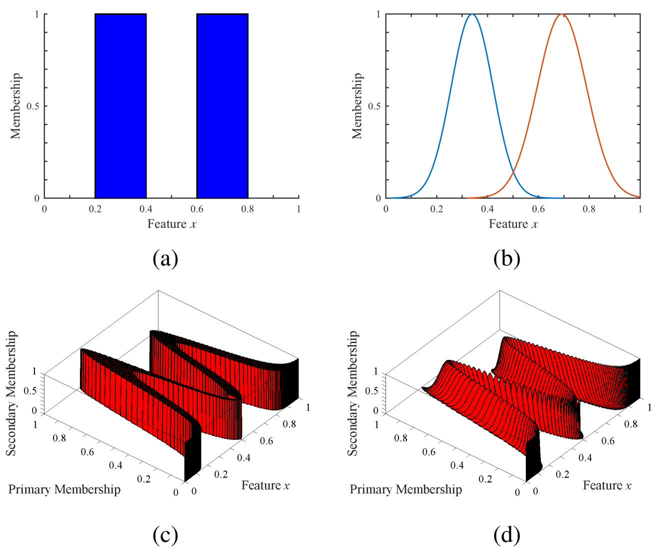


Fig. 12: MFs generated from Sample B1: (a) crisp, (b) T1 fuzzy, (c) IT2 fuzzy, and (d) SGT2 fuzzy.

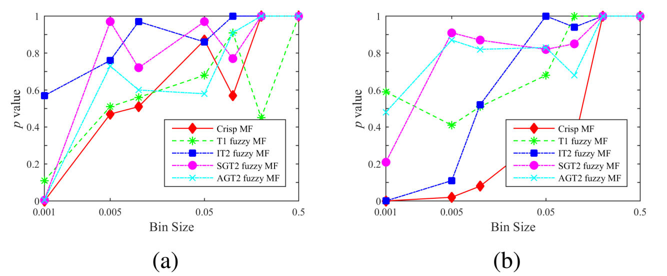


Fig. 13: Wilcoxon test statistics for comparison of Sample B1 with data generated from the corresponding crisp, T1, IT2 SGT2, and AGT2 fuzzy MFs: (a) Signed rank test and (b) Rank sum test.

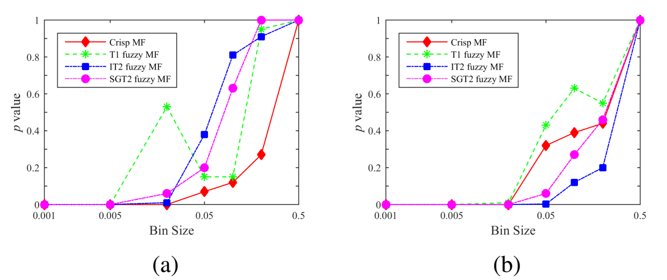


Fig. 14: Wilcoxon test statistics for comparison of Sample B2 with data generated from the corresponding crisp, T1, IT2, and SGT2 fuzzy MFs: (a) Signed rank test and (b) Rank sum test.

shaped clusters of different sizes. On comparing the generated data sets with Sample B2, the results obtained are given in Fig. 14. Two observations can be made from the statistics.

First, it may be noted that when the bin size becomes n = 0.02, Wilcoxon Rank-Sum test is able to disprove the null hypothesis for the data sets generated from all the MFs. However, since the p value for T1 fuzzy MF is significantly higher than the other MFs, it may be hypothesized that using the T1 fuzzy MF gives the most similar data.

Second, from the statistics obtained by comparing IT2 and SGT2 data with the original data set, we find that the

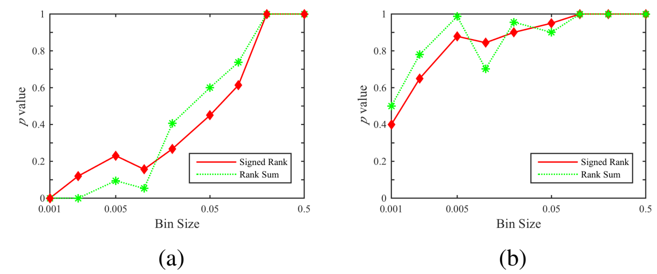


Fig. 15: Wilcoxon test statistics for data sets generated from (a) crisp MF and (b) T1 fuzzy MF of Sample B2.

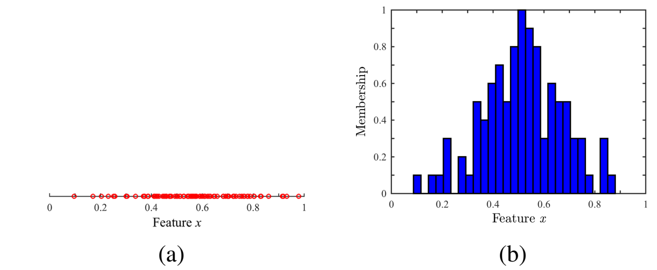


Fig. 16: Illustration of 1-D data set Sample C: (a) scatter plot and (b) the corresponding normalized histograms.

SGT2 data is more similar. This may be due to the fact that while generating data from IT2 fuzzy MF, we select a membership uniformly at random from the FOU. However, with a Gaussian distribution as the secondary membership, the probability that the chosen membership value is closer to the T1 fuzzy membership is higher. Therefore, data generated from a SGT2 fuzzy MF is closer to T1 fuzzy MF data than that generated from an IT2 fuzzy MF.

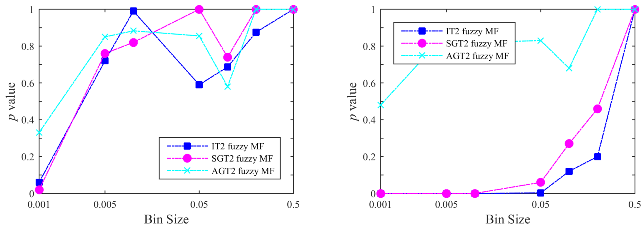
Normal distribution with noise in 1-D:

T2 fuzzy MFs assign uncertainty to primary memberships. From Fig. 14, the following observations can be made about the use of different kinds of T2 fuzzy MFs.

1) When the distribution of data is such that a simple Gaussian can accurately describe it without any added uncertainty in the form of deviation from the function, a T1 fuzzy MF is best suited to represent the data.
2) When random noise is added to a normal Gaussian distribution of data, such that data samples can lie anywhere between an UMF and an LMF and its location is equiprobable inside this range, an IT2 fuzzy MF describes the data set most accurately.
3) When the data set shows slight deviation from a normal Gaussian distribution, but small enough such that the majority of the data lies along the T1 fuzzy MF, the data set is most accurately represented by a SGT2 fuzzy MF.

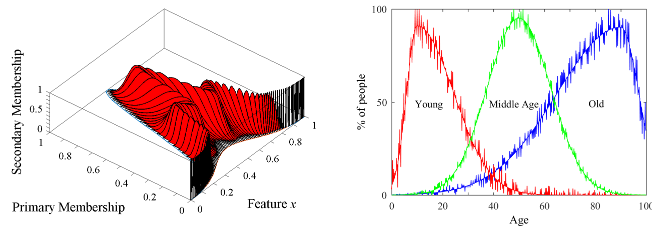
Let us now consider the 1-D data set Sample C shown in Fig. 16(a). The normalized histogram for the same is shown in Fig. 16(b). The IT2 and SGT2 fuzzy MFs for this data are similar to those shown in Fig. 8(a) and 9(a), and the AGT2 “snaky” fuzzy MF is shown in Fig. 18(a).

Let S1, S2, and S3 denote the data sets generated from IT2, SGT2, and AGT2 fuzzy MFs, respectively. Fig. 17 summarizes



(a) (b)

Fig. 17: Wilcoxon test statistics for comparison of Sample C with data generated from the corresponding IT2, SGT2, and AGT2 fuzzy MFs: (a) Signed rank test and (b) Rank sum test.



(a) (b)

Fig. 18: Illustration of a "snaky" T2 fuzzy MF: (a) the MF and (b) an example of data that can be modeled using it.

Wilcoxon test statistics when these are compared with Sample C. We find that the p value for n = 0.0005 is two orders of magnitude better in the case of a AGT2 "snaky" fuzzy MF data set, implying that it is most similar to Sample C.

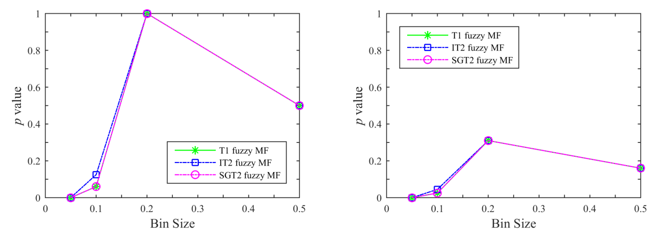
The "snaky" fuzzy MF may also be used to model situations when a variable is partitioned into a number of terms. For this, the domain of the variable can be normalized between 0 and 1, and the local points of maxima (peaks of the "snaky" Gaussian) are interpreted as the memberships of each of the partitions. Since the T1 fuzzy membership of the "snaky" surface is obtained by superimposing a periodically varying composite function over a normalized Gaussian [39], it may represent any number of partitions and any amount of variation in a partition. As an illustration, Fig. 18(b) shows a sample variable, namely "age," which can be partitioned into 3 fuzzy sets, i.e., "young," "middle age," and "old."

B. Iris data

The Iris data set [51] consists of 150 data samples of four features and three classes (50 samples for each class). One class is linearly separable from the other 2; the latter are not linearly separable from each other. From previous work [52], it has been found that smooth 2-D Gaussians can be fitted on to features 3 and 4. Hence, it may be reasonable to argue that if the samples are extended to all four features, hypergeometric Gaussian functions may approximate the samples. Therefore, we obtain the T1, IT2 and SGT2 fuzzy MFs for the 4-D Iris data set. From the Wilcoxon test statistics in Fig. 19, we find that the data generated from IT2 fuzzy MF is the most similar to the original data set. It is to be noted that although the null hypothesis is violated for all the comparisons when size of bins becomes less than 0.1, the p value, which indicates probability of finding similar data at random positions, is the highest for IT2, albeit only slightly. An IT2 fuzzy MF, hence, approximates the data set in the best possible manner. This result has also been verified earlier by Hwang and Rhee [15], [53].

C. Yeast data

The Yeast data set [51] consists of 1484 samples containing eight real-valued features. Since the number of bins increases exponentially with the number of dimensions and the partitioning along each dimension, this value becomes very large



(a) (b)

Fig. 19: Wilcoxon test statistics for Iris data vs data sets generated from T1, IT2, and SGT2 fuzzy MFs: (a) Signed rank test and (b) Rank sum test.

for dimensions greater than 10. Further, if the number of samples is small (approximately 0.02 times, experimentally) compared to the total number of bins, the data distribution becomes too sparse to be efficiently approximated. In such cases, our proposed method fails to approximate the original data using any of the MFs. This can be seen from the p values corresponding to n = 0.2, in Fig. 20.

From the figure, it may be observed that for n = 0.25, an SGT2 fuzzy MF is able to approximate the Yeast data most closely, as validated by the p values. We found from experiments that fitting 2 Gaussian hypergeometric surfaces on the samples generated the best possible approximation of the original data.

D. Madelon data

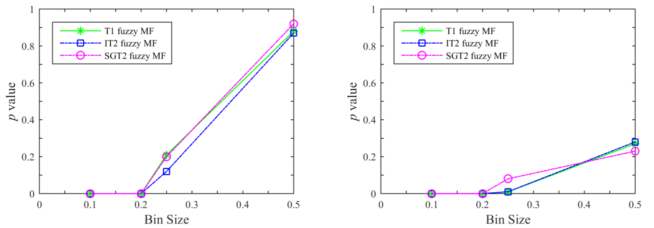
The Madelon data [54] is a two-class artificial data set which was part of the NIPS 2003 feature selection challenge. It has continuous input variables, consisting of 4400 instances and 500 features. With dimensionality reduction for efficiency, the results obtained are shown in Fig. 21. It may be observed that both an IT2 and SGT2 perform equally well for the given dataset.

E. Image segmentation

Image segmentation is a process of partitioning an image into some non-intersecting regions such that each region is homogeneous and the union of no two adjacent regions is homogeneous [55]. Fuzzy sets have been incorporated into a number of clustering algorithms for image segmentation [56], [57], especially when class boundaries are uncertain.

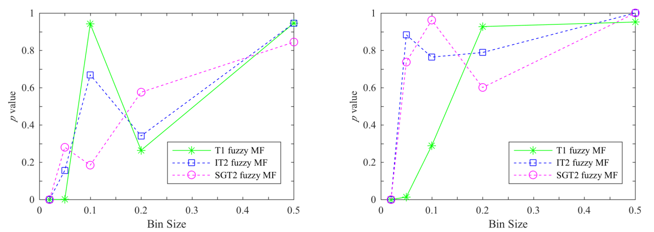
For an image consisting of multiple segments, applying filters such as median filter, Gaussian blur, squared variance,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2017.2688342, IEEE Transactions on Fuzzy Systems



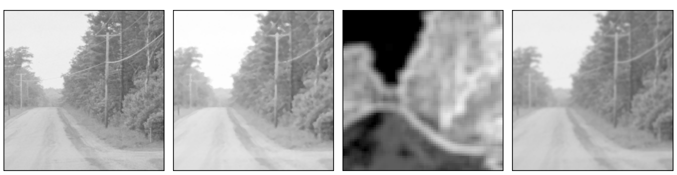
(a) (b)

Fig. 20: Wilcoxon test statistics for Yeast data vs data sets generated from T1, IT2, and SGT2 fuzzy MFs: (a) Signed rank test and (b) Rank sum test.



(a) (b)

Fig. 21: Wilcoxon test statistics for Madelon vs data sets generated from T1, IT2, and SGT2 fuzzy MFs: (a) Signed rank test and (b) Rank sum test.

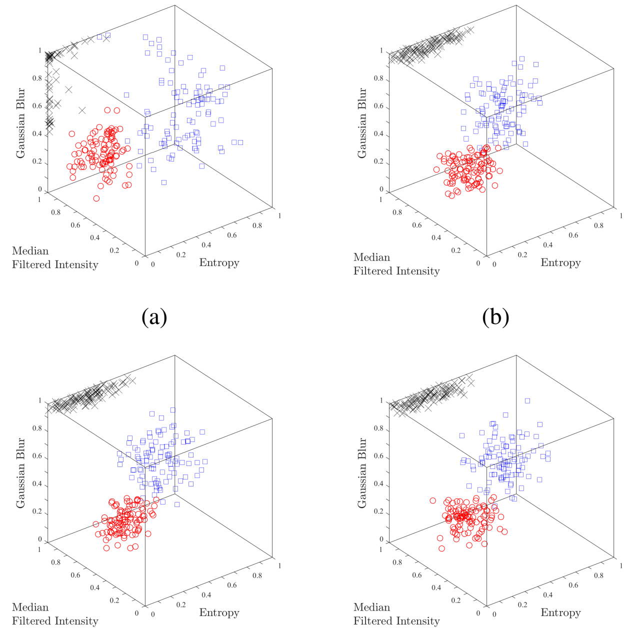


(a) (b) (c) (d)

Fig. 22: Feature images for natural scene: (a) original image, (b) median filtered intensity, (c) entropy, and (d) Gaussian blur.

to name a few, results in a multi-label classification/clustering problem with a multidimensional data set. Since each pixel may have some degree of membership in every class, this is a possible application for our proposed method.

Let us consider a natural scene image consisting of 3 regions, namely sky, road, and forest, shown in Fig. 22(a). Feature images are extracted by applying mask filters (i.e., median filtered intensity, entropy, and Gaussian blur) to the image. Figs. 22(b), 22(c), and 22(d) show the feature images obtained from the scene image. We randomly select 100 sample pixels from each region of the feature images and obtain the 3-D scatter plot given in Fig. 23(a). Further, on obtaining the T1, IT2, and SGT2 fuzzy MFs from the data and generating the data back from the MFs, we obtain the scatter plots shown in Fig. 23(b), 23(c), and 23(d), respectively. From Fig. 24, it may be observed that although the p-values obtained from T1, IT2, and SGT2 fuzzy MFs are relatively close for different values of n, only IT2 is able to withstand a null hypothesis test for n = 0.1. Therefore, an IT2 fuzzy MF may be considered the most suitable for the given image for the specified filters.



(a) (b) (c) (d)

Fig. 23: MFs generated from Sample B1: (a) crisp, (b) T1 fuzzy, (c) IT2 fuzzy, and (d) SGT2 fuzzy.

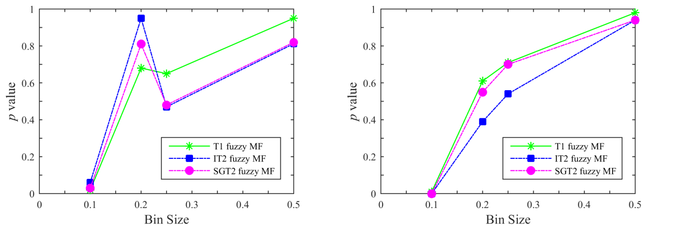
VI. Conclusion

In this paper, we have proposed a three-stage method for determining which MF, namely crisp, T1 fuzzy, IT2 fuzzy, or GT2 fuzzy, is most suitable for representing a data set, given some application requirements. In particular, we have looked at the generation of different types of MFs from data sets, and described an iterative algorithm for determining the similarity of generated data with the original data. For practical applications, the MF chosen to represent a data set may depend on the requirement of data to be generated, which can be broadly classified into three categories.

Category 1) Cases when the data follows a uniform random distribution, a crisp MF best represents the data set. For a crisp MF, every element in the fuzzy set has an equal probability of occurrence, hence satisfying the requirement of a uniform random distribution. No further flexibility may be obtained in this case, since it is not possible to bias a uniform random fuzzy without adding error into the generated data.

Category 2) Cases when the data distribution closely represents a continuous function of the constituent features without many deviations, a T1 fuzzy MF may be best suited for its representation. By definition, a T1 fuzzy MF is considered as a continuous function on the constituent features. Further, if flexibility is required with each instance of the generated data, an IT2 fuzzy MF may be used as it incorporates a uniform uncertainty with the primary membership.

Category 3) Cases when the data distribution is such that it vaguely follows a well-defined function on the constituent features and a significant number of data samples deviate from it. Here it may be useful to represent the data using a T2 fuzzy MF. Further, if the deviations themselves follow an observable pattern, such as a Gaussian, a SGT2 fuzzy MF may be used.



(a) (b)

Fig. 24: Wilcoxon test statistics for image segmentation features vs data sets generated from T1, IT2, and SGT2 fuzzy MFs: (a) Signed rank test and (b) Rank sum test.

We obtained MFs for 1-D as well as multidimensional data sets, and subsequently inferred results from comparison using Wilcoxon Signed Rank and Wilcoxon Rank Sum tests, as described in Fig. 1 which illustrated this technique of data generation and analysis for a 1-D data sets. In the case of IRIS data, the AGT2 "snaky" fuzzy MF was not generated, since the improvement in approximation using this fuzzy MF may be insignificant compared to the increase in computational complexity. However, using advanced distributed computing algorithms such as those for fast computation of separable functions using a randomization approach, it may be possible to implement this fuzzy MF for complex data distributions as well. Such an implementation may be useful in applications which require high precision of representation, such as medical applications and time-series forecasting. In applications like classification of coded A/V streams and autonomous navigation systems which require flexibility and ease of computation over accuracy, it may be convenient to use less complex MFs such as T1 or IT2. This categorization may be considered advantageous for identifying applications where T2 FLSs may outperform T1 FLSs, or vice versa.

Finally, it is to be noted that depending upon the application, subsets of features of multidimensional data may also be considered for analysis using the methods described. For example, the methods described for 1-D and 2-D data sets may be applied for multidimensional data where subsets of features are either mutually independent or pairwise dependent, respectively.

ACKNOWLEDGMENT

This work was supported by the Technology Innovation Program of the Korea Institute for Advancement of Technology (KTA) granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea (No. 2015-122).

REFERENCES

[1] L. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[2] J. Mendel, *Uncertain Rule-based Fuzzy Logic System: Introduction and New Directions*. Prentice–Hall PTR, 2001.

[3] ——, "General type-2 fuzzy logic systems made simple: a tutorial," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 5, pp. 1162–1182, 2014.

[4] J. Mendel and R. John, "Type-2 fuzzy sets made simple," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 2, pp. 117–127, 2002.

[5] J. Mendel, "Computing derivatives in interval type-2 fuzzy logic systems," *IEEE Trans. Fuzzy Syst.*, vol. 12, no. 1, pp. 84–98, 2004.

[6] J. Mendel and X. Liu, "Simplified interval type-2 fuzzy logic systems," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 6, pp. 1056–1069, 2013.

[7] Q. Liang and J. Mendel, "Mpeg vbr video traffic modeling and classification using fuzzy technique," *IEEE Trans. Fuzzy Syst.*, vol. 9, no. 1, pp. 183–193, 2001.

[8] H. Hagras, "A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots," *IEEE Trans. Fuzzy Syst.*, vol. 12, no. 4, pp. 524–539, 2004.

[9] Q. Liang and J. Mendel, "Equalization of nonlinear time-varying channels using type-2 fuzzy adaptive filters," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 5, pp. 551–563, 2000.

[10] I. Türkşen, "Interval-representation of sets: Dempster-pawlak-turksen schema," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2002, pp. 602–606.

[11] N. Karnik and J. Mendel, "Applications of type-2 fuzzy logic systems to forecasting of time-series," *Info. Sci.*, vol. 120, no. 1, pp. 89–111, 1999.

[12] P. Innocent, R. John, I. Belton, and D. Finlay, "Type 2 fuzzy representations of lung scans to predict pulmonary emboli," in *Joint 9th IFSA World Cong. and 20th NAFIPS Int. Conf.*, 2001, pp. 1902–1907.

[13] R. John, P. Innocent, and M. Barnes, "Neuro-fuzzy clustering of radiographic tibia image data using type 2 fuzzy sets," *Info. Sci.*, vol. 125, no. 1, pp. 65–82, 2000.

[14] P. Innocent and R. John, "Type-2 fuzzy diagnosis," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2002, pp. 1326–1330.

[15] C. Hwang and F. C.-H. Rhee, "Uncertain fuzzy clustering: interval type-2 fuzzy approach to c-means," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 1, pp. 107–120, 2007.

[16] M. Raza and F. C.-H. Rhee, "Interval type-2 approach to kernel possibilistic c-means clustering," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2012, pp. 1–7.

[17] R. John and S. Coupland, "Type-2 fuzzy logic: challenges and misconceptions [discussion forum]," *IEEE Computational Intelligence Magazine*, vol. 7, no. 3, pp. 48–52, 2012.

[18] J. Mendel, "A quantitative comparison of interval type-2 and type-1 fuzzy logic systems: First results," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2010, pp. 1–8.

[19] D. Wu and W. Tan, "Type-2 FLS modeling capability analysis," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2005, pp. 242–247.

[20] L. Zadeh, "Discussion: Probability theory and fuzzy logic are complementary rather than competitive," *Technometrics*, vol. 37, no. 3, pp. 271–276, 1995.

[21] J. Bezdek, "A review of probabilistic, fuzzy, and neural models for pattern recognition," *Journal of Intelligent & Fuzzy Systems*, vol. 1, no. 1, pp. 1–25, 1993.

[22] B. Kosko, "Fuzziness vs. probability," *Int. Journal General Syst.*, vol. 17, no. 2-3, pp. 211–240, 1990.

[23] R. Zwick, E. Carlstein, and D. Budescu, "Measures of similarity among fuzzy concepts: A comparative analysis," *Int. Journal of Approx. Reasoning*, vol. 1, no. 2, pp. 221–242, 1987.

[24] C. Krumhansl, "Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density," *Psychological Review*, vol. 85, no. 5, pp. 445–463, 1978.

[25] A. Tversky, "Features of similarity," *Psychological Review*, vol. 84, no. 4, p. 327, 1977.

[26] D. Dubois and H. Prade, "The three semantics of fuzzy sets," *Fuzzy Sets and Systems*, vol. 90, no. 2, pp. 141–150, 1997.

[27] J. Kolodner, *Case-based reasoning*. Morgan Kaufmann, 2014.

[28] C. Murthy, S. Pal, and D. Majumder, "Correlation between two fuzzy membership functions," *Fuzzy Sets and Systems*, vol. 17, no. 1, pp. 23–38, 1985.

[29] P. Bonissone, "A fuzzy sets based linguistic approach: theory and applications," in *Proc. 12th Conf. Winter Simulation*, 1980, pp. 99–111.

[30] D. Dalalah, "Piecewise parametric polynomial fuzzy sets," *International journal of approximate reasoning*, vol. 50, no. 7, pp. 1081–1096, 2009.

[31] B. Choi and F. C.-H. Rhee, "Interval type-2 fuzzy membership function generation methods for pattern recognition," *Info. Sci.*, vol. 179, no. 13, pp. 2102–2122, 2009.

[32] F. C.-H. Rhee and R. Krishnapuram, "Fuzzy rule generation methods for high-level computer vision," *Fuzzy Sets and Systems*, vol. 60, no. 3, pp. 245–258, 1993.

[33] S. Pal and P. Mitra, "Case generation using rough sets with fuzzy representation," *IEEE Trans. Knowledge Data Eng.*, vol. 16, no. 3, pp. 293–300, 2004.

[34] J. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 1981.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2017.2688342, IEEE Transactions on Fuzzy Systems

[35] B. Stuckman and E. Easom, "A comparison of bayesian/sampling global optimization techniques," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 22, no. 5, pp. 1024–1032, 1992.

[36] D. Jones, M. Schonlau, and W. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global Optimization*, vol. 13, no. 4, pp. 455–492, 1998.

[37] N. Karnik and J. Mendel, "Introduction to type-2 fuzzy logic systems," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 1998, pp. 915–920.

[38] C. Walker and E. Walker, "The algebra of fuzzy truth values," *Fuzzy Sets and Systems*, vol. 149, no. 2, pp. 309–347, 2005.

[39] D. Raj, K. Tanna, B. Garg, and F. C.-H. Rhee, "Visual analysis and representations of type-2 fuzzy membership functions," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2016, pp. 550–554.

[40] J. Mendel, R. John, and F. Liu, "Interval type-2 fuzzy logic systems made simple," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 6, pp. 808–821, 2006.

[41] B. Flury, "Acceptance-rejection sampling made easy," *SIAM Review*, vol. 32, no. 3, pp. 474–476, 1990.

[42] R. Lewitt, "Reconstruction algorithms: transform methods," *Proc. IEEE*, vol. 71, no. 3, pp. 390–408, 1983.

[43] G. Marsaglia and T. Bray, "A convenient method for generating normal variables," *SIAM review*, vol. 6, no. 3, pp. 260–264, 1964.

[44] E. Golder and J. Settle, "The box-muller method for generating pseudo-random normal deviates," *Applied Statistics*, pp. 12–20, 1976.

[45] R. Randles, "Wilcoxon signed rank test," *Encyclopedia of Statistical Sci.*, 1988.

[46] B. Rosner, R. Glynn, and M. Lee, "The wilcoxon signed rank test for paired comparisons of clustered data," *Biometrics*, vol. 62, no. 1, pp. 185–192, 2006.

[47] W. Haynes, "Wilcoxon rank sum test," in *Encyclopedia of Systems Bio.* Springer, 2013, pp. 2354–2355.

[48] F. Wilcoxon, S. Katti, and R. Wilcox, "Critical values and probability levels for the wilcoxon rank sum test and the wilcoxon signed rank test," *Selected Tables in Mathematical Stats.*, vol. 1, pp. 171–259, 1970.

[49] J. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*. Springer, 2011.

[50] J. Higgins, *Intro. Modern Nonparametric Statistics*. Cengage Learning, 2003.

[51] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[52] R. Hathaway and J. Bezdek, "Fuzzy c-means clustering of incomplete data," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 31, no. 5, pp. 735–744, 2001.

[53] F. C.-H. Rhee, "Uncertain fuzzy clustering: insights and recommendations," *IEEE Comp. Intelligence Magazine*, vol. 2, no. 1, pp. 44–56, 2007.

[54] I. Guyon, S. Gunn, A. Ben-Hur, and G. Dror, "Result analysis of the nips 2003 feature selection challenge." in *NIPS*, vol. 4, 2004, pp. 545–552.

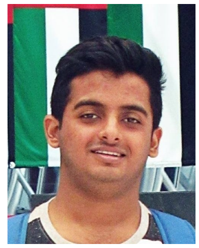
[55] N. Pal and S. Pal, "A review on image segmentation techniques," *Pattern Recognition*, vol. 26, no. 9, pp. 1277–1294, 1993.

[56] K. Chuang, H. Tzeng, S. Chen, J. Wu, and T. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *Computerized Medical Imaging and Graphics*, vol. 30, no. 1, pp. 9–15, 2006.

[57] O. Tobias and R. Seara, "Image segmentation by histogram thresholding using fuzzy sets," *IEEE Trans. Image Processing*, vol. 11, no. 12, pp. 1457–1465, 2002.

Desh Raj is an undergraduate in the Department of Computer Science & Engineering at IIT Guwahati, India. His research interests include natural language processing, deep learning, computer vision, and fuzzy logic systems.

He is currently working on attention-based neural networks for relation classification in biomedical data, text readability analysis using language modeling, and egocentric object recognition using spatial transformer networks.



Aditya Gupta is currently pusuing his B.Tech. degree in Mathematics and Computing from the Department of Mathematics, Indian Institute of Technology, Guwahati, India. His research interests are graph theory, unsupervised machine learning, and fuzzy logic systems.

He is currently working on devising efficient numerical schemes using Local Defect Correction in regions of high activity, in conjunction with Cubic Splines and Higher Order Compact(HOC) schemes.

Bhuvnesh Garg is currently pursuing his B.Tech. degree in Mathematics and Computing from the Dept. of Mathematics, Indian Institute of Technology, Guwahati, India.

His research interests are fuzzy logic systems, statistics, and mathematical finance with special emphasis on portfolio theory.

Kenil Tanna was born in Ahmedabad, India on February 1, 1996. He is currently pursuing his B.Tech. Degree in the Department of Computer Science and Engineering from Indian Institute of Technology, Guwahati where he started in 2013.

He is currently working on Video Surveillance using CNN and his current research interests include fuzzy logic, computer vision, machine learning, and deep learning.

Frank Chung-Hoon Rhee (S'91M'93) was born on March 3, 1963, in Seoul, Korea. He received the B.S. degree in electrical engineering from the University of Southern California (USC), Los Angeles, in 1985. He attended the University of Missouri, Columbia, from 1985 to 1987 and from 1990 to 1993, where he received the M.S. and Ph.D. degrees in electrical engineering, respectively.

From 1994 to 1995 he was a Senior Member of the Engineering Staff in the High-Speed Network Access Section at the Electronics and Telecommunications Research Institute (ETRI), Taejon, Korea, where his work involved applications of group communications. Since September 1995, he has been a faculty member in the School of Electrical Engineeering and Computer Science at Hanyang University, Korea, where he is currently a Professor. He served as the research Vice Dean of Engineering from 2010 to 2012. His current research interests include applications of computer vision, pattern recognition, and all aspects of computational intelligence.

Dr. Rhee was the program chair for the 2009 IEEE International Conference on Fuzzy Systems, Jeju, Korea. He has also served on several domestic and international conference program committees. He is the Korea representative for the International Fuzzy Systems Association (IFSA), and Vice President and a life member of the Korean Institute of Intelligent Systems.