On A Type-2 Fuzzy Clustering Algorithm

Leehter Yao and Kuei-Sung Weng
Dept. of Electrical Engineering
National Taipei University of Technology
Taipei, Taiwan
e-mail: ltyao@ntut.edu.tw; gsweng@mail.nihs.tp.edu.tw

***Abstract*—A Type-2 fuzzy clustering algoritm that integreates Type-2 fuzzy sets with Gustafson-Kessel algorithm is proposed in this paper. The proposed Type-2 Gustafson-Kessel algorithm (T2GKA) is essentially a combination of probabilistic and possibilistic clustering schemes. It will be shown that the T2GKA is less susceptive to noise than the Type-1 GKA. The T2GKA ignores the inlier and outlier interruptions. The clustering results show the robustness of the proposed T2GKA since a reasonable amount of noise data does not affect its clustering performance. A drawback of the conventional GKA is that it can only find clusters of approximately equal volume. To overcome this difficulty, this work uses an algorithm called The Directed Evaluation Ellipsoid Cluster Volume (DEECV) to effectively evaluate the proper ellipsoid volume. The proposed T2GKA is essentially a DEECV based learning algorithm integrated with T2GKA. The experimental results show that the T2GKA can learn suitable sized cluster volume along with a varying dataset structure volume.**

Keywords-ellipsoids; probabilistic; possibilistic; fuzzy c-means; Gustafson-Kessel algorithm; Type-2 fuzzy clustering

I. INTRODUCTION

Clustering shows powerful capabilities to determine a finite number of clusters for partitioning a dataset. Hruschka et al. [1] proposed a survey of evolutionary algorithms for clustering, we can see the clustering area profile by focusing more on those topics that have received more importance in the literature. Based on the partition-based concepts, the fuzzy clustering algorithm can be classified into probabilistic fuzzy clustering and possibilistic fuzzy clustering. The fuzzy c-means (FCM) algorithm proposed by Bezdek [2] is a widely used and efficient clustering method for clustering and classification. Because FCM employs the Euclidean norm to measure dissimilarity, it inherently imposes a spheroid onto the clusters regardless of the actual data distribution. In [3] and [4], Gustafson and Kessel proposed the G-K algorithm (GKA) using an adaptive distance norm based on the cluster center and data point covariance matrices to measure dissimilarity. Because the distance norm employed in the GKA is in the Mahalanobis norm form, GKA can be considered as utilizing ellipsoids to cluster prototype data points. However, GKA assumes fixed ellipsoid volumes before iteratively calculating the cluster centers.

FCM and GKA are probabilistic fuzzy clustering approaches. In a noise environment, the probabilistic fuzzy clustering will force noise to belong to one or more clusters, therefore seriously influencing the main dataset structure. To relieve the probabilistic clustering drawbacks, Krishnapuram and Keller proposed a possibilistic fuzzy clustering called the Possibilistic c-means (PCM) [5-6]. The possibilistic fuzzy clustering can evaluate a datum to a cluster depending only on the distance of the datum to that cluster, but not on its distance to other clusters. The possibilistic fuzzy clustering can alleviate the noise influence, but it is very sensitive to initialization, sometimes generating coincident clusters.

To avoid the various FCM and PCM problems, Pal et al. proposed a new model called the possibilistic fuzzy c-means (PFCM) model [7]. The PFCM is a hybridization of the PCM and FCM models. The PFCM solves the noise sensitivity defect of FCM and overcomes the coincident clusters problem of PCM. However, the PFCM model has four parameters that must be learned. For an uncertain environment how to search for the best four parameters is difficult. All aforementioned fuzzy clustering methods have membership values called Type-1 membership values. In a real application domain, the prototype data may have many uncertain factors. Owing to the Type-1 fuzzy sets, their membership functions are crisp and they cannot directly model the uncertainties. On the other hand, the Type-2 membership functions are fuzzy, and they can appropriately model the uncertainties.

The Type-2 fuzzy set concept was introduced by Zadeh [8]. The advances of the Type-2 fuzzy sets and systems [9] are largely attributed to their three-dimensional membership function to handle more uncertainties in real application problems. Recent researches [10-13] have shown that the uncertainty in fuzzy systems can be captured with Type-2 fuzzy sets. In [14], the interval Type-2 fuzzy set was incorporated into the FCM to observe the effect of managing uncertainty from the two fuzzifiers. Type-2 fuzzy sets have been used to manage the uncertainties in various domains where the performance of Type-1 fuzzy sets is not satisfactory. For instance, [15-17] used the Type-2 fuzzy set for handling uncertainty in pattern recognition. Zarandi et al.

[18] presented a systematic Type-2 fuzzy expert system for diagnosing human brain tumors.

When clustering methods are combined with Type-2 fuzzy sets the prototype data can be clustered more properly and accurately. We extend the Type-1 membership values to Type-2 by assigning a possibilistic-membership function to each Type-1 membership value. The possibility theory, introduced by Zadeh [19] appears as a mathematical counterpart of probability theory that deals with uncertainty using fuzzy sets. The Type-2 membership values are obtained by taking the difference between each Type-2 membership function area with the corresponding Type-1 membership value. In this paper we use the unbounded normal distributions Gaussian function as the secondary membership function [20-21].

Using the aforementioned cencepts we combined probabilistic and possibilistic methods to build Type-2 fuzzy sets. We present a Type-2 GKA (T2GKA) that is an extension of the conventional GKA. The membership values for each prototype datum are extended as Type-2 fuzzy memberships by assigning a membership grade to the Type-1 memberships. The higher the membership value for a prototype datum, the larger the prototype datum contribution possesses in determining the cluster center location. The experimental results show that the T2GKA was less susceptible to noise than the Type-1 GKA.

To overcome the T2GKA's inability to determine appropriate ellipsoid size, a Directed Evaluation Ellipsoid Cluster Volume (DEECV) scheme is proposed in this paper, so that the proper cluster volume can be directly evaluated instead of each cluster using equal cluster volume in the clustering learning. The Mahalanobis norm inducing matrix determinant is utilized in this paper to measure the ellipsoid size [22, 23]. The DEECV is developed to intelligently estimate the proper ellipsoid size value. With the proper ellipsoid size value determined by the proposed DEECV, the learning efficiency can be further improved. The proposed T2GKA is essentially a DEECV based learning algorithm integrated with T2GKA.

II. COMBINED PROBABILISTIC AND POSSIBILISTIC TO BUILD TYPE-2 FUZZY SET

We focus on providing a Type-2 fuzzy set model to avoid uncertain outliers affecting the clustering learning results. We explain how to build the Type-2 fuzzy sets based on the following concept. For every prototype data point, the ordered set of memberships to each of the clusters $\{\mu_1,\ldots,\mu_c\}$ spans a c-dimensional space. Sets of specific membership values in this space are represented as points. The possibility distribution transform of the Type-1 probability distribution on unbounded normal distributions Gaussian function around the Type-1 membership value. For each given point, the possibilistic type membership value indicates the strength of the attribution to any cluster independent from the rest. Figure 1 shows that two points x_1 and x_2 have the same Type-1 membership value but have different possibility values.

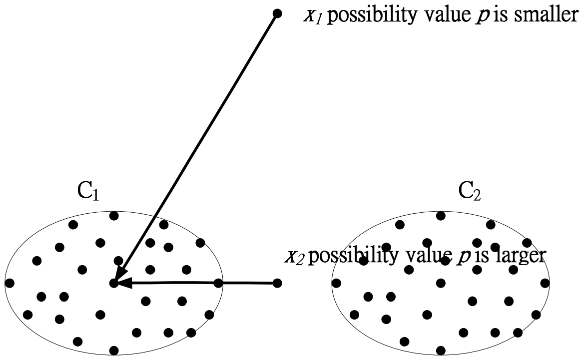

Figure 1. The points have the same membership value but have different possibility values

The idea in building Type-2 fuzzy sets is based simply on the fact that, for the same Type-1 membership value, the secondary membership function should make the larger possibility value more than the smaller possibility value. The secondary membership function based on the competitive learning theory proposed here originates from the rival-penalized competitive learning (RPCL) in [24]. The basic idea of RPCL is that, for each input, the winner unit is modified to adapt to the input and its rival is delearned using a smaller learning rate, so, RPCL rewards the winner and punishes the rival. A Type-2 fuzzy set is defined as an object $\widetilde{A}$ which has the following form:

$$\widetilde{A} \equiv \left\{ \left\langle u, t, \xi_A(\bullet) \right\rangle \right\}, \tag{1}$$

where $\xi_A(\bullet)$ is an unbounded normal distributions Gaussian function representing the secondary membership function of the element $(u,t), u \in \boldsymbol{U}, \xi_A(\bullet) \in [0,1]$ in $\widetilde{A}$. We set the Type-1 membership value and Type-2 membership value relation as following equations:

$$u = u \times \max(\xi_A(\bullet)), \tag{2}$$

$$t = u \times \xi_A(\bullet), \tag{3}$$

where u represents the primary membership value and t represents the Type-2 membership value. The $\xi_A(\bullet)$ is an unbounded normal distribution Gaussian function representing the secondary membership function:

$$\xi_A(\bullet) = \exp - \frac{1}{2}\left(\frac{a-b}{\sigma}\right)^2. \tag{4}$$

Under the aforementioned concepts, reducing the Type-2 fuzzy sets involves complicated operations. We use the input/output data points $\boldsymbol{x_k}$, $k = 1 \ldots N$, set as the possibility value, p_{ik} as the unbounded normal distribution Gaussian function standard deviation, σ and the $(p_{ik}-1)$ denotes the distance between p_{ik} to the central unbounded normal distribution Gaussian function, then design the secondary membership function $e^{-0.5\times(\frac{p_{ik}-1}{p_{ik}})^2}$.

The confidence intervals for varying possibilistic values p_{ik} built around the same prototype datum x_{ik} with

 ISBN: 978-1-61208-221-9

membership value μ_{ik} are nested. A unimodal numerical possibility distribution may also be viewed as a nested set of confidence intervals. The unbounded normal distribution Gaussian function's confidence intervals are 2σ and have a 95% confidence level. For example, the Type-1 membership value $\mu = 0.5$, has a secondary membership function with different possibility values as shown in Fig. 2.

The Type-2 membership values can be obtained using the following equation;

$$t_{ik} = \mu_{ik} \times \xi_A \Rightarrow t_{ik} = \mu_{ik} \times e^{-\frac{1}{2}\times\left(\frac{p_{ik}-1}{p_{ik}}\right)^2}, \quad (5)$$

where t_{ik} (μ_{ik}) denotes the Type-2(1) memberships, p_{ik} denotes the membership degrees for one datum resembling the possibility of its being a member of the corresponding cluster. For example, for the Type-1 membership value $\mu = 0.5$, the following evaluations process interprets that Type-2 fuzzy sets evaluate their secondary membership values with different possibility values. The prototype data points $\boldsymbol{x_k}$, $k = 1,\ldots,N$, have Type-1 membership value $\mu_{ik} = 0.5$ and possibility value $p_{ik} = 1.0$ then the Type-2 membership values $t_{ik} = 0.5$ are obtained using (5). For the same Type-1 membership value $\mu_{ik} = 0.5$, and possibility value $p_{ik} = 0.1$ we obtain the Type-2 membership values as $t_{ik} = 1.2884e - 018 \cong 0$.

We know that in our design the secondary membership function, for the same Type-1 membership value, a larger possibility value can make the Type-1 membership value larger than the smaller possibility value does. Using the aforementioned concepts, we combined the probability and possibility membership values and propose the Type-2 Gustafson-Kessel Algorithm (T2GKA).

III. THE TYPE-2 G-K ALGORITHM (T2GKA):

To overcome the drawback of the GK algorithm, it is used to find only clusters of approximately equal volumes. In this paper an algorithm called The Directed Evaluation Ellipsoid Cluster Volume (DEECV) is proposed to effectively evaluate the proper ellipsoid volume. The proposed T2GKA is essentially a DEECV based learning algorithm integrated with T2GKA.

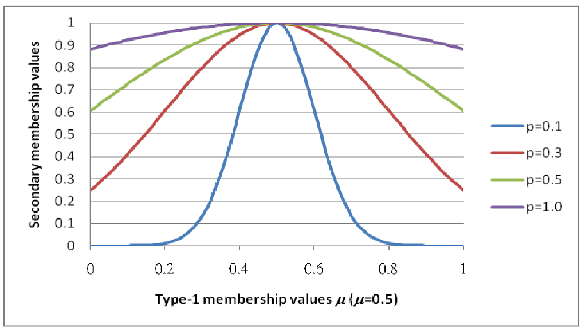

Figure 2. The secondary membership function with the different possibility values

A. *The Type-2 G-K Algorithm (T2GKA):*

Based on the prototype data points $\boldsymbol{x_k}$, $k = 1,\ldots,N$, given the random initial Type-1 fuzzy partition matrix $\boldsymbol{U}^{(0)} = \boldsymbol{T}^{(0)}$, T2GKA is to learn the Type-2 fuzzy partition matrix $\boldsymbol{T}$, the coordinates of all cluster centers $\boldsymbol{V}$ and the norm inducing matrix $\boldsymbol{A_i}$ by minimizing , $i = 1,\ldots, c$

$$J_{T2GKA}(\boldsymbol{T},\boldsymbol{V},\boldsymbol{A}) = \sum_{i=1}^{c}\sum_{k=1}^{N}(t_{ik})^m \boldsymbol{D}^2_{\boldsymbol{ikA_i}} + \sum_{i=1}^{c}\omega_i(|\boldsymbol{A}_i| - \rho_i) + \sum_{k=1}^{N}\gamma_k(\sum_{i=1}^{c}t_{ik} - 1), \quad (6)$$

where t_{ik} has the same meaning of membership and constraints as FCM. The distance between the *k-th* prototype data point and the *i-th* cluster center is defined as the Mahalanobis norm:

$$\boldsymbol{D}_{\boldsymbol{ikA_i}} = ((\boldsymbol{x_k} - \boldsymbol{v_i})^T \boldsymbol{A_i}(\boldsymbol{x_k} - \boldsymbol{v_i}))^{1/2}. \quad (7)$$

For the *i-th* cluster, the ellipsoid $\phi_i(\cdot)$ is defined as

$$\phi_i(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{v_i})^T \boldsymbol{A_i}(\boldsymbol{x} - \boldsymbol{v_i}) = 1,\ i = 1,\ldots, c. \quad (8)$$

Since the volume of $\phi_i(\cdot)$ is inversely proportional to the determinant of $\boldsymbol{A_i}$, det($\boldsymbol{A_i}$) is thus utilized as a measure of the ellipsoid volume for T2GKA. If the determinant of $\boldsymbol{A_i}$ is given as ρ_i, $\boldsymbol{A_i}$ is constrained by

$$\det(\boldsymbol{A_i}) = \rho_i,\ \rho_i > 0,\ i = 1,\ldots, c. \quad (9)$$

The optimization in ($\boldsymbol{T},\boldsymbol{V},\boldsymbol{A}$) can be solved using differentiations as follows:

$$\boldsymbol{A}_i = \left[\rho_i \det(\boldsymbol{F}_i)\right]^{1/n} \boldsymbol{F}_i^{-1} \quad i = 1,\ldots,c, \quad (10)$$

$$\boldsymbol{F}_i = \frac{\sum_{k=1}^{N}(t_{ik})^m (x_k - v_i)(x_k - v_i)^T}{\sum_{k=1}^{N}(t_{ik})^m}. \quad (11)$$

To avoid the covariance matrix being singular in the iterative process, a scaled identity matrix is added to the covariance matrix, i.e.,

$$\boldsymbol{F}_i = (1-\kappa)\boldsymbol{F}_i + \kappa \det(\boldsymbol{F_0})^{1/n}\boldsymbol{I}, \quad (12)$$

where $\kappa \in [0,1]$ is a tuning factor with a small value and $\boldsymbol{F_0}$ is the whole data set covariance matrix with fixed value. The coordinate of each cluster center as well as the membership element in the partition matrix can be updated using the following equations:

$$\boldsymbol{v_i} = \frac{\sum_{k=1}^{N}(t_{ik})^m \boldsymbol{x_k}}{\sum_{k=1}^{N}(t_{ik})^m}, \quad (13)$$

$$t_{ik} = \left(\sum_{j=1}^{c}\left(\frac{\boldsymbol{D}_{\boldsymbol{ikA_i}}}{\boldsymbol{D}_{\boldsymbol{jkA_i}}}\right)^{2/(m-1)}\right)^{-1}, 1 \le i \le c; 1 \le k \le N. \quad (14)$$

 ISBN: 978-1-61208-221-9

For each given point, the possibilistic type membership value, indicating the strength of the attribution to any cluster, is independent from the rest. We calculate the possibilistic type membership value simultaneously using

$$p_{ik} = \frac{1}{1+\left(\frac{\boldsymbol{D}^2_{ikA_i}}{\varepsilon_i}\right)^{\frac{1}{(m-1)}}}. \tag{15}$$

We determine the reasonable number of ε_i by computing

$$\varepsilon_i = K\frac{\sum_{k=1}^{N} t_{ik}^m \boldsymbol{D}^2_{ikA_i}}{\sum_{k=1}^{N} t_{ik}^m}, \tag{16}$$

usually K=1 is chosen. For each given point, using the possibilistic type membership value, the Type-2 membership values can be updated using equation (5).

B. *The Directed Evaluation Ellipsoid Cluster Volume (DEECV)*

Without knowing the prototype data point distribution range a priori, a tentative value ρ_a is first assigned to every parameter, ρ_i, $i = 1,\ldots, c$. With $\rho_i = \rho_a$, $i = 1,\ldots, c$, T2GKA is applied to calculate the tentative ellipsoid $\hat{\phi}_i$ with center $\hat{\boldsymbol{v}}_i$, the covariance matrix $\hat{\boldsymbol{F}}_i$, and the norm inducing matrix $\hat{\boldsymbol{A}}_i$, $i = 1,\ldots, c$. Denote B_i as the set of data points belonging to the cluster corresponding to $\hat{\phi}_i$ and $\boldsymbol{x}_j^i$ as the *j-th* data point belonging to B_i. Let $\hat{\boldsymbol{x}}^i$ be the data point with the largest Mahalanobis distance $\hat{L}_i$ among all data points in B_i, i.e.

$$\hat{\boldsymbol{x}}^i = \underset{\boldsymbol{x}_j^i \in B_i}{Argmax}(\|\boldsymbol{x}_j^i - \hat{\boldsymbol{v}}_i\|_{\hat{A}_i}) \tag{17}$$

and $$\hat{L}_i = \underset{\boldsymbol{x}_j^i \in B_i}{max}(\|\boldsymbol{x}_j^i - \hat{\boldsymbol{v}}_i\|_{\hat{A}_i}), \tag{18}$$

where $\|\bullet\|_{\hat{A}_i}$ denotes the Mahalanobis norm with the norm inducing matrix $\hat{\boldsymbol{A}}_i$ as in (7). According to (7) and (10),

$$(\hat{\boldsymbol{x}}^i - \hat{\boldsymbol{v}}_i)^T(\rho_a \det(\hat{\boldsymbol{F}}_i))^{1/n}\hat{\boldsymbol{F}}_i^{-1}(\hat{\boldsymbol{x}}^i - \hat{\boldsymbol{v}}_i) = \hat{L}_i. \tag{19}$$

It is thus obvious that if the initialization process appropriately adjusts the initial ellipsoid volumes so that the farthest data point $\hat{\boldsymbol{x}}^i$ with the largest Mahalanobis norm is right on the initialized ellipsoid, all of the ellipsoid volumes will be scaled to the range of solutions. As shown in (8), the data points on the ellipsoids have a Mahalanobis distance of 1. Divide $\hat{L}_i$ at both sides of (19),

$$(\hat{\boldsymbol{x}}^i - \hat{\boldsymbol{v}}_i)^T(\frac{\rho_a}{\hat{L}_i^n}\det(\hat{\boldsymbol{F}}_i))^{1/n}\hat{\boldsymbol{F}}_i^{-1}(\hat{\boldsymbol{x}}^i - \hat{\boldsymbol{v}}_i) = 1. \tag{20}$$

Therefore, the appropriate initial volume for the *i-th* ellipsoid leading to the result that all data points are included by the ellipsoid with tentative value ρ_a can thus be defined as:

$$\rho_{i\_initial} = \frac{\rho_a}{\hat{L}_i^n}, \text{ i = 1,…, c.} \tag{21}$$

It is worth noting that if $\hat{x}^i$ is an outlier for the cluster corresponding to $\hat{\phi}_i$, $\hat{L}_i$ will be unreasonably large. This results in an inaccurate initial ellipsoid volume ρ_i according to (21). For the data points with too much noise, an outlier detection scheme is required to determine the outliers and filter them out before applying the directed initialization. Let $\bar{d}_i$ be the average Mahalanobis distance among all data points belonging to B_i, then

$$\bar{d}_i = \frac{\sum_{j=1}^{|B_i|}\|\boldsymbol{x}_j^i - \hat{\boldsymbol{v}}_i\|_{\hat{A}_i}}{|B_i|}, \tag{22}$$

where $|B_i|$ denotes the number of data points in B_i. For all data points in B_i, the farthest data point and its maximum Mahalanobis distance can be respectively determined using (17) and (18). Removing the outliers affects the clustering learning results. With a predetermined threshold γ, any data point $\boldsymbol{x}^i$ belonging to the *i-th* cluster and its possibility membership value P_{ik} is larger than a predetermined threshold possibility membership value $P_{ik} \geq \alpha$ (in this paper, we set $\alpha = 0.1$), satisfies the following criterion:

$$\frac{\|\boldsymbol{x}^i - \hat{\boldsymbol{v}}_i\|_{\hat{A}_i}}{\bar{d}_i} \geq \gamma \tag{23}$$

is considered as an outlier and can be removed from B_i. The outlier detection scheme, as shown in (22) and (23), is recursively applied to every cluster of data points until no outlier has been detected. After filtering out the outliers in every cluster, the accuracy of calculating proper ellipsoid volume according to (21) for T2GKA's directed evaluation can be greatly improved.

IV. COMPUTER SIMULATIONS

We used the following computational conditions for all datasets: 1. The termination tolerance $\varepsilon = 0.000001$, the $\boldsymbol{D}_{ikA_i}$ for the FCM, FCMPCM, and PFCM is the Euclidean norm. 2. The $\boldsymbol{D}_{ikA_i}$ for the GK and T2GKA is the Mahalanobis norm. 3. The number of c clusters c is 7 for 7cluster. 4. c is 5 for 5 same-circle and sinusoidal sets. 5. c is *2* for all other datasets.

Example 1: The artificial 2-dimensional datasets X_{400} and X_{550} are designed. The X_{400} is a mixture of two 2-variate normal distributions with mean vectors $\begin{pmatrix}5.0\\6.0\end{pmatrix}$ and $\begin{pmatrix}5.0\\12.0\end{pmatrix}$. Each cluster has 200 points, while X_{550} is an augmented version of X400 with an additional 150 points uniformly distributed over $[0,15]\times[0,11]$. For data set X_{400} the clustering results in Table I show that the terminal centroids learned by all five algorithms produce good centroids.

When we cluster dataset X_{550}, we hope that the 150 noise points can be ignored and the cluster center will be found closer to the true centroids $\boldsymbol{V}_{true}$. From Table I, we can see that all five algorithms clustered the dataset X_{550} terminal centroids. Because PCM is very sensitive to initialization and it sometimes generates coincident clusters, we utilized the FCM clustering results to initialize PCM. The other four clustering methods ran the algorithm directly. To make a rough assessment of how each method accounted for inliers and outliers, we estimated $E_A = \|\boldsymbol{V}_{true} - \boldsymbol{V}_A\|^2$, where A denotes FCM, FCMPCM, PFCM, GK, and T2GKA. The E_{FCM}=0.4173, E_{FCMPCM}=0.0001, E_{PFCM}=0.3714 (a=1, b=0.1, m=2, η=2), E_{PFCM}=0.1699 (a=1, b=1, m=2, η=2), E_{GKA}=0.4825, and E_{T2GKA}=0.0066. The T2GKA clustering results with the proper cluster volumes for the datasets X_{550} are shown in Fig. 3. We compared the five clustering method's E_A values, the E_{FCMPCM} value is smaller than that in other methods, but its membership values are independent of the other clusters. We cannot depend on the membership values to classify the data points belonging to which cluster. Except for the E_{FCMPCM}, the E_{T2GKA} value is smaller than that in other methods. The clustering results show the robustness of the proposed T2GKA because a reasonable amount of noise data does not affect its clustering performance.

Example 2: To verify that the proposed method can accord the prototype dataset structure to learn the proper cluster centers, 5 same-circles were designed with each cluster containing 300 prototype data points. The dataset 5 same-circle is a mixture of two 2-variate normal distributions with mean vectors $\begin{pmatrix}0.0\\3.0\end{pmatrix}, \begin{pmatrix}5.0\\3.0\end{pmatrix}, \begin{pmatrix}0.0\\-3.0\end{pmatrix}, \begin{pmatrix}5.0\\-3.0\end{pmatrix}$, and $\begin{pmatrix}2.5\\0.0\end{pmatrix}$.

The T2GKA clustered results with the proper clusters centers for the 5 same-circle datasets are shown in Fig. 4. For the 5 same-circle datasets, the E_{FCM}=0.0042, E_{FCMPCM}=0.0003, E_{PFCM}=0.0039 (a=1, b=0.1, m=2, η=2), E_{PFCM}=12.2009 (a=1, b=1, m=2, η=2), E_{GKA}=0.0036, and E_{T2GKA}=0.0026. We compared the five clustering method's E_A values. Except for the E_{FCMPCM}, the E_{T2GKA} value is smaller than that in other methods. The clustering results show the robustness of the proposed T2GKA because a reasonable amount of noise data does not affect its clustering performance.

Example 3: To verify that the proposed method can accord the prototype dataset structure to learn the proper cluster volumes, 2 artificial datasets named 7cluster and sinusoidal were designed. There are 700 and 200 prototype data points in the 7cluster and sinusoidal datasets, respectively. There are 700 prototype data points in the 7cluster datasets clustered into 7 clusters with different sizes and orientations. Each cluster contains 100 prototype data points. The 7cluster dataset is a mixture of two 2-variate distributions with varying deviation, its mean vectors are $\begin{pmatrix}5.0\\1.0\end{pmatrix}, \begin{pmatrix}1.0\\5.0\end{pmatrix}, \begin{pmatrix}1.0\\1.0\end{pmatrix}, \begin{pmatrix}5.0\\5.0\end{pmatrix}, \begin{pmatrix}2.0\\-2.0\end{pmatrix}, \begin{pmatrix}-2.0\\2.0\end{pmatrix}$, and $\begin{pmatrix}4.5\\3.0\end{pmatrix}$.

The prototype data points in the dataset sinusoidal are generated by $x_2 = 10^{-4}\, sin(0.001x_1^2)x_1^3 + \varepsilon$, where $x_1 \in [0,100]$ and $\varepsilon \sim Normal(0,25)$ is a normally distributed random noise. The T2GKA clustered results with the proper clusters volumes for the 7cluster and sinusoidal datasets are shown in Figs. 5 and 6, respectively. The proposed T2GKA is essentially a DEECV based learning algorithm integrated with the T2GKA. The experimental results show that the T2GKA can learn suitable sized cluster volume along with dataset varying structure volume.

TABLE I. THE TERMINAL CENTROIDS LEARNED BY FCM, FCMPCM, PFCM, GK, AND T2GKA IN THE DATASETS X_{400} AND X_{550}, EXAMPLE 1

Clustering Algorithm	Data sets			
	X400 (centroid)		X550 (centroid)	
	x1	x2	x1	x2
FCM: m=2	4.9794	5.9531	5.5711	5.4143
	4.9407	12.0593	5.1885	11.6395
FCMPCM: η=2	5.0017	6.0094	5.0076	6.0091
	4.9973	12.0102	4.9968	12.0103
PFCM: a=1, b=1, m=2, η=2	4.9843	5.9746	5.3716	5.7308
	4.9566	12.0506	5.1281	11.6642
PFCM: a=1, b=0.1, m=2, η=2	4.9800	5.9558	5.5410	5.4604
	4.9427	12.0582	5.1804	11.6445
GKA: m=2	4.9782	5.9538	5.1064	5.4443
	4.9397	12.0568	5.5502	11.4151
T2GKA: m=2	5.0048	6.0239	5.0137	5.9593
	5.0097	12.0837	4.9743	12.1031

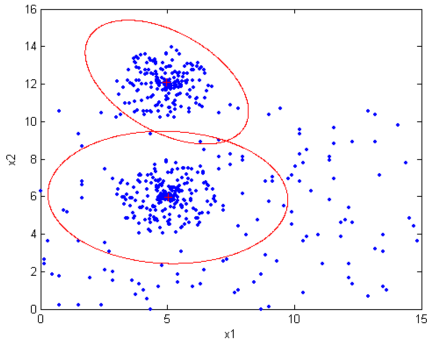

Figure 3. The T2GKA clustering results with the proper clusters volumes for the dataset X_{550}, Example 1

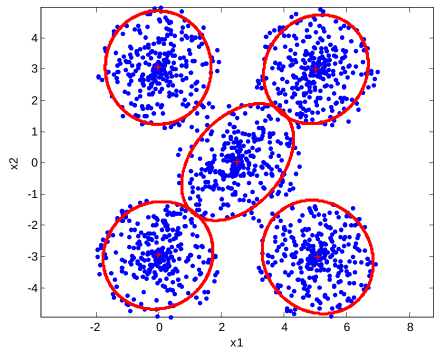

Figure 4. Clustering results using 5 ellipsoids for the prototype data points in the dataset 5samecircle, Example 2

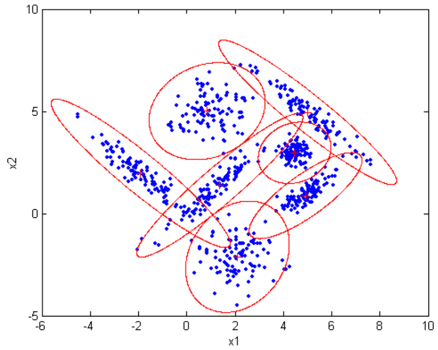

Figure 5. Clustering results using 7 ellipsoids for the prototype data points in the dataset 7cluster, Example 3

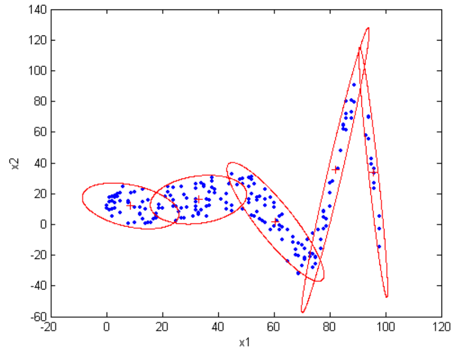

Figure 6. Clustering results using 5 ellipsoids for the prototype data points in the dataset sinusoidal, Example 3

V. CONCLUSIONS

This paper presented an efficient combined probabilistic and possibilistic method for building Type-2 fuzzy sets. Utilizing this concept we presented a Type-2 GKA (T2GKA) that is an extension of the conventional GKA. The experimental results showed that the T2GKA was less susceptible to noise than the Type-1 GKA. The clustering results showed the robustness of the proposed T2GKA because a reasonable amount of noise data does not affect its clustering performance.

The DEECV is proposed to effectively evaluate proper ellipsoid volume. The proposed T2GKA is essentially a DEECV-based learning algorithm integrated with T2GKA. The experimental results showed that the T2GKA can learn suitable sized clusters volume along with varying dataset structure volume.

REFERENCES

[1] E. R. Hruschka, R. J. G. B. Campello, A. A. Freitas and A. C. P. L. F. de Carvalho, “A Survey of Evolutionary Algorithm for Clustering,” IEEE Trans. Syst., Man, Cybern., pt. C, vol. 39, no. 2, pp.133-155, March 2009.

[2] J. Bezdek, Pattern Recognition with Fuzzy Objective Function, Plenum Press, New York, 1981.

[3] D. E. Gustafson and W. C. Kessel, “Fuzzy clustering with a fuzzy covariance matrix,” in Proc. IEEE Conf. Decision Contr., San Diego, CA, pp. 761-766, 1979.

[4] R. Babuška, Fuzzy modeling for control, Kluwer Academic Publishers: Massachusetts, 1998.

[5] R. Krishnapuram and J. Keller, “A possibilistic approach to clustering,” IEEE Trans. Fuzzy Sys., vol. 1, no. 2, pp. 98-110, May 1993.

[6] R. Krishnapuram and J. Keller, “The possibilistic c-Means algorithm: Insights and recommendations,” IEEE Trans. Fuzzy Sys., vol. 4, no. 3, pp. 385-393, August 1996.

[7] N. R. Pal, K. Pal, J. M. Keller and J. C. Bezdek, “A Possibilistic Fuzzy c-Means Clustering Algorithm,” IEEE Trans. Fuzzy Sys., vol. 13, no. 4, pp. 517-530, August 2005.

[8] L. A. Zadeh, “The concept of a linguistic variable and its application to approximate reasoning-I,” Inform. Sci., vol. 8, no. 3, pp. 199-249, 1975.

[9] J. Mendel, “Advances in Type-2 fuzzy sets and systems,” Inform. Sci., vol. 177, pp. 84-110, 2007.

[10] N. N. Karnik, J. M. Mendel and Q. Liang, “Type-2 fuzzy logic systems,” IEEE Trans. Fuzzy Sys., vol. 7, no. 6, pp. 643-658, December 1999.

[11] Q. Liang and J. M. Mendel, “Interval Type-2 fuzzy logic systems: theory and design,” IEEE Trans. Fuzzy Syst., vol. 8, no. 5, pp. 535-550, October 2000.

[12] J. M. Mendel, Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions, Upper Saddle River, NJ: Prentice-Hall, 2001.

[13] S. Coupland and R. John, “Geometric Type-1 and Type-2 fuzzy logic systems,” IEEE Trans. Fuzzy Sys., vol. 15, no. 1, pp. 3-15, February 2007.

[14] C. Hwang and F. C. H. Rhee, “Uncertain Fuzzy Clustering: Interval Type-2 Fuzzy Approach to C-Means,” IEEE Trans. Fuzzy Sys., vol. 15, no. 1, pp. 107-120, February 2007.

[15] H. B. Mitchell, “Pattern recognition using type-II fuzzy sets,” Inform. Sci., vol. 170, pp. 409-418, 2005.

[16] J. Zeng and Z. Q. Liu, “Type-2 Fuzzy Sets for Pattern Recognition: The State-of-the-Art,” Journal of Uncertain Systems, vol. 1, no. 3, pp. 163-177, 2007.

[17] J. Zeng, L. Xie and Z. Q. Liu, “Type-2 fuzzy Gaussian mixture models,” Pattern Recognition, vol. 41, pp. 3636-3643, 2008.

[18] M. H. Fazel Zarandi, M. Zarinbal and M. Izadi, “Systematic image processing for diagnosing brain tumors: A Type-II fuzzy expert system approach,” Applied Soft Computing, vol. 11, pp. 285-294, January 2011.

[19] L. A. Zadeh, “Fuzzy Sets as a Basis for a Theory of Possibility,” Fuzzy Sets and Systems, vol. 1, no. 1, pp. 3–28, 1978.

[20] D. Dubois, L. Foulloy, G. Mauris and H. Prade, “Probability-possibility transformations, triangular fuzzy sets and probabilistic inequalities,” Reliab. Comput., vol. 10, no. 4, pp. 273-297, 2004.

[21] G. Mauris, “Expression of Measurement Uncertainty in a Very Limited Knowledge Context: A Probability Theory-Based Approach,” IEEE Trans. Instr. Measu., vol. 56, no. 3, pp. 731-735, June 2007.

[22] L. Yao, “Nonparametric learning of decision regions via the genetic algorithm,” IEEE Trans. System, Man, and Cybernetics, vol. 26, no. 2, pp. 313-321, April 1996.

[23] L. Vandenberghe, S. Boyd and S. P. Wu, “determinant maximization with linear matrix inequality constraints,” J. SIAM, vol. 19, no. 2, pp. 499-533, 1998.

[24] L. Xu, A. Krzyak and E. Oja, “Rival penalized competitive learning for clustering analysis, RBF net, and curve detection,” IEEE Trans. Neural Netw., vol. 4, no. 4, pp. 636-649, July 1993.