On the Robustness of Dimensionality Reduction Methods in Selecting Type-1 and Type-2 Fuzzy Membership Functions for High-dimensional Data

Annway Samal*, Arnav Gupta* and Frank Chung-Hoon Rhee†, *Senior Member, IEEE*
*Department of Mathematics
Indian Institute of Technology Guwahati, Guwahati, India 781039
Email: {s.annway, arnavg3010}@iitg.ac.in
†Department of Electrical Engineering
Hanyang University, Ansan-Si, Korea 15588
Email: frhee@hanyang.ac.kr

***Abstract*—The Wilcoxon Minimal Bin Sizes algorithm has been recently proposed as a deterministic method for selecting the most suitable fuzzy membership function for representing a multi-dimensional dataset, based on null hypothesis testing. Although the method is theoretically robust, it may require large amounts of computations in practice for high-dimensional data, consisting of a few hundred dimensions. In this paper, we extend the algorithm for such datasets by evaluating several dimensionality reduction techniques to select the most suitable which may be integrated with the method without significantly affecting its accuracy. The motivation behind this approach is that high-dimensional datasets are gaining popularity in various fields such as medicine, image processing, geolocation, biochemistry, and computational linguistics, and for most real datasets, careful selection of a small number of features may localize enough variance such that the data can be effectively represented. We validate the robustness of our method on the 8-dimensional Yeast dataset and demonstrate its applicability on the 500-dimensional Madelon dataset.**

I. Introduction

Clustering and pattern recognition in high-dimensional spaces are fundamental challenges in data analysis [1]. Fuzzy Logic Systems (FLSs) [2] have gained widespread adoption for these tasks due to their ability to model the inherent uncertainty and vagueness in real-world data. Type-1 (T1) fuzzy sets [3] are limited by precise membership functions (MFs) that struggle with high noise [4]. Consequently, Type-2 (T2) [5], [6] and Interval Type-2 (IT2) [7] sets were developed (Fig. 1), offering a Footprint of Uncertainty (FOU) for superior modeling in noisy environments [8]–[11]. These MFs are employed in various algorithms like fuzzy C-means [12]–[14], document categorization [15], and parameter adaptation [16].

However, a critical dilemma for researchers is selecting the most appropriate MF (e.g., T1, IT2, or General T2) for a given dataset. Choosing a more complex MF increases computational cost, while a simpler MF may fail to represent the data's structure accurately. While T2 systems often provide better modeling, they do not always outperform T1 systems [17]. To address this, Raj et al. [18] proposed the Wilcoxon Minimal Bin Size (WMBS) algorithm, a deterministic method that uses statistical hypothesis testing [19], [20] and visual analysis [21] to identify the optimal MF. While theoretically robust, the WMBS algorithm suffers from the "curse of dimensionality": as the number of feature dimensions increases, the number of bins required for the test grows exponentially, rendering the method computationally infeasible for high-dimensional datasets [22] (e.g., hundreds of features).

This paper addresses this limitation by investigating the intersection of Dimensionality Reduction (DR) and Fuzzy MF selection. We premise that if a DR technique can compress high-dimensional data while preserving its fuzzy signature, the inherent uncertainty structure-it can serve as a helper to make the WMBS algorithm feasible. Historically, DR has evolved from linear methods like Principal Component Analysis (PCA) [23], factor analysis [24], and classical scaling [25] to spectral [26] and nonlinear methods [27]. In comparative studies [28], linear techniques often prove most reliable. It is crucial to clarify that our primary goal is not merely data compression or feature selection for its own sake, but rather optimal Fuzzy MF selection. We assess usage in the context of representational fidelity: finding the MF that best captures the geometric and statistical properties of the data for downstream tasks like classification.

To this end, we propose an extended framework that integrates DR techniques with the WMBS algorithm. The primary contributions of this paper are:

- Formulating a methodology to extend the WMBS algorithm to high-dimensional datasets ($d \gg 10$) by evaluating DR techniques as preprocessing steps.
- Introducing the Resonance criterion to determine if a DR technique preserves fuzzy characteristics for optimal MF selection.
- Performing a comparative analysis of four distinct DR methods-PCA, Kernel PCA [29], Probabilistic PCA [30], and t-Stochastic Neighborhood Embedding (t-SNE) [31] to determine which is most resonant for fuzzy modeling based on information sufficiency [32].
- Empirically validate this approach on the 8-D Yeast

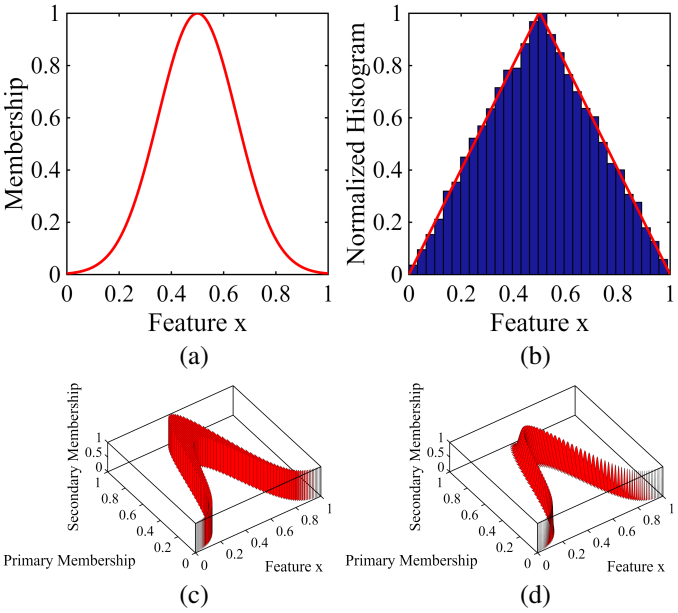

Fig. 1: Different types of fuzzy membership functions: (a) Gaussian T1, (b) triangular T1, (c) IT2 and (d) symmetric Gaussian T2.

dataset [33] and the 500-dimensional Madelon dataset [34], demonstrating that specific DR techniques can enable accurate MF selection in high-dimensional spaces where standard approaches fail.

The remainder of this paper is organized as follows. In Section II, we briefly discuss the WMBS algorithm for multi-dimensional (multi-D) datasets, and summarize the different methods for dimensionality reduction. Section III describes our proposed extension to the algorithm in some detail. In Section IV, we discuss the datasets and the experimental setup, and then proceed to describe the results obtained on the different datasets. We conclude with a discussion of possible improvements to the method in Section V.

II. Background

A. WMBS algorithm

The WMBS algorithm is an iterative, deterministic method that selects the most suitable fuzzy MF by testing the statistical similarity between original data and data regenerated from the MF. The process follows three stages: 1) *Forward stage*: Generating MFs from the dataset; 2) *Reverse stage*: Regenerating synthetic data via Inverse Transform Sampling; and 3) *Similarity analysis*: Using Wilcoxon Signed-Rank (WSR) [20] and Rank-Sum (WRS) [21] tests to find the minimum bin size (maximum precision) where the p-value remains above a threshold (e.g., 0.05) . As shown in Algorithm 1, the bin size is halved iteratively, making the complexity logarithmic.

B. Data Generation from Fuzzy MFs

A key component of the WMBS algorithm is the Reverse Stage, where data is generated back from a membership function to test its representational fidelity. To generate data from a specific Fuzzy MF (e.g., Gaussian T1), we employ a rejection sampling technique (Inverse Transform Sampling). The domain is sampled uniformly, and points are accepted based on their membership grade $\mu(x)$, effectively creating a synthetic dataset that statistically mirrors the uncertainty modeled by the MF. Unlike fuzzy control systems that partition individual variables, this study focuses on multivariate fuzzy membership functions (e.g., a multivariate Gaussian or interaction of separate dimensions) to model the geometric structure of the high-dimensional data cluster as a whole. This treats the problem as one of fuzzy pattern recognition rather than linguistic approximation.

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**
    $m \leftarrow \lceil maxValue/binSize \rceil$
    $A, B \leftarrow$ arrays of size m initialized to 0
    for $0 \leq i < n$ **do**
      $A[x_{1i}/binSize] \leftarrow A[x_{1i}/binSize] + 1$
      $B[x_{2i}/binSize] \leftarrow B[x_{2i}/binSize] + 1$
    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

C. Dimensionality reduction techniques

PCA: Consider a random vector $X = [X_1, ..., X_p]^T$ with population variance-covariance matrix [23]

$$var(X) = \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{pmatrix}. \quad (1)$$

Each of the linear combinations

$$Y_1 = e_{11}X_1 + e_{12}X_2 + \cdots + e_{1p}X_p \quad (2)$$
$$Y_2 = e_{21}X_1 + e_{22}X_2 + \cdots + e_{2p}X_p \quad (3)$$
$$\vdots \quad (4)$$
$$Y_p = e_{p1}X_1 + e_{p2}X_2 + \cdots + e_{pp}X_p \quad (5)$$

may be thought of as a linear regression, predicting Y_i from $X_1, X_2, ..., X_p$. Y_i has a population variance

$$var(Y_i) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{ik}e_{il}\sigma_{kl} = e_i'\Sigma e_i. \quad (6)$$

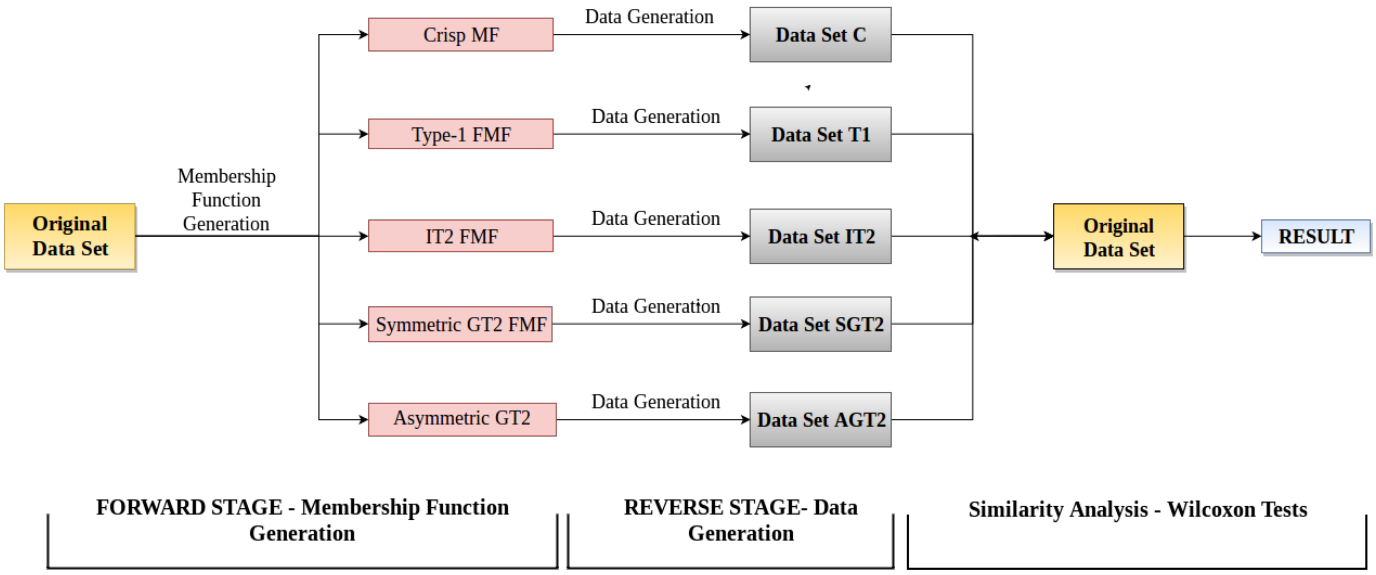

Fig. 2: Flow diagram for the WMBS algorithm for determining the most suitable fuzzy MF for a given dataset.

Moreover, Y_i and Y_j have a population covariance $cov(Y_i, Y_j) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{ik}e_{jl}\sigma_{kl} = e_i'\Sigma e_j$ where the coefficients e_{ij} are collected into the vector $(e_{i1}, ..., e_{ip})^T$.

The first principal component is the linear combination of variables that has maximum variance. More formally, we select $e_{11}, e_{12}, ..., e_{1p}$ that maximizes $var(Y_1) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{1k}e_{1l}\sigma_{kl} = e_1'\Sigma e_1$ subject to the constraint that $e_1'e_1 = \sum_{j=1}^{p} e_{1j}^2 = 1$.

For the i^{th} principal component, we select $e_{i1}, e_{i2}, ..., e_{ip}$ that maximizes $var(Y_i) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{ik}e_{il}\sigma_{kl} = e_i'\Sigma e_i$, subject to the constraints $e_i'e_i = \sum_{j=1}^{p} e_{ij}^2 = 1$ and $cov(Y_j, Y_i) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{jk}e_{il}\sigma_{kl} = e_j'\Sigma e_i = 0, \forall j < i$, i.e., the new component is uncorrelated with all the previously defined components.

Kernel PCA: Given N points $x_i \in \mathbb{R}^d (d < N)$, if we map them to an N-dimensional space using the transformation function $\Phi(x_i)$, where $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^N$, it becomes convenient to obtain a hyperplane that divides the points into arbitrary clusters [29]. Further, to avoid computations in high-dimensional spaces, we define an $N \times N$ kernel K as

$$K = k(x, y) = (\Phi(x), \Phi(y)) = \Phi(x)^T\Phi(y) \tag{7}$$

which represents the inner product space. The kernel-formulation of PCA is restricted in that it computes not the principal components themselves, but the projections of our data onto those components. To evaluate the projection from a point in the feature space $\Phi(x)$ onto the k^{th} principal component V^k, we compute

$$V^{k^T}\Phi(x) = (\sum_{i=1}^{N} a_i{}^k\Phi(x_i))^T\Phi(x). \tag{8}$$

To calculate a_i^k, we solve the eigenvector equation

$$N\lambda a = Ka, \tag{9}$$

where N is the number of data points in the set, and λ and a are the eigenvalues and eigenvectors of K. Then, to normalize the eigenvectors a^k, we require that

$$1 = (V^k)^T V^k. \tag{10}$$

In practice, a large dataset leads to a large K, such that storage becomes an issue. To alleviate this, clustering is performed on the dataset, and means of the clusters are used to populate the kernel. Since even this method may yield a relatively large K, it is common to compute only the top P eigenvalues and eigenvectors of K.

Probabilistic PCA: An arbitrary set of centered data with n observations and d dimensions, $Y = [Y_1, ..., Y_n]^T$ may be viewed as having a linear relationship with some embedded latent-space data x_n [30]. Assume Gaussian noise as $y_n = Wx_n + n$, where x_n is the q-dimensional latent variable associated with each observation, and $W \in \mathbb{R}^{D\times q}$ is the transformation matrix which relates the observed space to the latent space.

Considering a spherical Gaussian distribution for the noise with a mean of zero and a covariance of $\beta^{-1}I$, the likelihood for a particular observation y_n can be expressed as $p(y_n|x_n, W, \beta) = N(y_n|Wx_n, \beta^{-1}I)$ where N denotes the distribution function of a multivariate normal distribution for y_n with, in this case, mean Wx_n and covariance matrix $\beta^{-1}I$. We define a conditional probability model for the high dimensional data along with some Gaussian noise. To obtain a solution for W, we take the standard approach of marginalizing x_n, putting a Gaussian prior on W and solving using maximum likelihood.

To obtain the marginal likelihood we integrate over the latent variables, i.e.,

$$p(y_n|W, b) = \int p(y_n|x_n, W, b)p(x_n)dx_n, \tag{11}$$

with a unit covariance, zero mean Gaussian distribution as the prior distribution over x_n. The marginal likelihood of y_n is then described as $p(y_n|W,\beta) = N(y_n|0, WW^T + \beta^{-1}I)$. Assuming identical and independent data, the total likelihood is the product of the individual probabilities, i.e. $p(Y|W,\beta) = \prod_{n=1}^{N} p(y_n|W,\beta)$. A solution for W is obtained by maximizing this probability.

t-SNE: Given N high-dimensional samples $x_1, ..., x_N$, t-SNE first computes probabilities p_{ij} that are proportional to the similarity of objects x_i and x_j [31], as

$$p_{j|i} = \frac{exp(-||x_i - x_j||^2/2\sigma_i^2)}{\sum_{k\neq i} exp(-||x_i - x_k||^2/2\sigma_i^2)} \tag{12}$$

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N} \tag{13}$$

The bandwidth σ_i of the Gaussian kernels is chosen such that the perplexity of the conditional distribution equals a predefined perplexity. For this reason, smaller (larger) values of σ_i are used in denser (sparser) parts of the data space. The objective of t-SNE is to learn a d-dimensional map $y_1, ..., y_N$ ($y_i \in \mathbb{R}^d$) that accurately reflects the similarities p_{ij}. For this, it measures similarities q_{ij} between two points in the map y_i and y_j as

$$q_{ij} = \frac{(1 + ||y_i - y_j||^2)^{-1}}{\sum_{k\neq m}(1 + ||y_k - y_m||^2)^{-1}}. \tag{14}$$

The locations of the points y_i are determined by minimizing the (non-symmetric) KL divergence [32] of the distribution Q from the distribution P, i.e.,

$$KL(P||Q) = \sum_{ij} p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{15}$$

The result of this optimization, which is performed using gradient descent, is a map that adequately reflects the similarities between the high-dimensional inputs.

III. Proposed Method

In this section, we outline our proposed methodology for comparing the different dimensionality reduction techniques as a preprocessing module for the WMBS algorithm [18], [21]. Let us first define some new terms for this purpose.

1) A dimensionality reduction technique is said to be in resonance with the WMBS algorithm, if the output of the algorithm on the original dataset is the same as its output on the dataset obtained after removing a significant number of features.
2) A d-dimensional dataset is said to be sufficiently representable with respect to the WMBS algorithm, if it may be reduced without significant loss in variance, using some dimensionality reduction technique, to N features, such that $N < k$ for some k which depends on the system configuration such as computation power and storage space.
3) Any technique which may reduce a significantly representable dataset is called a helper.

We now state the following hypothesis without proof: The WMBS algorithm is robust, in practice, for any sufficiently representable dataset if it has at least one helper which is in resonance with the algorithm.

With this definition in place, our comparison methodology has two objectives: (i) to identify a helper technique for reducing a given dataset, subject to the system constraint k, and (ii) to show that the identified helper is resonant with the WMBS algorithm.

In most modern systems and for current trends in memory sizes, k has been found to be approximately 10-12. Further, for most datasets (where d is in the order of a few hundreds) which have been organized from real sensors, the variance in the features may be sufficiently correlated such that a subset of features (smaller than the aforementioned k) may be able to adequate for capturing the variance information. For such datasets, most of the popular dimensionality reduction techniques may be identified as helpers, and therefore our first objective is met.

The remaining task, therefore, is to identify the helper that is in resonance with the WMBS algorithm. Without loss of generality, suppose there are two helpers P and Q for a given d-dimensional dataset X. Let us assume that the helpers may be able to reduce X to either of dimensions d_1 or d_2 ($d_1, d_2 < k$, where k is system-dependent). Let $X_1^P(X_1^Q)$ and $X_2^P(X_2^Q)$ be the reduced datasets containing d_1 and d_2 features, respectively, obtained from X using helper P (Q).

Now, if each of these datasets are provided as input to the WMBS algorithm, suppose the most appropriate fuzzy MF that may represent them (i.e., the output of the WMBS algorithm) be denoted by M_1^P, M_2^P, M_1^Q and M_2^Q, respectively. Then, we have the following.

If $M_1^P(M_1^Q)$ is found to be the same as $M_2^P(M_2^Q)$, then P (Q) may be considered to be in resonance with the WMBS algorithm. If both P and Q are resonant helpers, either method may be chosen for dimensionality reduction depending upon their computational complexity. If neither P or Q are resonant, the WMBS algorithm may not be guaranteed to be robust on the dataset.

If M_1^P is found to be the same as M_2^P, this suggests that the reduced datasets X_1^P and X_2^P represent almost similar variance despite having different number of features. The smaller d_i ($i \in \{1, 2\}$) may then be considered to contain sufficient information so as to represent the entire dataset X. With such an assumption, it may be argued then that the WMBS algorithm, given the original dataset X, would provide the same result as either X_1^P or X_2^P. The helper P, therefore, is in resonance with the algorithm and may be used for dimensionality reduction. This method of determining whether a helper is resonant is also outlined in Fig. 3.

IV. Experimental Results

A. Experimental Setup

To ensure reproducibility, we detail the hyperparameters used in our experiments.

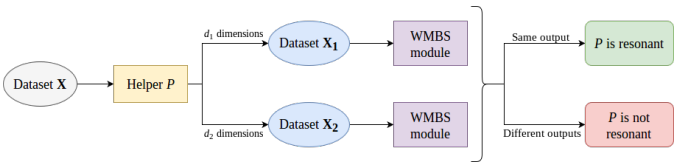

Fig. 3: Outline of the method to determine if a helper P is resonant with the WMBS algorithm for a dataset X.

TABLE I: Hyperparameter Settings

Algorithm	Parameter Setting
PCA	95% Variance Retained
k-PCA	RBF Kernel, $\sigma = 1.0$
t-SNE	Perplexity = 30, Learning Rate = 200
WMBS	Initial Normalized Bin Size = 0.1
WMBS	Significance Threshold $\alpha = 0.05$

We first validate our proposed method and hypothesis on the 8-D Yeast dataset [33], obtained from the UCI machine learning repository. The dataset consists of 1484 instances and the classification objective is to predict the cellular localization sites of proteins. Raj et al. [18] have earlier shown that for bin size $n = 0.25$, an SGT2 fuzzy MF is able to approximate the Yeast data most closely.

Fig. 4 represents p-statistic vs. bin size n obtained on applying the WMBS algorithm on the original and the reduced Yeast datasets (using the PCA method). Conceivably, the values of the p-statistic for the reduced data are sufficiently close to the original data, especially for the 4-dimensional reduced data, such that PCA may be considered as a resonant helper for the Yeast dataset. Moreover, the statistics suggest that an SGT2 fuzzy MF may be the most suitable for representing the reduced data, hence justifying the validity of our hypothesis.

Let us now consider the Madelon data [34]. It is an artificial dataset which was part of the NIPS 2003 feature selection challenge. This is a two-class classification problem consisting of 4400 instances and 500 features. We applied dimensionality reduction on this dataset using the 4 methods described earlier, and the data was reduced to 3 and 4 dimensions. These were then given to a WMBS module and the results obtained are summarized in Table II.

From the table, we may observe that different fuzzy MFs are adjudged most suitable for representing the 3 and 4-dimensional reduced Madelon data, when reduction is done using PCA and kernel PCA. Consequently, these methods are not resonant with the WMBS algorithm for the Madelon dataset. Further, t-SNE only reduces a data to 2 and 3 dimensions for visualization. Therefore, we may conclude that probabilistic PCA is a resonant helper for the given dataset, and hence it may be used in conjunction with the WMBS algorithm without loss in accuracy.

V. Conclusion and Future Work

In this paper, we proposed a deterministic method for selecting one of several dimensionality reduction techniques to be used in conjunction with the WMBS algorithm for datasets consisting of large number of features. In addition

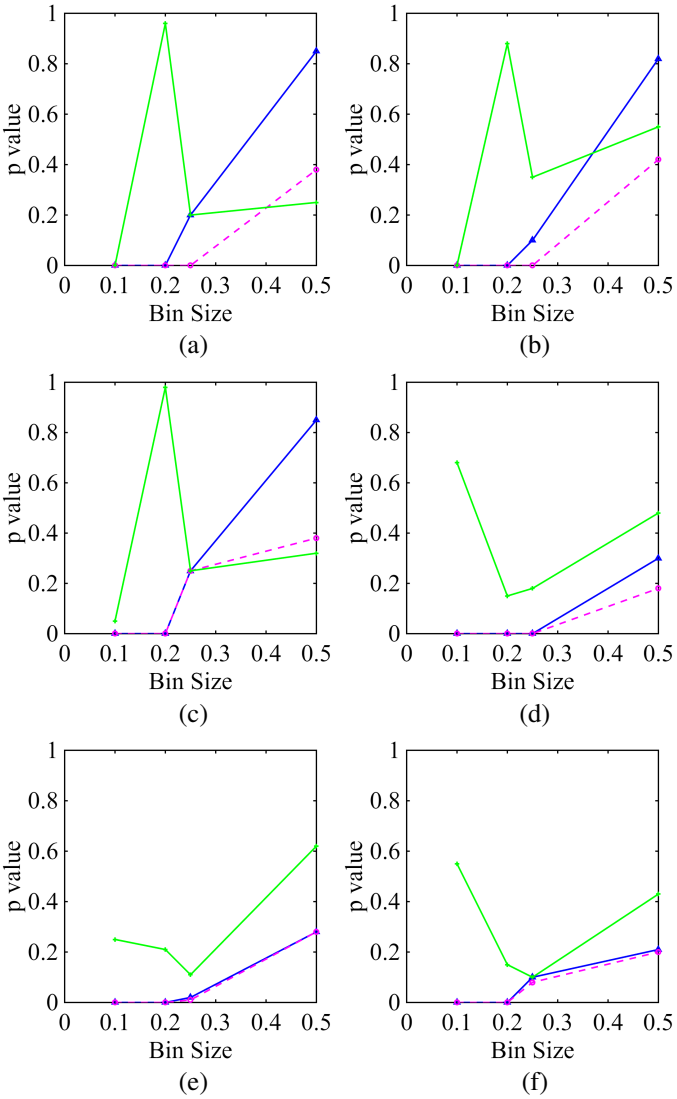

Fig. 4: Comparison of p-statistic obtained using Wilcoxon's tests on the 8-D (blue) Yeast dataset with the reduced 3-D (purple) and 4-D (green) data. Signed-rank test statistics for (a) T1 fuzzy, (b) IT2 fuzzy, and (c) SGT2 fuzzy MFs; and Rank-sum test statistics for (d) T1 fuzzy, (e) IT2 fuzzy, and (f) SGT2 fuzzy MFs.

to establishing a framework for such an analysis, we also showed the validity of our hypothesis on the 8-D Yeast dataset. Without providing a rigorous proof, the intuition behind the method was evident from discussions and experiments. From results obtained, it was found that different techniques may be suitable to reduce different datasets, since PCA and probabilistic PCA performed well on the Yeast and Madelon datasets, respectively.

Since the t-SNE algorithm has recently shown immense promise in dimensionality reduction, we believe it may be suitable for visualizing large datasets in order to estimate the choice of fuzzy MF to represent the data. For instance, if the 2-D or 3-D scatter obtained with t-SNE consists of several ellipsoidic clusters, a Gaussian T1 or SGT2 fuzzy MF may be considered most suitable for the representation. Further, in case of a randomly distributed high-dimensional dataset, an

TABLE II: Wilcoxon test statistics for Madelon data vs data sets generated from T1, IT2 and SGT2 Fuzzy MFs

DR Method	Type	3 Dimensions		4 Dimensions	
		Bin Size	p-WSR / p-WRS	Bin Size	p-WSR / p-WRS
PCA	T1	0.1	0.9418 / 0.2886	0.5	0.9588 / 0.9358
	IT2	0.05	0.2562 / 0.8834	0.5	0.8767 / 0.8850
	SGT2	0.05	0.1801 / 0.7375	0.5	0.8767 / 0.8951
k-PCA	T1	0.1	0.8749 / 0.2377	0.5	0.3314 / 0.2702
	IT2	0.05	0.6112 / 0.7714	0.5	0.0645 / 0.2222
	SGT2	0.05	0.1486 / 0.3162	0.5	0.2695 / 0.1836
p-PCA	T1	0.1	0.7672 / 0.2073	0.5	1 / 0.8065
	IT2	0.2	0.0676 / 0.5904	0.5	0.8361 / 0.6376
	SGT2	0.1	0.5488 / 0.1643	0.5	1 / 0.7774
t-SNE	T1	0.2	0.1634 / 0.7829	-	-
	IT2	0.2	0.3062 / 0.7372	-	-
	SGT2	0.2	0.1563 / 0.7842	-	-

ensemble of several techniques may be applied to obtain a better low-dimensional representation of the data than what is obtained from a single approach. An analysis of the results obtained from the different module may then be used to assign weights to the various components in the ensemble. This is currently under investigation.

ACKNOWLEDGMENT

The authors would like to thank Desh Raj, Aditya Gupta, Kenil Tanna and Bhuvnesh Garg for their valuable assistance in the development and preparation of this manuscript.

REFERENCES

[1] L. Jimenez and D. Landgrebe, "Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data," *IEEE Trans. Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 28, no. 1, pp. 39–54, 1998.

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

[3] J. Mendel, *Uncertain rule-based fuzzy logic systems: introduction and new directions*. Prentice Hall PTR Upper Saddle River, 2001.

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

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

[6] Q. Liang and J. Mendel, "Interval type-2 fuzzy logic systems: theory and design," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 5, pp. 535–550, 2000.

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

[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] 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.

[10] 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.

[11] 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.

[12] 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.

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

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

[15] V. Pham, L. Ngo, and W. Pedrycz, "Interval-valued fuzzy set approach to fuzzy co-clustering for data classification," *Knowledge-Based Systems*, 2016.

[16] C. Peraza, F. Valdez, and O. Castillo, "Interval type-2 fuzzy logic for dynamic parameter adaptation in the harmony search algorithm," in *Proc. IEEE Int. Conf. Intelligent Syst.*, 2016, pp. 106–112.

[17] 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.

[18] D. Raj, A. Gupta, B. Garg, K. Tanna, and F. C.-H. Rhee, "Analysis of data generated from multidimensional type-1 and type-2 fuzzy membership functions," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 681–693, 2017.

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

[20] 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.

[21] 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.

[22] W. Wang and M. A. Carreira-Perpinan, "The role of dimensionality reduction in classification," in *AAAI*, 2014, pp. 2128–2134.

[23] K. Pearson, "On lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.

[24] C. Spearman, "General intelligence: Objectively determined and measured," *The American Journal of Psychology*, vol. 15, no. 2, pp. 201–292, 1904.

[25] W. Torgerson, "Multidimensional scaling: 1. theory and method," *Psychometrika*, vol. 17, no. 4, pp. 401–419, 1952.

[26] L. Saul, K. Weinberger, J. Ham, F. Sha, and D. Lee, "Spectral methods for dimensionality reduction," *Semisupervised learning*, pp. 293–308, 2006.

[27] J. Lee and M. Verleysen, *Nonlinear dimensionality reduction*. Springer Science & Business Media, 2007.

[28] L. Van Der Maaten, E. Postma, and J. Van den Herik, "Dimensionality reduction: a comparative," *J Mach Learn Res*, vol. 10, pp. 66–71, 2009.

[29] B. Schölkopf, A. Smola, and K. Müller, "Kernel principal component analysis," in *International Conference on Artificial Neural Networks*. Springer, 1997, pp. 583–588.

[30] M. Tipping and C. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, no. 3, pp. 611–622, 1999.

[31] L. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[32] S. Kullback and R. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

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

[34] 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.