Information Sciences 424 (2018) 69–90

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Uncertain fuzzy self-organization based clustering: interval type-2 fuzzy approach to adaptive resonance theory

Shakaiba Majeed[a], Aditya Gupta[b], Desh Raj[c], Frank Chung-Hoon Rhee[d,*]

[a] Department of Computer and Software, Hanyang University, Korea
[b] Department of Mathematics, IIT Guwahati, Guwahati - 781039, Assam, India
[c] Department of Computer Science and Engineering, IIT Guwahati, Guwahati - 781039, Assam, India
[d] Computational Vision and Fuzzy Systems Laboratory, Department of Electronic Engineering, Hanyang University, Korea

ARTICLE INFO

Article history:
Received 16 December 2016
Revised 24 September 2017
Accepted 27 September 2017
Available online 28 September 2017

Keywords:
Adaptive resonance theory
Fuzzy ART
Unsupervised learning
Interval type-2 fuzzy clustering
Vigilance parameter
Type reduction

ABSTRACT

Conventional unsupervised learning algorithms require knowledge of the desired number of clusters beforehand. Even if such knowledge is not required in advance, empirical selection of the parameter values may limit the adaptive capability of the algorithm, thereby restricting the clustering performance. An inherent uncertainty in the number and size of clusters requires integration of fuzzy sets into a clustering algorithm. In this paper, we propose a type-1 (T1) fuzzy ART method that adaptively selects the vigilance parameter value, which is critical in determining the network dynamics. This results in improved clustering performance due to the added flexibility in dynamic selection of the number of clusters with the use of fuzzy sets. To further manage the uncertainty associated with memberships, we extend the proposed T1 fuzzy ART with adaptive vigilance to an interval type-2 (IT2) fuzzy ART method. Type reduction and defuzzification are then performed using the KM algorithm to obtain a defuzzified vigilance parameter value. We evaluate our proposed methods on several data sets to validate their effectiveness.

1. Introduction

In most unsupervised clustering algorithms, the desired number of clusters may not be estimated efficiently by the algorithm itself, since this value is required as an input parameter, such as in K-means clustering, or computation of complex statistics must be performed for this decision, such as the CH-index [2] and gap statistic [45] for hierarchical clustering. For this reason, various methods have been proposed to achieve self-organizing clustering, such as the adaptive resonance theory (ART) [4].

ART networks have emerged as intelligent and autonomous learning algorithms due to the following properties: self-organization, self-stabilization, plasticity, and real-time learning, where numerous network variants have been proposed, such as fuzzy ART [6], ARTMAP [5], ARTMAP-pi, dART, dARTMAP [3], and falcon ART [25], to name a few.

Although ART-based methods may provide a satisfactory baseline solution to the problem of assigning dynamic number and size of clusters, their performance is still restricted due to the requirement of selecting an empirical vigilance parameter value. For this reason, integrating fuzzy sets with ART methods, as in fuzzy ART [6], have proved to be successful in imitating

* Corresponding author.
E-mail address: frhee@fuzzy.hanyang.ac.kr (F. C.-H. Rhee).

https://doi.org/10.1016/j.ins.2017.09.062

the ambiguity in the number and size of clusters. They have since been utilized in a number of applications in diverse fields such as gene expression analysis [46] and image processing [8].

In order to improve the clustering performance of conventional ART, we propose a method to define the vigilance parameter as a function of the relative distance between the current centers of the candidate cluster and the input vector. Such a definition is based on the radial nature of the clusters present in most real datasets. The proposed vigilance function can then be viewed as a type-1 (T1) fuzzy membership function, where for each incoming pattern a suitable vigilance value may be obtained.

For any pattern recognition application, it may not always be possible to extract perfect knowledge from a pattern set. This ambiguity may lead to an uncertain choice of parameters values to represent data that may result in the formation of inappropriate prototypes. To mitigate this problem, many researchers have applied type-2 (T2) fuzzy methods to several applications, for example, medical diagnosis [18], community transport scheduling [19], reactive robot navigation [14,26], chemical analysis [50], and survey processing [20], to name a few. Accordingly, the uncertainty associated with the vigilance parameter may be managed by extending it to a T2 fuzzy set (FS). In general, T2 FSs are capable of modelling various uncertainties that cannot be properly managed by T1 FSs, at the cost of added computational complexity.

To reduce this complexity, interval type-2 (IT2) FSs were proposed, since all secondary grades in these FSs are uniformly weighted (i.e., all equal to one). These MFs are also convenient for defuzzification, since the T1 fuzzy MF obtained after type reduction is an interval, which may be characterized by two fixed values. In light of this, we propose an IT2 fuzzy ART approach to further improve the performance of the proposed T1 fuzzy ART using the adaptive vigilance method, by incorporating the associated uncertainty of the data using upper and lower vigilance membership functions.

Our key contributions in this paper may be summarized as follows.

- We propose novel methods (T1 fuzzy ART with adaptive vigilance and IT2 fuzzy ART approach) for dynamic clustering that also takes into account the uncertainty associated with clusters in most real world applications.
- We model the vigilance parameter in ART clustering using T1 fuzzy and IT2 fuzzy MFs, and exploit the KM algorithm for defuzzification. Our method differs from conventional fuzzy ART in that the MF may be selected heuristically depending upon an estimate of the number and size of the clusters.
- We argue that the IT2 fuzzy approach may be suitable for extending any clustering algorithm that requires dynamic selection of clusters, due to its inherent ability to model uncertainties.

The remainder of this paper is organized as follows. We provide a brief review of the related research in Section 2. In Section 3, we discuss the conventional fuzzy ART algorithm and the effect of the vigilance parameter on the network dynamics. We also summarize the concepts of IT2 fuzzy sets, type reduction, and defuzzification. In Section 4, we describe our proposed methods, namely T1 fuzzy ART and IT2 fuzzy ART. Experimental results showing the effectiveness and improved performance of the algorithms are presented in Section 5. We conclude with a discussion of future work in Section 6.

2. Related research

Previously, complex fuzzy methods for adapting the vigilance parameter for ART-II network have been proposed for telecommunication signals [24]. However, the converging time for this method can be slow and requires specification of an empirical value of the initial vigilance that may affect the performance accordingly. In a previous work, evolutionary computation techniques were used to automate the selection of parameters for fuzzy ART [23]. However, no criterion has been specified to stop the genetic cycle of generation of parameters. An enhanced fuzzy ART network has also been proposed to dynamically control the vigilance parameter [22]. Unfortunately, such a method may only be used in a supervised learning environment.

In a recent work, the available domain knowledge has been incorporated into the clustering process to develop a knowledge-driven Mahalanobis distance-based ART algorithm [44]. Due to the knowledge-driven approach, the number of clusters may be determined automatically, which may effectively improve the clustering performance. Outside the domain of ART, there has been significant work on novel clustering algorithms. These include searching for structural consistency in data using proximity matrices [33], and using instance-level constraints for fuzzy clustering [12]. ART-based approaches have also been recently used for image indexing [41] and electric load balancing [1].

IT2 FSs have been successfully integrated into many well-known clustering algorithms such as fuzzy C-means (FCM) clustering [17,28,39], fuzzy probabilistic C-means [42], collaborative clustering [11], and fuzzy c-spherical shell clustering [16]. Furthermore, they have also been incorporated into co-clustering algorithms for color segmentation, multispectral image classification, and document categorization [37], in addition to time series prediction [7] and dynamic parameter adaptation for various optimization techniques [32,36]. There have also been attempts to model fuzzy neural networks using evolutionary algorithms [31,35], and to construct FCM cluster-based information granules for fuzzy modeling [34].

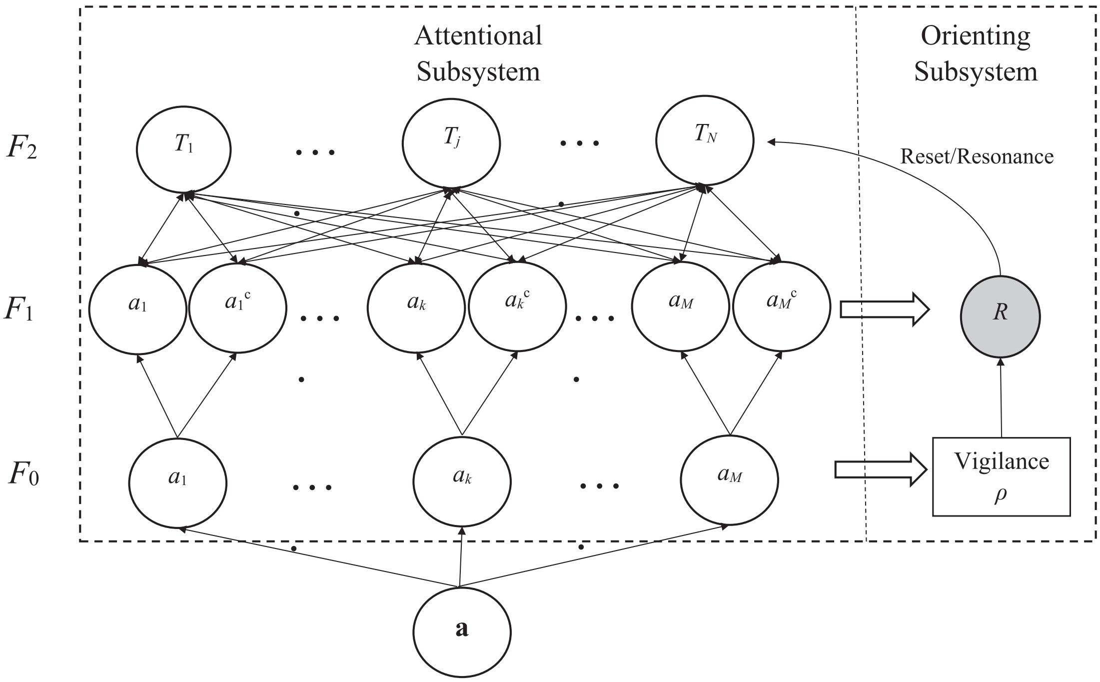


Fig. 1. Fuzzy ART architecture consisting of two sub-systems: attentional sub-system and orienting sub-system.

3. Background

3.1. Fuzzy ART algorithm

The fundamental feature of all ART systems is a matching process that results in either of two states: a resonant state that focuses attention and triggers stable prototype learning, or a self-regulating parallel memory search [6]. The architecture of the fuzzy ART network, shown in Fig. 1, consists of two subsystems: an attentional subsystem and an orienting subsystem.

1. The attentional subsystem consists of an input layer F_0, which receives an M-dimension input vector, a membership computing layer F_1, which calculates the degree of membership of an input pattern in each of the clusters, and an output layer F_2, consisting of N nodes, which represents the clusters generated by the network. The nodes in F_1 and F_2 are connected with bottom-up and top-down weights.
2. The orienting subsystem consists of a single node R, whose output is either a resonance or reset signal. The output signal affects the F_2 layer.

The algorithm is described in Algorithm 1 below. In this algorithm, $\mathbf{c}$ denotes the set consisting of all the cluster centers.

In fuzzy ART, three user-defined parameters, namely the choice parameter α, learning rate β, and vigilance parameter ρ determine the number and structure of the clusters generated. These parameters are kept constant, and chosen by the user empirically.

3.1.1. Choice parameter

The choice parameter α scales the degree of match between the input vector $\mathbf{I}$ and weight vector $\mathbf{w}_j$. In general, a higher value of α favors the creation of uncommitted nodes over the selection of a committed node.

3.1.2. Learning rate

The learning parameter β represents the degree to which the weight vector $\mathbf{w}_j$ will learn the characteristics of an input pattern being assigned to node j. Depending on the value of β, the network may employ fast learning, slow learning, or fast commit-slow recode properties.

3.1.3. Vigilance parameter

The vigilance parameter ρ determines the minimum amount of similarity that must be maintained within patterns that belong to the same category. If the match between the template of the current node and input pattern is less than the vigilance parameter, a reset signal is sent to the F_2 layer to disable the node and the network searches for a better match (i.e. a different node with a higher choice function value).

Algorithm 1 Fuzzy ART Algorithm.

Set all necessary initial parameters (α, β, ρ, $\mathbf{w}$, $\mathbf{c} = \phi$)
for input pattern $\mathbf{I}_i$ **do**
    Normalize input pattern and apply complement coding
    for output node j of F_2 **do**
        Calculate choice function $T_j = \frac{\|\mathbf{I}_i \wedge \mathbf{w}_j\|_1}{\alpha + \|\mathbf{w}_j\|_1}$
    end for
    while Vigilance_criterion = false **do**
        Select category j, $T^* = \max\{T_j\}$
        Calculate distance $d = \|\mathbf{c}_j - \mathbf{I}_i\|_1$
        if Nodes remaining **then**
            if $\frac{\|\mathbf{I}_i \wedge \mathbf{w}_j\|}{\|\mathbf{I}_i\|_1} \geq \rho$ **then**
                Update weight vector $\mathbf{w}_j^{(new)} = \beta(\mathbf{I}_i \wedge \mathbf{w}_j^{(old)}) + (1-\beta)\mathbf{w}_j^{(old)}$
                Assign pattern $\mathbf{I}_i$ to category j
                Update cluster center $\mathbf{c}_j^{(new)}$
                Vigilance_criterion = true
            else
                Set T_j = -1
                Vigilance_criterion = false
            end if
        else
            Create new F_2 node
            $\mathbf{c} = \mathbf{c} \bigcup \mathbf{I}_i$
            Vigilance_criterion = true
        end if
    end while
end for

Since the vigilance parameter is a key factor in determining the learning performance of the network, using a fixed value may restrain the performance since some inputs may require greater vigilance than others. For this research endeavour, we show that by using a suitable, non-uniform value of vigilance for every incoming pattern the performance of the network can be significantly improved.

In order to focus on the effect of the vigilance parameter ρ, the other two parameters are kept fixed by training the network for fast learning in a conservative limit (i.e. $\beta = 1$ and $\alpha = 0^+$, i.e., slightly larger than 0).

3.2. Interval type-2 fuzzy sets

A T2 FS in a universe of discourse $\mathbf{X}$, denoted as $\tilde{\mathbf{A}}$, is characterized by a T2 fuzzy membership function (MF) $\mu_{\tilde{\mathbf{A}}}(x, u)$, where $x \in \mathbf{X}$ and $u \in J_x \subseteq [0,1]$, i.e.,

$$\tilde{\mathbf{A}} = \left\{ \left((x, u), \mu_{\tilde{\mathbf{A}}}(x, u) \right) \mid \forall x \in \mathbf{X}, \forall u \in J_x \subseteq [0, 1] \right\} \tag{1}$$

in which $0 \leq \mu_{\tilde{\mathbf{A}}}(x, u) \leq 1$ and J_x is the primary membership of x which is the domain of the secondary MF [9,29]. Fig. 2(a) illustrates a Gaussian T2 fuzzy MF with fixed mean and uncertain standard deviation. The shaded region bounded by two T1 fuzzy MFs, the upper membership function (UMF) and lower membership function (LMF) of $\tilde{\mathbf{A}}$, is called the footprint of uncertainty (FOU). If we take a vertical slice of $\mu_{\tilde{\mathbf{A}}}(x, u)$, we obtain the secondary MF, which may be in the form of an interval (see Fig. 2(b)), or a Gaussian (see Fig. 2(c)). The secondary MF in IT2 fuzzy MFs are interval sets expressed as

$$\tilde{\mathbf{A}} = \int_{x \in \mathbf{X}} \int_{u \in J_x} 1/(x, u) = \int_{x \in \mathbf{X}} \left[\int_{u \in J_x} 1/u \right] / x \quad J_x \subseteq [0, 1]. \tag{2}$$

Moreover, FOU($\tilde{\mathbf{A}}$) can also be expressed as

$$\mathrm{FOU}(\tilde{\mathbf{A}}) = \bigcup_{\forall x \in \mathbf{X}} \left[\underline{\mu}_{\tilde{\mathbf{A}}}, \overline{\mu}_{\tilde{\mathbf{A}}} \right]. \tag{3}$$

Fig. 2(b) shows the secondary MF of an IT2 FS that is all uniformly weighted equaling one for each primary membership of x. Therefore, the IT2 FS requires only simple interval arithmetic for calculating operations on T2 FS.

The type reduction methods for T2 FSs involve computation of their centroid which is a T1 FS given by

$$Y_{TR}(\mathbf{x}') = [y_l, y_r] = \int_{y^1 \in [y_l^1, y_r^1]} \cdots \int_{y^M \in [y_l^M, y_r^M]} \int_{f^1 \in [\underline{f}^1, \overline{f}^1]} \cdots \int_{f^M \in [\underline{f}^M, \overline{f}^M]} 1 / \frac{\sum_{i=1}^{M} f^i y^i}{\sum_{i=1}^{M} f^i}. \tag{4}$$

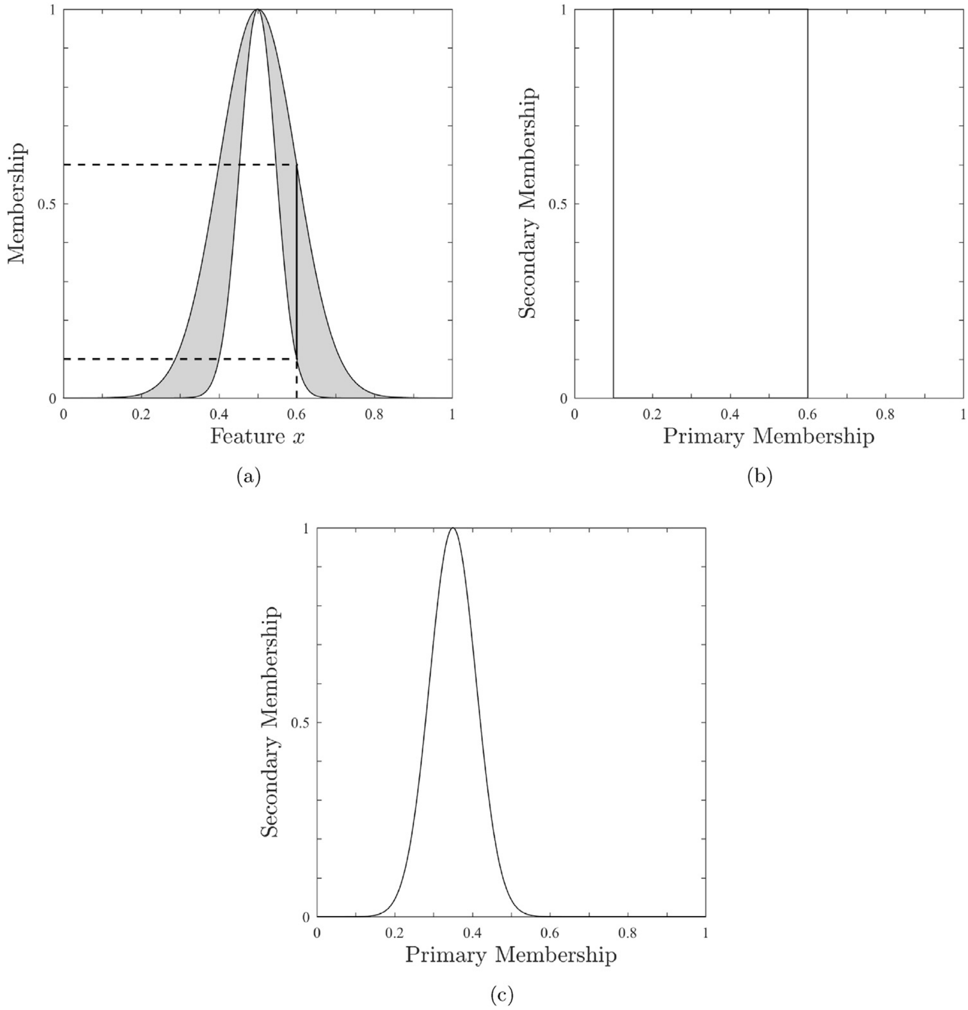


Fig. 2. Illustration of a T2 fuzzy MF: (a) FOU bounded by upper and lower membership functions, (b) interval secondary MF, and (c) Gaussian secondary MF.

In general, computing the centroid of any T2 FS may be computationally intensive. However, in the case of an IT2 FS, the centroid can be described by an interval FS, which may further be completely characterized by its left and right end points y_l and y_r, and hence it may be computed exactly if these two points are known. The KM iterative algorithm, proposed by Karnik and Mendel [21], makes use of this feature of IT2 FSs to efficiently compute their centroid. Several researchers have also come up with various alternative algorithms [10,21] to reduce the complexity of computing the centroid of IT2 FSs.

4. Proposed methods

In this section, we describe our proposed methods, namely T1 fuzzy ART with adaptive vigilance and IT2 fuzzy ART clustering.

4.1. Proposed T1 fuzzy ART with adaptive vigilance

We propose a method to adapt the vigilance parameter ρ throughout the clustering process by incorporating the distance information into the vigilance criterion. This is achieved by updating the cluster centers as the patterns are continuously assigned to the clusters, and then selecting a *suitable* value of ρ depending upon its distance from the cluster centers, instead of using a fixed value for all input patterns. Here, by *suitable* we mean that if the distance of the pattern from the cluster center is small (large), we select a relatively small (large) value of vigilance. It should be noted that all patterns are normalized between 0 and 1 to facilitate an appropriate qualitative categorization of the distance. Therefore, for a given match value, the vigilance criterion is easily obtained. Fig. 3 illustrates this concept, where $\rho_1 < \rho_2 < \ldots < \rho_{\max}$.

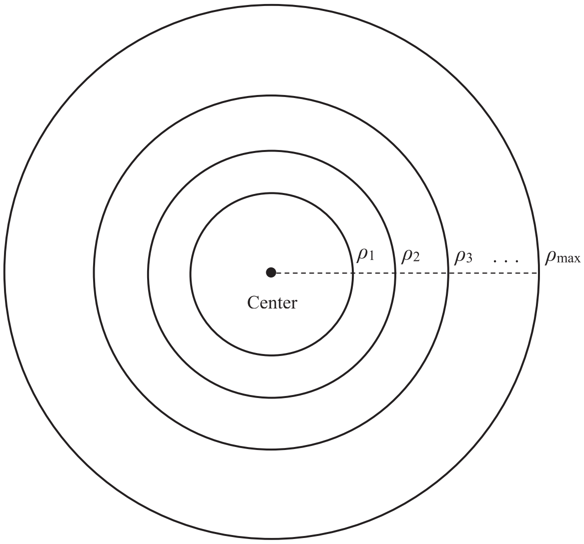


Fig. 3. Illustration of vigilance in terms of relative distance from cluster center.

For this purpose, we require a vigilance function which is defined in terms of the distance between the input pattern and cluster center, and subscribes to the property stated above. A candidate for such a function is

$$f(x) = (1 - e^{-mx}), \tag{5}$$

where m is the fuzzifier for the vigilance function. Section 4.1.3 describes various reasons for selecting this function as the vigilance.

As a result, the overall vigilance for a $2M$-dimensional input pattern $\mathbf{I}$ can be given by

$$\rho = \sum_{k=1}^{2M} \rho_k, \quad \text{where } \rho_k = (1 - e^{-md_k}), \tag{6}$$

$d_k = |c_{jk} - I_k|$, and c_{jk} is feature (element) k of cluster center j. It must be noted that while (5) fulfils the criteria for the vigilance function, the user may heuristically select any other function such as triangular, trapezoidal, Gaussian, S-function, and π-function [9] that better suits the requirements of the application concerned.

4.1.1. Regulating the value of ρ

There are several methods to regulate the value of ρ. Some of them are as follows.

1. Set $\rho^* = \frac{\rho}{2M}$, where M denotes the dimension of input data.
   (i.e., $\rho = \sum_{k=1}^{2M} \rho_k$, where $\rho_k < 1\ \forall k$ and $\frac{\rho}{2M} < 1$).
   This guarantees ρ^* to always be less than 1.
2. Set $m \geq \ln(\frac{2M}{2M-1})$.
   Find maximum value of function $(1 - e^{-md_k})$, i.e., 1 - $\min\limits_{d_k \in [0,1]} \{e^{-md_k}\}$.
   Since minimum value of e^{-md_k} occurs at $d_k = 1$,
   $\max\{\rho_k\} = 1$ - e^{-m}, or $\rho_{\max} = 2M(1 - e^{-m})$.
   Setting $\rho_{\max} = 1$,
   $e^{-m} = \frac{2M-1}{2M}$, or
   $m = \ln(\frac{2M}{2M-1})$.

4.1.2. Relation between m and ρ

If the input pattern is near (far from) the cluster center, the value of $\rho \to 0$ ($\rho \to 1$). This is in accordance with the fact that a pattern closer to the cluster center should have high membership of belonging to that cluster, and vice versa. Therefore, the choice of parameter m is extremely crucial for an optimal cluster structure as it decides the growth of the value of ρ as the input pattern moves away from the cluster center. Thus, depending on the choice of m in the proposed T1 fuzzy ART algorithm, the network self-organizes itself into appropriate number and structure of clusters that represents the input data. Hence, m is a critical factor in determining the network dynamics.

4.1.3. Reasons for selecting the vigilance function as $f(x) = (1 - e^{-mx})$

1. This vigilance function is simple since it involves only a single parameter which can be adjusted according to the requirement of the application. A higher (lower) value of m causes the algorithm to generate comparatively more (less) number of clusters, i.e., a more (less) vigilant clustering process.
2. The function outputs the value of vigilance from the interval $[0, \rho_{\max}]$, where 0 corresponds to the vigilance for the input parameter at the center and $\rho_{\max}$ corresponds to the patterns at relative distance 1 from the center.
3. The parameter m controls rate of change in vigilance for a given distance. With a small (large) value of m, the value of vigilance increases at a slower (faster) rate, resulting in the formation of generalized (specialized) categories.
4. This vigilance function is effective for partitioning samples into radial clusters since the vigilance only depends on the value of the fuzzifier m.

With this definition of a vigilance parameter, the conventional fuzzy ART can be modified to obtain the proposed T1 fuzzy ART with an adaptive vigilance algorithm by adding the following lines prior to the vigilance test in the Algorithm 1 (i.e., after line 11) described earlier.

Select category j, $T^* = \max\{T_j\}$
Calculate distance $d_{ik} = |c_{jk} - I_k|$
Estimate vigilance $\rho = \sum_{k=1}^{2M}(1 - e^{-md_{ik}})$
Calculate ρ^*, as explained in Section 4.1.1
Use ρ^* in place of the standard ρ in the vigilance test

4.2. Proposed IT2 fuzzy ART

For some pattern recognition applications, the prototypes and decision boundaries are estimated by the information present in the given pattern set. However, it may not always be possible to extract the necessary information from the pattern set for such an estimation, due to various defects in the collection, organization, or reporting of the data. This unavailability of information may then lead to the selection of unsatisfactory parameters for the data, resulting in the formation of incompatible prototypes. For example, when performing clustering with fuzzy sets, the assigned T1 fuzzy membership values may not be appropriate, due to the selection of parameters such as distance measure and fuzzifier.

In the previous section, we defined an adaptive vigilance function in an attempt to improve the conventional fuzzy ART clustering algorithm. The choice of parameter m is critical in assigning vigilance values for any given data set. Often it may be difficult to use a single value of m to represent all the clusters in a given pattern space (e.g., clusters that differ in cluster size and density), which suggests that uncertainty may exist in the choice of parameter m.

For this research endeavor, we manage the uncertainty associated with the parameter m in the proposed vigilance function by extending it to an IT2 fuzzy MF. As a result, the proposed IT2 fuzzy ART algorithm can be illustrated in Fig. 4.

4.2.1. Extensions to an IT2 fuzzy MF

After applying the input pattern to the fuzzy ART algorithm, a potential cluster j is chosen to encode this pattern using the category choice function. Next, we calculate the distance from the cluster center for all the patterns that have already been categorized into cluster j, including the current input pattern. This constitutes a pattern set such that d_{ik} is the distance of pattern i from the center of cluster j for feature k. The vigilance for each pattern can then be computed using the proposed function $\rho = \sum_{k=1}^{2M}(1 - e^{-md_{ik}})$. Now, in order to extend the vigilance function to an IT2 fuzzy MF, we use two different values of the parameter m to define the footprint of uncertainty (FOU). Under such an assumption, the upper and lower vigilance membership values for pattern i are computed as

$$\overline{\rho}_i = 1 - e^{-m_1 d_{ik}} \quad \text{and} \quad \underline{\rho}_i = 1 - e^{-m_2 d_{ik}}, \tag{7}$$

respectively.

Here, m_1 and m_2 are the fuzzifiers that represent different degrees of vigilance for any given input pattern. We first determine the upper and lower membership functions for the interval corresponding to the primary membership of the input pattern. Similarly, by using (7) and the KM algorithm, we obtain the upper and lower vigilance parameter values for any pattern.

Fig. 5(a) shows an example of the FOU, expressed as the shaded horizontal region formed by using two values of the fuzzifier m. The secondary memberships for each primary membership (which, in our case, is the vigilance) are all uniformly weighted. Fig. 5(b) shows a vertical slice of the IT2 fuzzy MF.

4.2.2. Type reduction and defuzzification

As mentioned above, we extend the vigilance to an IT2 fuzzy MF, and then compute its centroid by performing a type reduction. The KM iterative algorithm is used to compute the centroid of an IT2 fuzzy MF which is an interval-valued set given by

$$\mathbf{v} = [\mathbf{v}^L, \mathbf{v}^R]. \tag{8}$$

The centroid of an IT2 fuzzy MF can be completely characterized by its left and right end points, $\mathbf{v}^L$ and $\mathbf{v}^R$), respectively. The determination of $\mathbf{v}^L$ ($\mathbf{v}^R$) also gives an estimate for the left (right) vigilance value ρ_i^L (ρ_i^R) for all the patterns, since the

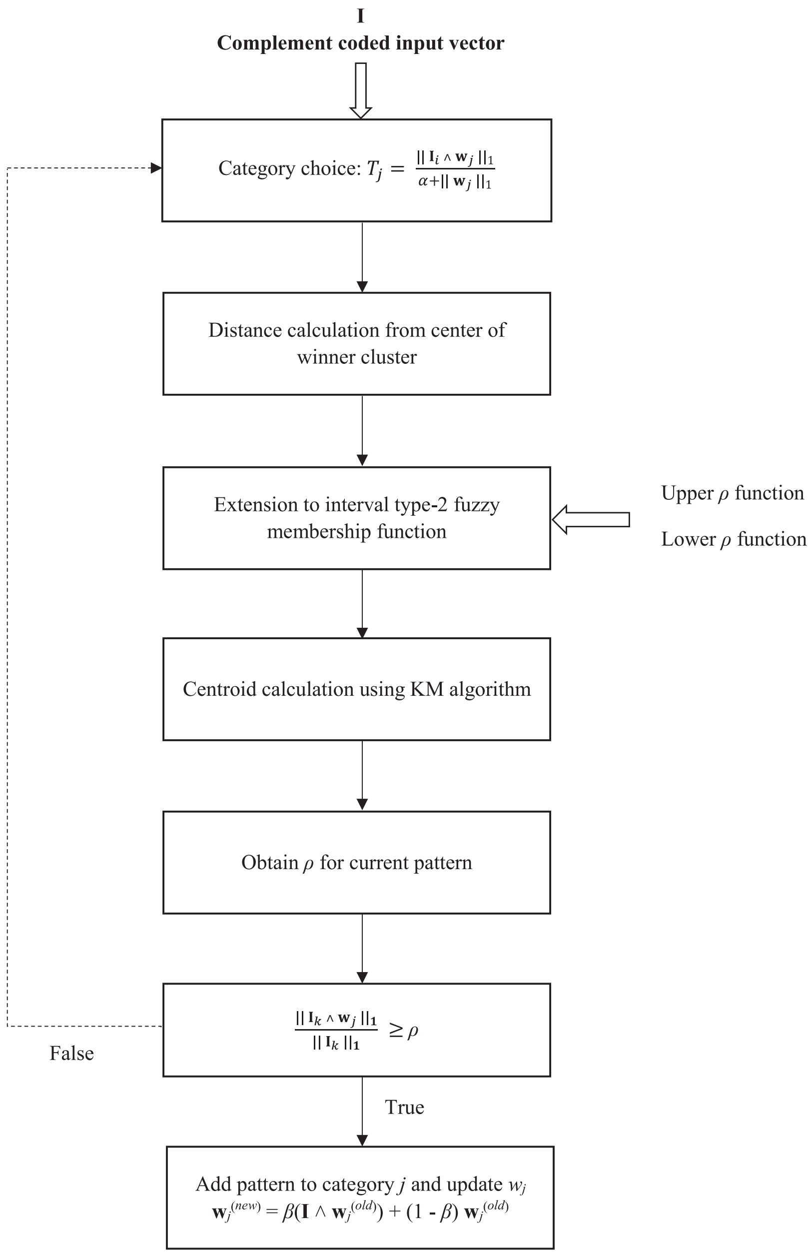


Fig. 4. Proposed IT2 fuzzy ART algorithm.

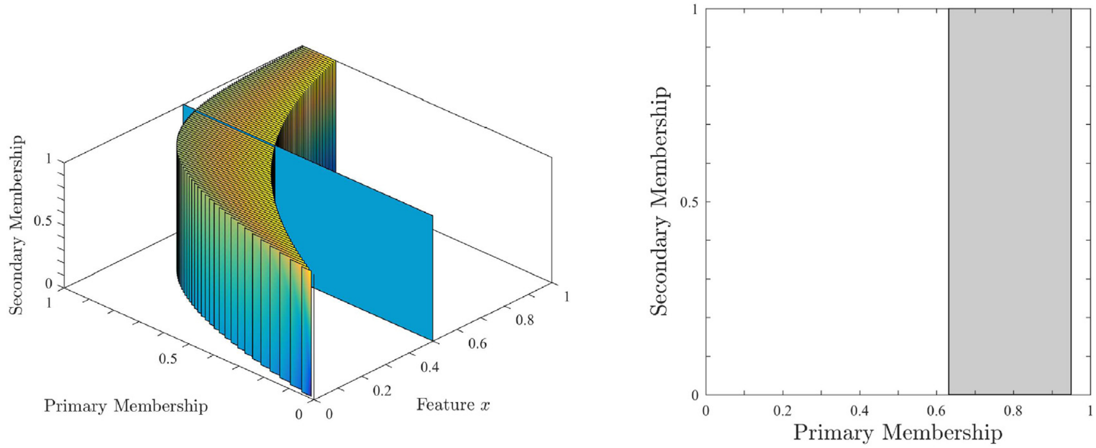
vigilance parameters are considered to be equal to the value of the centroids. We can then extract the vigilance value for the current input pattern and use it in the conventional fuzzy ART algorithm to perform a vigilance test. The KM iterative algorithm to calculate the minimum of centroid $\mathbf{v}^L$ is described in Algorithm 2 below.

The right centroid $\mathbf{v}^R$ can be obtained using the previous algorithm and replacing the "**if**" statement (i.e., line 15) by

if $(i < l)$ **then**
  Set $\rho_i = \underline{\rho}_i$
else
  Set $\rho_i = \overline{\rho}_i$
end if

In addition, the statement on line 29 should be replaced with

Set $\mathbf{v}^R = \mathbf{v}'_j = (c'_{j1}, \ldots, c'_{j2M})$



(a) (b)

Fig. 5. Illustration of our proposed IT2 fuzzy vigilance MF: (a) FOU with vertical slice and (b) interval secondary MF.

Algorithm 2 Iterative Algorithm for finding Left Centroid $\mathbf{v}^L$ for Category j.

Set all necessary initial parameters $(\alpha, \beta, \rho, m_1, m_2, \mathbf{w}, \mathbf{c} = \phi)$
for feature k of pattern set **do**
    for all patterns $\mathbf{I}_i$ in category j **do**
        Calculate distance $d_{ik} = |c_{jk} - I_{ik}|$
    end for
    Compute $\rho_i = \frac{\overline{\rho}_i + \underline{\rho}_i}{2}$
    Compute centroid using $c_{jk} = \frac{\sum_{i=1}^{P} I_{ik}\rho_i}{\sum_{i=1}^{P} \rho_i}$
end for
Sort pattern indexes for all P patterns, for each of $2M$ elements in ascending order
Set comparison = false
for feature k **do**
    while comparison = false **do**
        Find interval index l $(1 \le l \le P-1)$, such that $I_{lk} \le c_{jk} \le I_{(l+1)k}$
        for all patterns in the category j **do**
            if $i \le l$ **then**
                Set $\rho_i = \overline{\rho}_i$
            else
                Set $\rho_i = \underline{\rho}_i$
            end if
        end for
        Compute left value of centroid $c'_{jk} = \frac{\sum_{i=1}^{P} I_{ik}\rho_i}{\sum_{i=1}^{P} \rho_i}$
        if $c'_{jk} = c_{jk}$ **then**
            Set comparison = true
        else
            Set $c'_{jk} = c_{jk}$
        end if
    end while
end for
Set $\mathbf{v}^L = \mathbf{v}'_j = (c'_{j1}, \ldots, c'_{j2M})$

Once we have computed the interval fuzzy MF for the cluster centroid, its defuzzified value can be expressed as

$$\mathbf{v}_j = \frac{\mathbf{v}^L + \mathbf{v}^R}{2}. \tag{9}$$

In computing $\mathbf{v}^L$ and $\mathbf{v}^R$, we have evaluated ρ_i^L and ρ_i^R for the input pattern $\mathbf{I}_i$ for all the patterns currently present in the category j. Hence, the defuzzified value of the vigilance parameter used in the vigilance test for the current input $\mathbf{I}_i$ can be evaluated as

$$\rho_i = \frac{\rho_i^L + \rho_i^R}{2}. \tag{10}$$

The resulting vigilance value is then used to compare the similarity measure between input $\mathbf{I}_i$ and node j, which is declared as the winner node in the category choice step of conventional fuzzy ART.

4.3. Convergence and complexity of proposed methods

For the conventional ART system, complement coding and fast learning ensures the creation of stable hypercubes along with the following consequences [6].

1. The category hypercube is bounded as $|R_j| \leq M(1-\rho)$, where R_j, M, and ρ denote the category hypercube, dimensionality, and vigilance parameter, respectively.
2. The weight vector $\mathbf{w}_j$ is monotonically decreasing.

Our proposed T1 fuzzy ART and IT2 fuzzy ART methods follow complement coding and fast learning. Furthermore, it exploits an update rule similar to that used in the conventional ART, which ensures that $\mathbf{w}_j$ decreases monotonically. Since we deal with a fixed, finite data set, it may be argued that it is always possible to find a fuzzifier parameter that sets the minimum value of ρ to the desired value $\rho_{\min}$ for any worst case scenario of category creation. Consequently, we can have an upper bound on the hypercube where $|R_j| \leq M(1-\rho_{\min})$ and $\rho_{\min} \in (0, 1)$, such that the hypercube is bounded. This, in addition to the monotonically decreasing weight vector, ensures that the methods converge in a finite number of iterations. Finally, normalization of the inputs using complement coding allows our proposed methods to overcome the category proliferation problem while retaining the stable coding properties of the weight update rule.

To analyze the computational complexity of our proposed methods, we can visualize it as a module within another module. The outer module is the traditional ART clustering algorithm which uses a fixed value of the vigilance parameter ρ. Within this module, we have another module that obtains a defuzzified value for ρ using the KM iterative algorithm for type reduction.

For the outer module, the complexity depends on the maximum number of iterations of the training set. For each iteration, the choice function and the vigilance criterion is evaluated for each input pattern against the categories already created. Therefore, the complexity $T_{ART} \sim ENC$, where E, N, and C denote the maximum number of iterations, size of training set, and number of categories, respectively. The maximum number of iterations is usually small when compared to the size of the training set ($E \ll N$). Therefore, the running time complexity of the conventional ART is of the order $O(NC)$.

In the inner module, we have used the KM algorithm that follows an iterative routine, the only available convergence approximation [21] for which is highly pessimistic (approximately the number of sampled values of the primary variable). Thus, an upper bound on the running time complexity T_{KM} would be $O(N^2)$. However, in recent times the convergence observations have recently been quantified to prove super-exponential convergence for the algorithms [30].

The overall complexity of our proposed methods is simply the product of T_{ART} and T_{KM}. Finally, we have that $T_{IT2ART} = T_{ART} \times T_{KM}$.

5. Experimental results

In this section, we illustrate various experimental results to compare the clustering performance of the conventional fuzzy ART algorithm, the proposed T1 fuzzy ART with adaptive vigilance, and the IT2 fuzzy ART algorithm. We begin by demonstrating the clustering procedure for simple synthetic data and then proceed to higher dimensional data sets. We also provide comparisons of the clustering methods for image segmentation problems.

For the conventional fuzzy ART algorithm, experimental results for various values of ρ are first obtained, and then the value of ρ for which the output is *optimal* (the value of ρ which generates the required number of categories with the most suitable result) is selected. For T1 fuzzy ART with adaptive vigilance, the clustering performance is controlled using the fuzzy parameter m, while for IT2 fuzzy ART, m_1 and m_2 are used for this purpose.

For all experiments, our proposed methods focus on selecting an adaptive choice for the *vigilance parameter* ρ, and the remaining parameters of the fuzzy ART algorithm are kept constant (i.e., choice parameter α=0.0001 and learning rate β=1). It is to be noted that clusters generated by the conventional fuzzy ART algorithm may be considered order dependent, which implies that the algorithm may in general provide satisfactory results when neighboring patterns belonging to the same category are sequentially presented. For this reason, we present randomly ordered input data to all three algorithms for all experiments, for an unbiased evaluation of their performance.

5.1. Similar "circle" data

Our first experiment shows the clustering results for 60 patterns consisting of two features that are distributed as two disjoint circular clusters of similar volume and an equal number of patterns. Identical results were obtained for all three methods for a large range of fuzzifier values, as shown in Fig. 6. The clusters were sufficiently distinct such that any clustering algorithm was able to identify them effectively.

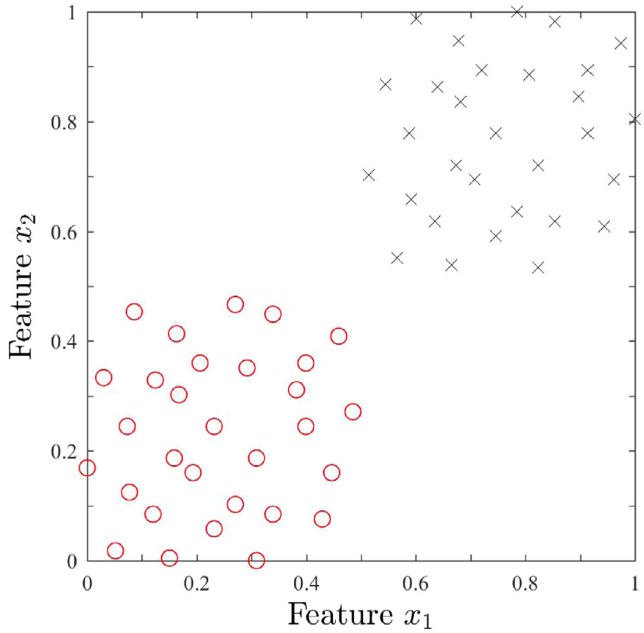


Fig. 6. Clustering results on a pattern set consisting of symmetric disjoint circular shaped data of similar volume for fuzzy ART (ρ= 0.48), proposed T1 fuzzy ART with adaptive vigilance (m = 1.5), proposed IT2 fuzzy ART (m_1 = 2.0, m_2 = 0.7).

Table 1
Recognition rate (R_rate) and confusion matrix for Iris data, for fuzzy ART, T1 fuzzy ART with adaptive vigilance (Proposed T1), and IT2 fuzzy ART (Proposed IT2).

		Recognized class								
		Fuzzy ART			Proposed T1			Proposed IT2		
		ρ=0.12			m=1.80			(m_1, m_2)=(3.30,1.40)		
R_Rate (%)		66.00			76.00			89.33		
		1	2	3	1	2	3	1	2	3
True class	1	50	0	0	50	0	0	50	0	0
	2	5	44	1	7	43	0	3	43	4
	3	1	44	5	1	28	21	0	9	41

5.2. Asymmetric “circle” data

In this example, the pattern distribution consists of two asymmetric circles (30 patterns each) of different volume, as shown in Fig. 7(a). From the results obtained using the conventional fuzzy ART, the algorithm was unable to correctly categorize five patterns (Fig. 7(b)), which were misclassified as part of the smaller cluster. For the proposed T1 fuzzy ART with adaptive vigilance, the number of misclassified patterns was slightly reduced to three (Fig. 7(c)). Finally, the proposed IT2 fuzzy ART *successfully* trained the network for all 60 patterns, i.e., the network was capable of classifying the data in accordance with the original data set (Fig. 7(d)). In each of these figures, the centers of the resulting clusters are represented by the green square (□).

5.3. “Iris” data

The “Iris” data set consists of 150 patterns of four features and three classes (50 patterns in each class). The patterns belonging to class 1 are known to be separable from the other two. However, some patterns in class 2 and class 3 are considered to overlap.

For this series of experiments, the network was trained using 149 samples from the Iris data (i.e., all patterns except 1) for all 4 features, and then used to recognize the remaining single test sample. The experiment was repeated 150 times and each sample was used exactly once to test the network, while the remaining 149 were used for training. The recognition rate and confusion matrix obtained by performing the above mentioned experiments is presented in Table 1.

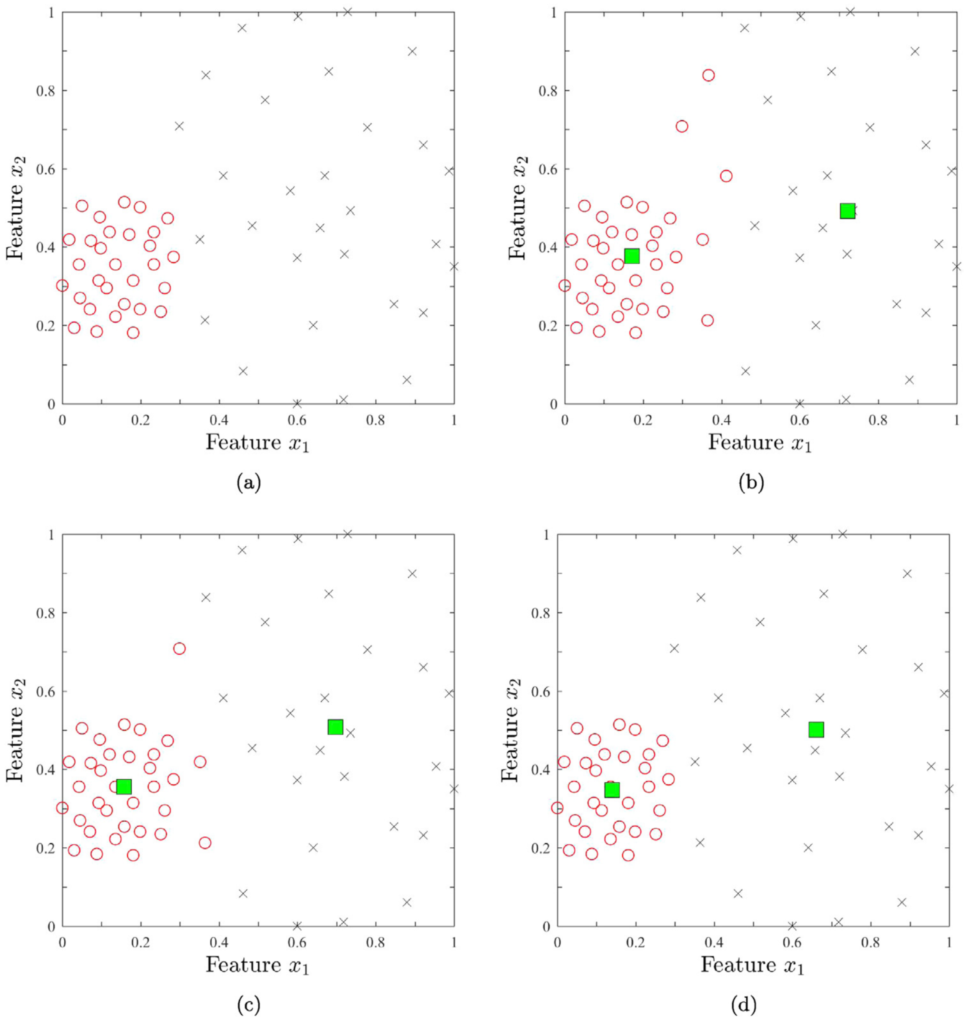


Fig. 7. Clustering results on a pattern set consisting of two asymmetric circular shaped data: (a) scatter plot, (b) fuzzy ART ($\rho = 0.18$), (c) proposed T1 fuzzy ART with adaptive vigilance ($m = 0.5$), and (d) proposed IT2 fuzzy ART ($m_1 = 0.3$, $m_2 = 0.1$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

All methods gave 100% correct results in clustering patterns belonging to class 1. For the remaining classes (i.e., class 2 and 3), the results were not satisfactory for the conventional fuzzy ART algorithm. However, by using our proposed T1 fuzzy ART with adaptive vigilance, the network was able to improve the clustering patterns in class 2 and 3, as indicated in Table 1. The results obtained by the proposed IT2 fuzzy ART algorithm were further improved.

This is in accordance with results in clustering of the Iris data set using fuzzy methods for training neural networks. In two such endeavors, a fuzzy min-max clustering neural network produced 14 clusters from the data [43], while an improvement in the learning rule for fuzzy ART was found to be over 90% accurate for the original improved fuzzy ART [15]. Our results also compare favorably with existing unsupervised clustering approaches for the Iris data. In [47], integrating fuzzy sets into the Kohonen clustering algorithm reduced the number of misclassifications from 25 to 15, similar to our observations.

5.4. "Wine" data

In a classification context, this data set is a well posed problem since it contains "well behaved" class structures, but requires supervision for clustering. For this reason, it is suitable for initial testing of classifiers. The data set consist of three classes each having 59, 71, and 48 data points, respectively. Each pattern consists of 13 features, where most of the patterns are overlapping for the three clusters. The conventional fuzzy ART algorithm was not able to cluster the data effectively, while the clustering performance was improved by employing our proposed algorithms, considering all 13 features. A comparison of the three methods is shown in Table 2.

Table 2
Recognition rate (R_Rate) and confusion matrix for Wine data, for fuzzy ART, T1 fuzzy ART with adaptive vigilance (Proposed T1), and IT2 fuzzy ART (Proposed IT2).

		Recognized class								
		Fuzzy ART			Proposed T1			Proposed IT2		
		ρ=0.19			m=0.32			(m_1, m_2)=(0.30,0.24)		
R_Rate (%)		54.49			65.73			78.09		
		1	2	3	1	2	3	1	2	3
True class	1	40	16	3	54	1	4	53	6	0
	2	20	31	20	33	17	21	10	60	1
	3	21	1	26	2	0	46	3	19	26

Table 3
Precision, Recall, and F1 score for the task Gesture Phase Segmentation, for fuzzy ART, T1 fuzzy ART with adaptive vigilance (Proposed T1), and IT2 fuzzy ART (Proposed IT2).

Performance metric (%)	Fuzzy ART ρ=0.03	Proposed T1 $m = 1.3$	Proposed IT2 $(m_1, m_2) = (1.7, 1.1)$
Precision	77.93	75.67	80.83
Recall	69.89	73.72	74.81
F1 Score	73.69	74.68	77.70

Table 4
Recognition rate (R_Rate) and confusion matrix for Shuttle image, with recognized classes shuttle (S) and background (B), for fuzzy ART, T1 fuzzy ART with adaptive vigilance (Proposed T1), and IT2 fuzzy ART (Proposed IT2).

		Recognized class					
		Fuzzy ART		Proposed T1		Proposed IT2	
		ρ=0.02		m=1.2		(m_1, m_2)=(1.3,0.2)	
R_Rate (%)		20.66		51.21		83.71	
		S	B	S	B	S	B
True class	S	274	1052	679	647	1110	216
	B	0	38,674	6	38,668	123	38,551

Although there are other unsupervised methods such as k'-clustering [49], maximum certainty data partitioning [40], and Mercer kernel-based clustering [13], that have recently shown high degrees of accuracy (only four misclassifications) with the Wine data set. We argue that the performance of our method does not compare with these methods due to an underlying limitation with the ART network rather than the IT2 approach. This is evident from the observation that the recognition rate for our proposed IT2 fuzzy ART method may be considered up to 40% better than that of the existing fuzzy ART. This further suggests that the integration of an IT2 fuzzy MF with a more accurate base model may perform significantly better than existing state-of-the-art approaches.

5.5. Gesture phase segmentation

As an example of a large, high-dimensional data set, we consider the problem of gesture phase segmentation. For this, we used a dataset of samples collected from seven videos containing gesturing people [27], captured by a Microsoft Kinect sensor. There are 9900 labeled instances arising from the seven videos, and each instance consists of 50 numeric features.

A single video, represented by a sequence of n frames $S = \{f_1, f_2, \ldots, f_n\}$, is input to a segmentation strategy aimed at identifying gesture phases. The segmentation problem consists of the input frame representation f_i and classification of these frames into one of the classes $c_i = \{E, P, S, H, R\}$, corresponding to rest, preparation, stroke, hold, and retraction.

Here, we perform the task of classifying rest positions: input $f_i \in S$, and output $c_i = \{E, G\}$, where $G \supset \{P, S, H, R\}$ corresponds to G gesture units. Table 3 shows the precision [1], recall [2] and F1 score [3] for the three methods. Our proposed

[1] The fraction of true instances among all instances classified into a label.
[2] The fraction of instances classified correctly among all instances belonging to a label.
[3] Harmonic mean of precision and recall.

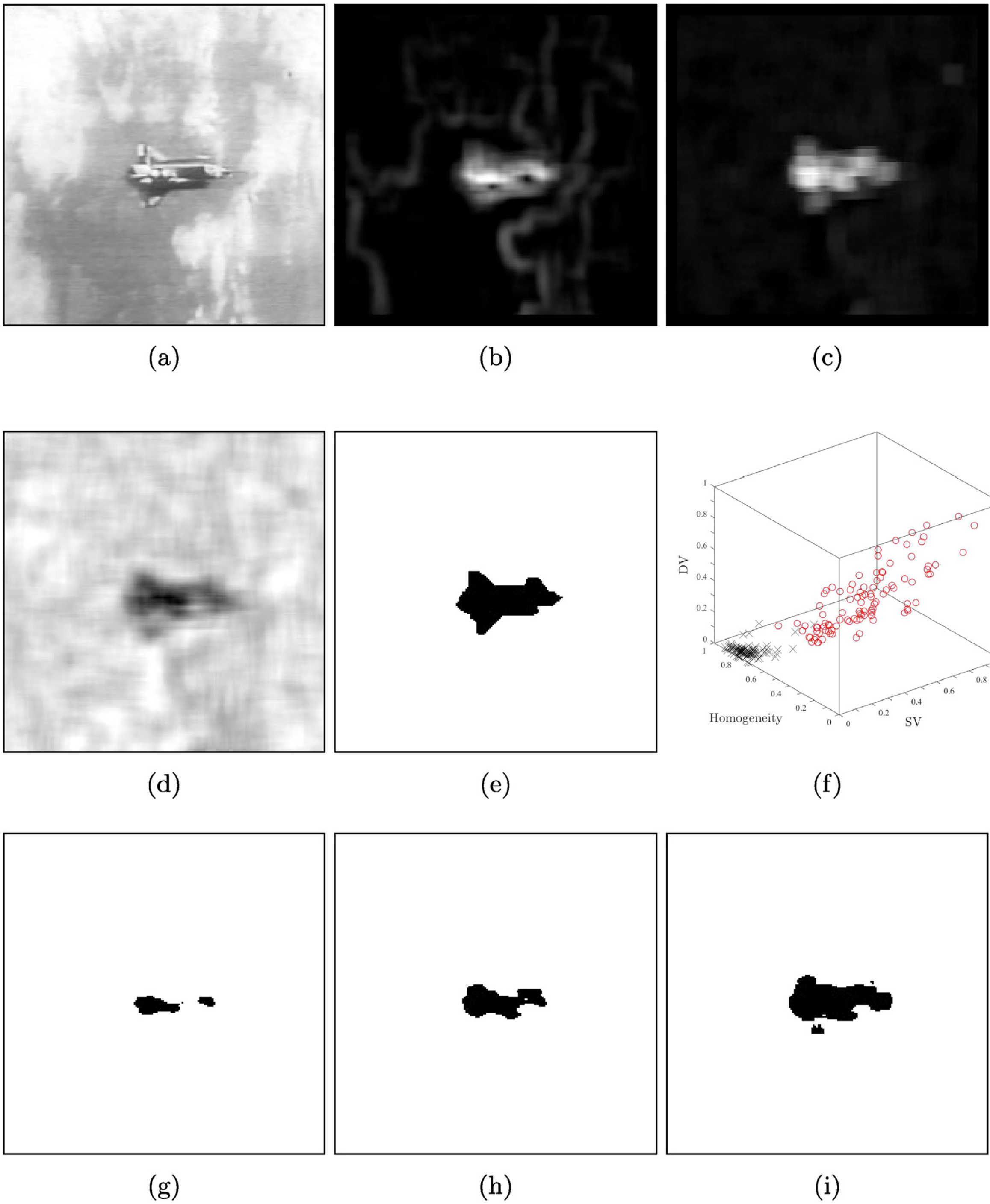


Fig. 8. Image segmentation for "Shuttle": (a) original image, feature images ((b) squared variance (SV), (c) difference variance (DV), and (d) homogeneity), (e) mask image, (f) scatter plot of sample patterns, segmentation results for (g) fuzzy ART, (h) proposed T1 fuzzy ART with adaptive vigilance, and (i) proposed IT2 fuzzy ART.

fuzzy ART methods outperform the conventional fuzzy ART on all the metrics, which sufficiently shows its specificity and sensitivity (see Table 3).

In previous research, motion capture techniques [38] and temporal structures [48] have been used to obtain classification accuracy of 80–90% in this task. Although our method does not fare as well as these techniques, it still gives better generalization than using complex hand-crafted features.

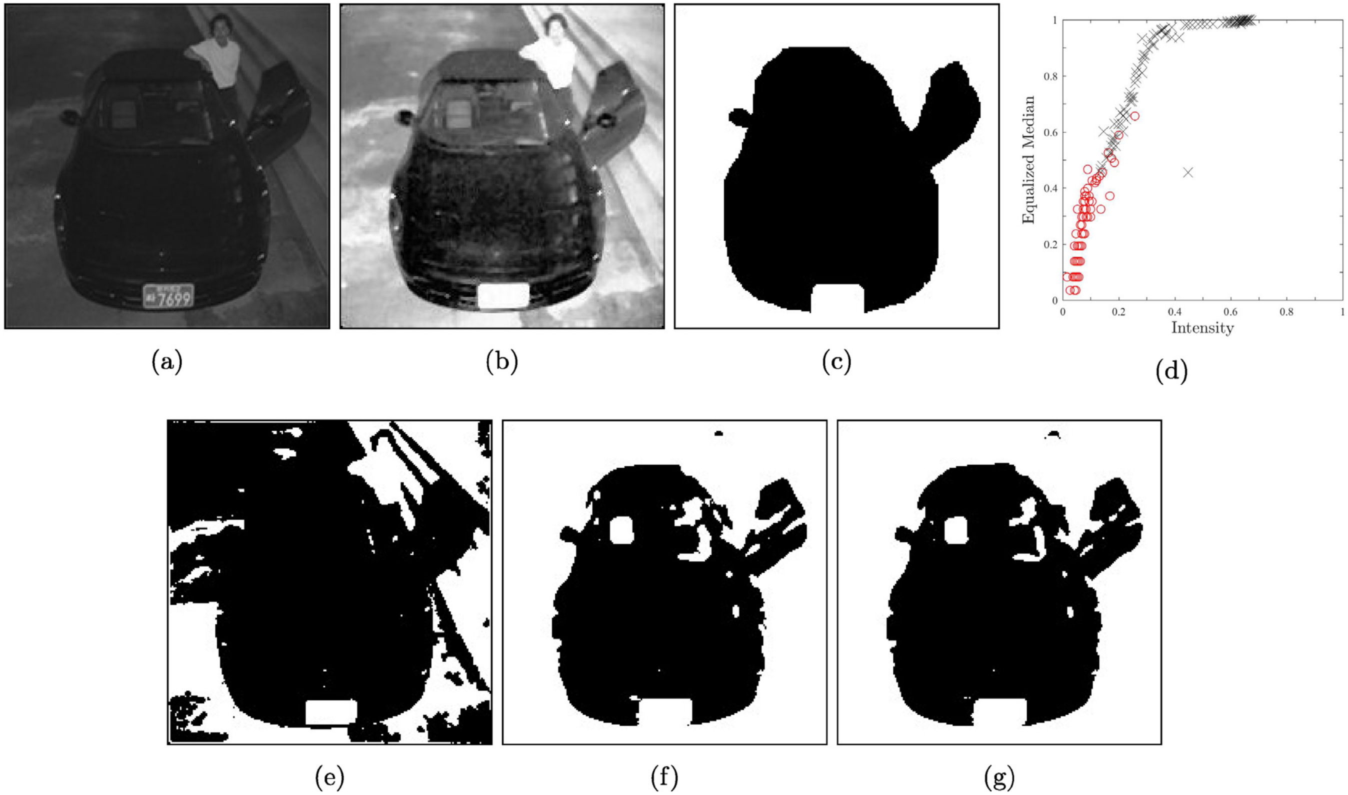


Fig. 9. Image segmentation for "Elan": feature images ((a) intensity and (b) median filtered intensity), (c) mask image, (d) scatter plot of sample patterns, segmentation results for (e) fuzzy ART, (f) proposed T1 fuzzy ART with adaptive vigilance, and (g) proposed IT2 fuzzy ART.

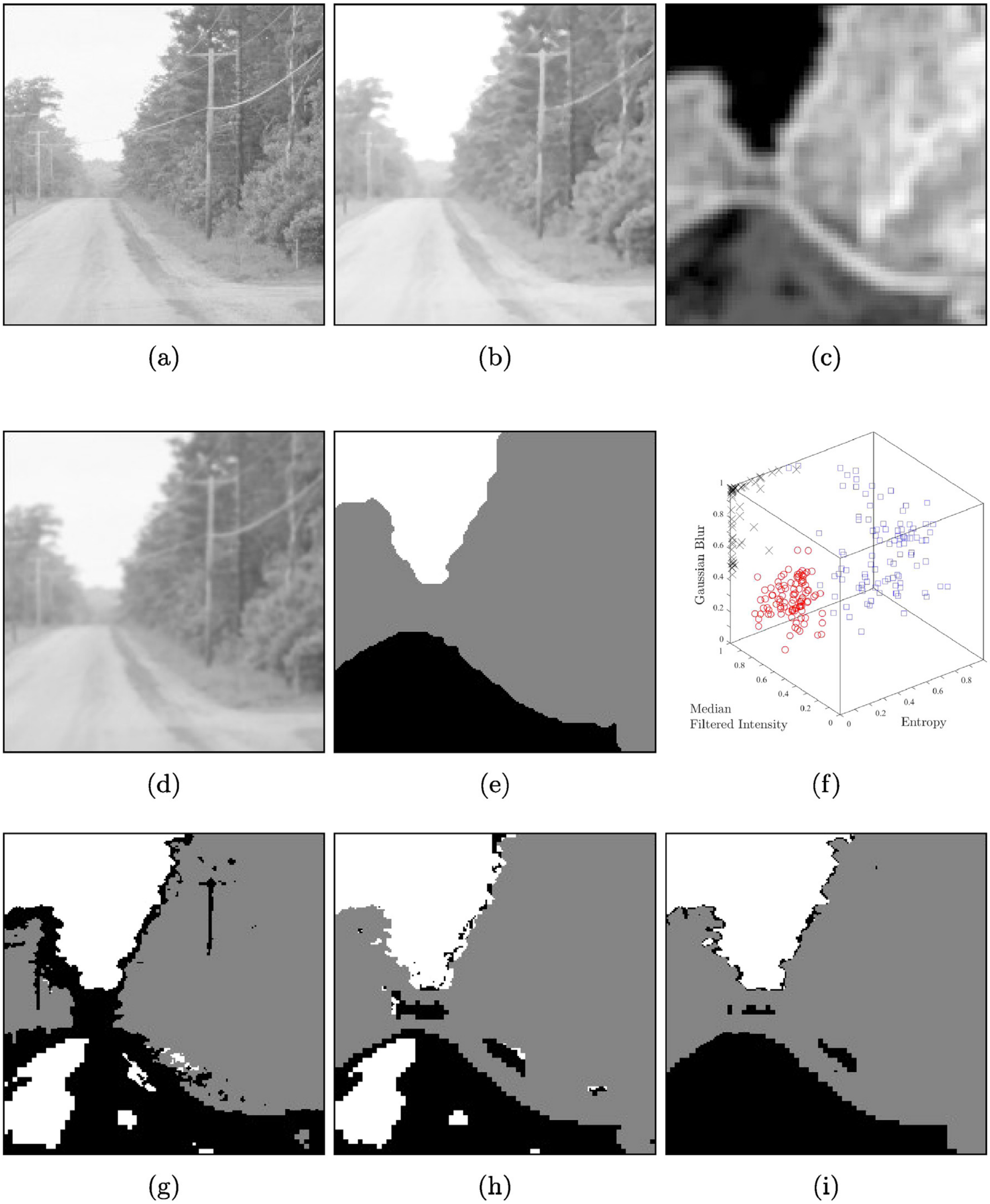


Fig. 10. Image segmentation for Scene 1: (a) original image, feature images ((b) median filtered intensity, (c) entropy, and (d) Gaussian blur), (e) mask image, (f) scatter plot of sample patterns, segmentation results for (g) fuzzy ART, (h) proposed T1 fuzzy ART with adaptive vigilance, and (i) proposed IT2 fuzzy ART.

5.6. 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. For this example, we partitioned pixels of a grayscale image into salient regions. The process was carried out in the following steps.

Table 5
Recognition rate (R_Rate) and confusion matrix for Elan image, with recognized classes car (C) and background (B), for fuzzy ART, T1 fuzzy ART with adaptive vigilance (Proposed T1), and IT2 fuzzy ART (Proposed IT2).

		Recognized class					
		Fuzzy ART		Proposed T1		Proposed IT2	
		ρ=0.38		m=1.6		(m_1, m_2)=(1.7,1.6)	
R_Rate (%)		74.11		94.74		95.81	
		C	B	C	B	C	B
True class	C	18,714	145	16,852	2007	17,333	1526
	B	10,211	10,930	99	21,042	151	20,990

- *Step* 1 *Pre-processing*: Feature images are generated, namely median filter, gaussian blur, squared variance (SV), difference variance (DV), and homogeneity, to create a higher dimensional data set by incorporating the pixel information from these images.
- *Step* 2 *Learning and Categorization*: From step 1, we collect n sample data patterns uniformly at random from each region of the image that is considered to be labelled, to train the networks. The networks are then used to categorize the input data.
- *Step* 3 *Image segmentation*: A segmented (grayscale) image is generated from the cluster categories obtained in Step 2, by assigning a unique grayscale value (i.e., 0–255) to each cluster.

5.6.1. Shuttle image

For this example, we consider a 200 × 200 shuttle image consisting of two regions: shuttle and background, as shown in Fig. 8(a). Figs. 8(b), 8(c), and 8(d) show the three feature images, namely, SV, DV, and homogeneity, generated from the shuttle image. The networks are trained by randomly selecting 100 3-dimensional sample pixels from each region of the feature images. In Fig. 8(f), the patterns labelled as " × " are extracted from the background region and the patterns labelled as "○" are extracted from the shuttle region.

The pixels in the shuttle image are classified into two regions by using the trained network. The segmentation results corresponding to the three algorithms are then compared with the mask image shown in Fig. 8(e). Fig. 8(g) shows that the conventional fuzzy ART network misclassified most of the pixels in the shuttle region. Performance was improved by using the proposed T1 fuzzy ART with adaptive vigilance, however, several samples in the shuttle region are still misclassified (see Fig. 8(h)). However, by using the proposed IT2 fuzzy ART algorithm, however, the classification performance improved significantly, as shown in Fig. 8(i). The confusion matrix showing the results and recognition rate is given in Table 4. Since the number of samples in the shuttle region are significantly less than those in the background, the recognition rate for the shuttle is used as the measure of accuracy since the overall recognition rate may become high even with many false negatives.

5.6.2. Elan image

In this example, we consider a 200 × 200 Elan image consisting of two regions: car and background, as shown in Fig. 9(a). We used the original intensity image, and median filtered image after histogram equalization, as features. Fig. 9(a) and (b) show the two feature images, and Fig. 9(d) shows the scatter plot after selecting 100 random data points from the two regions. The patterns labelled as " × " are extracted from the background region and the patterns labelled as "○" are extracted from the car region.

Fig. 9(e)–(g) show result for the various methods, and it can be observed that the conventional fuzzy ART incorrectly classifies a large number of pixels in the background region due to the similarity in the intensity values, as justified by the overlapping " × " and "○" samples in Fig. 9(d). In Fig. 9(f), the performance is improved where the number of misclassified pixels is reduced significantly. A further enhancement in the classification is observed by using the proposed IT2 fuzzy ART method as shown in Fig. 9(g). Table 5 shows the recognition results for the three methods.

5.6.3. Natural scenes

As our final example, we select four 200 × 200 natural scene images. All images consist of three regions: road, sky, and forest. The feature images used for this example are median filtered intensity, entropy, and gaussian blur. We randomly select 100 sample pixels from each region of the feature images, thus obtaining a total of 300 samples for each of the four scenes. Fig. 10(b)–(d) show the feature images for Scene 1. Note that the feature images for the other scenes are omitted due to redundancy.

Fig. 10(f) shows the scatter plot for the sample patterns of the three features used for Scene 1. The pattern labelled as "○, " "□," and " × " were extracted from the road, forest, and sky region, respectively. Figs. 11(c), 12(c), and 13(c) show the scatter plot for Scene 2, 3, and 4, respectively.

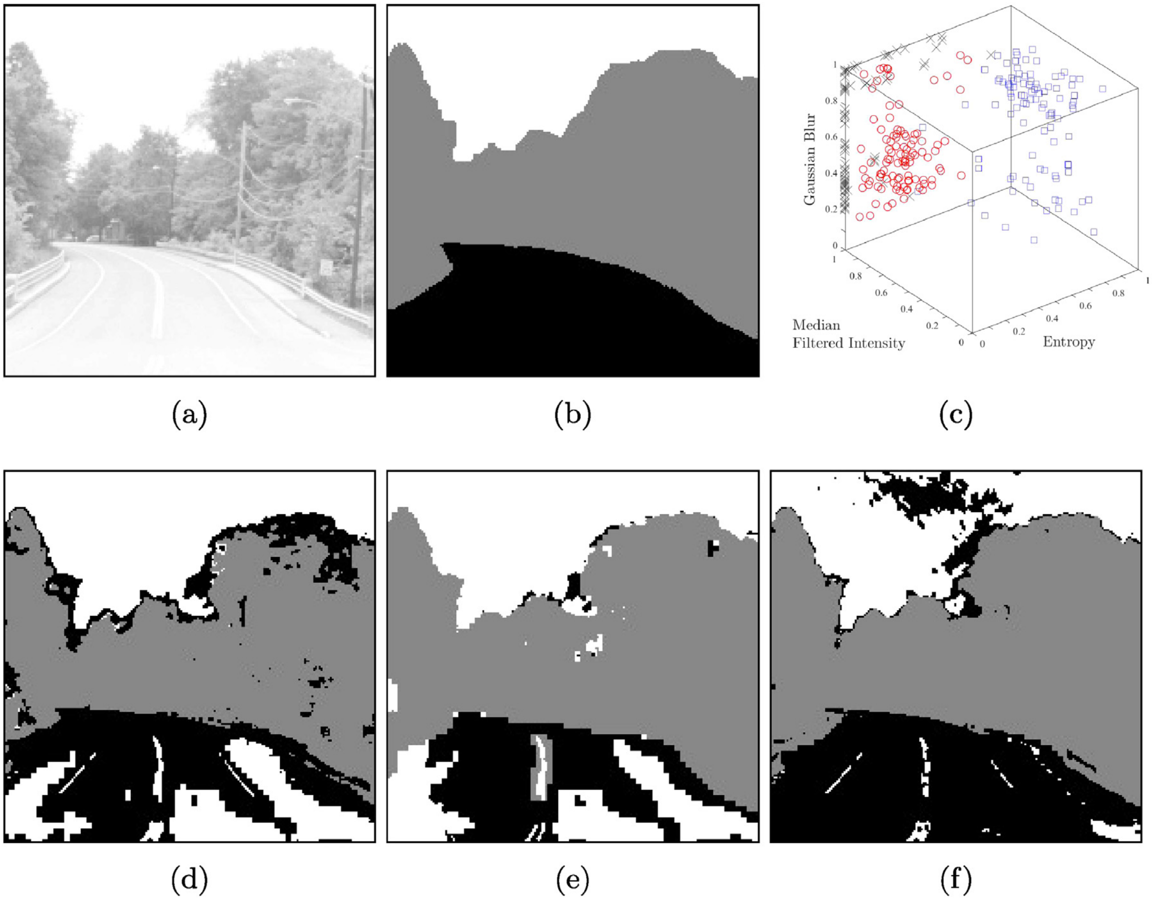


Fig. 11. Image segmentation for Scene 2: (a) original image, (b) mask image, (c) scatter plot of sample patterns, segmentation results for (d) fuzzy ART, (e) proposed T1 fuzzy ART with adaptive vigilance, and (f) proposed IT2 fuzzy ART.

Table 6
Recognition rate (R_Rate) and confusion matrix for Scene 1 (S1), when segmented into Road (R), Forest (F), and Sky (S), for fuzzy ART, T1 fuzzy ART with adaptive vigilance (Proposed T1), and IT2 fuzzy ART (Proposed IT2).

		Recognized class								
		Fuzzy ART			Proposed T1			Proposed IT2		
R_Rate (%)		79.96			90.22			96.26		
		R	F	S	R	F	S	R	F	S
True class	R	5790	2	2428	6084	3	2133	8204	16	0
	F	5031	19,870	545	1090	23,790	566	896	23,976	574
	S	10	0	6324	119	0	6215	10	0	6324

A trained network was used to classify all the pixels in Scene 1 and the same process is repeated for the other images (see Figs. 10–13). All the pixels are classified into three groups. The results for each method were compared with corresponding mask image of each scene. The confusion matrices given in Tables 6–9 suggest that our proposed algorithms (i.e., T1 fuzzy ART with adaptive vigilance and IT2 fuzzy ART) outperform the conventional fuzzy ART algorithm for this example as well. Table 10 gives a summary of the recognition rates for all four scenes. For each of the four scenes, a vigilance parameter $\rho = 0.35$ was used for the fuzzy ART algorithm. The fuzzifier m for the proposed T1 fuzzy ART was 1.1, whereas the fuzzifiers for the UMF and LMF of the proposed IT2 fuzzy ART were 1.7 and 0.9, respectively.

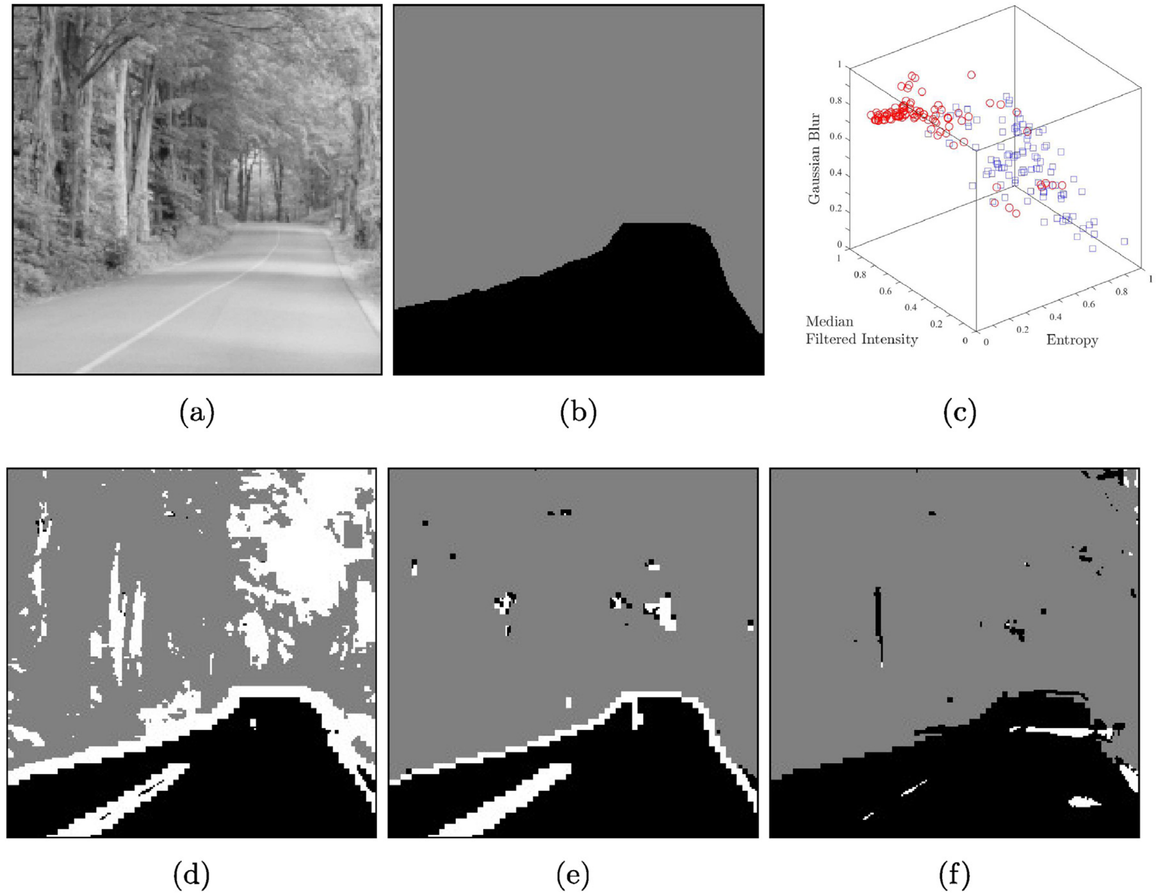


Fig. 12. Image segmentation for Scene 3: (a) original image, (b) mask image, (c) scatter plot of sample patterns, segmentation results for (d) fuzzy ART, (e) proposed T1 fuzzy ART with adaptive vigilance, and (f) proposed IT2 fuzzy ART.

Table 7
Recognition rate (R_Rate) and confusion matrix for Scene 2 (S2), when segmented into Road (R), Forest (F), and Sky (S), for fuzzy ART, T1 fuzzy ART with adaptive vigilance (Proposed T1), and IT2 fuzzy ART (Proposed IT2).

		Recognized class								
		Fuzzy ART			Proposed T1			Proposed IT2		
R_Rate (%)		80.09			85.02			92.38		
		R	F	S	R	F	S	R	F	S
True class	R	6805	14	3958	6012	810	3955	10,103	234	440
	F	3345	16,305	232	299	19,069	514	602	19,189	91
	S	415	0	8926	129	285	8927	1578	102	7661

Table 8
Recognition rate (R_Rate) and confusion matrix for Scene 3 (S3), when segmented into Road (R), Forest (F), and Sky (S), for fuzzy ART, T1 fuzzy ART with adaptive vigilance (Proposed T1), and IT2 fuzzy ART (Proposed IT2).

		Recognized class								
		Fuzzy ART			Proposed T1			Proposed IT2		
R_Rate (%)		76.12			92.78			96.37		
		R	F	S	R	F	S	R	F	S
True class	R	9442	25	2060	9245	521	1761	10,720	580	227
	F	98	21,007	7368	271	27,865	337	488	27,829	156

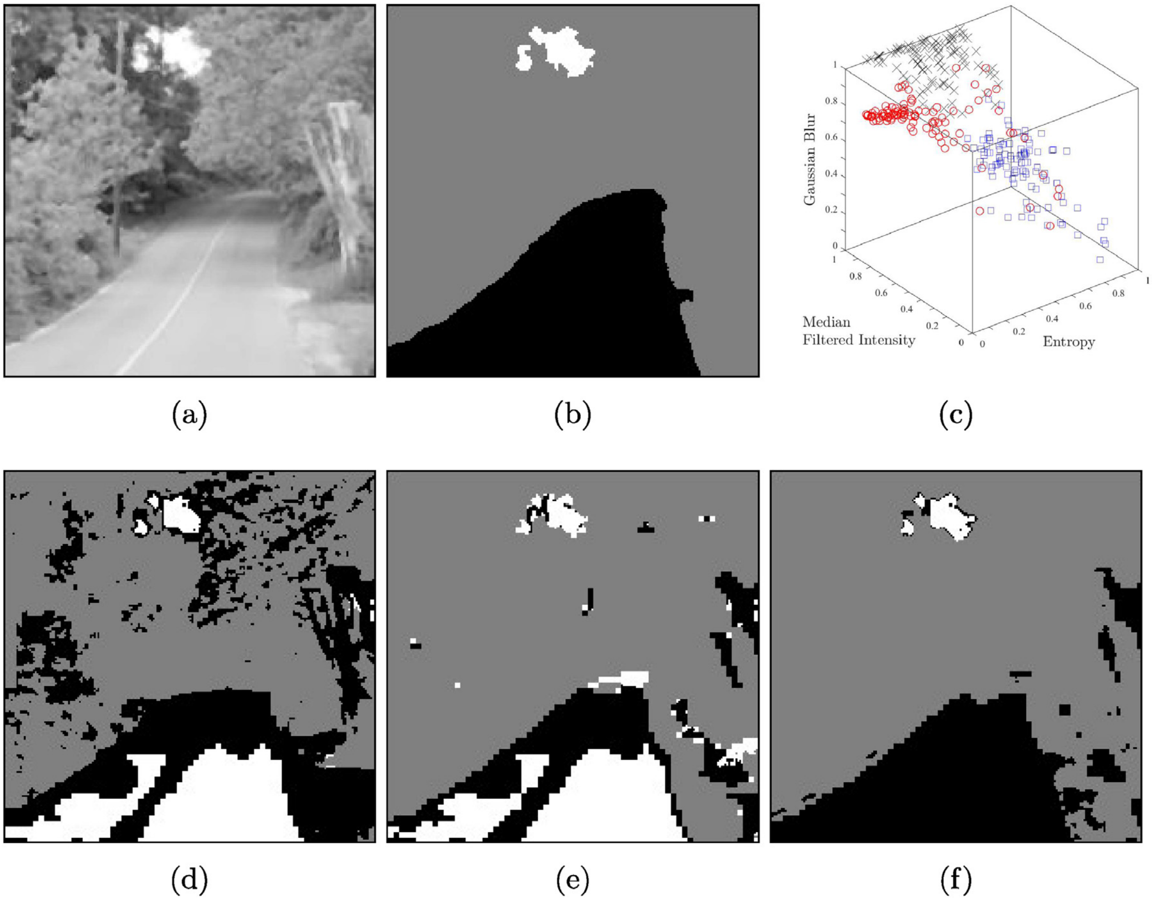


Fig. 13. Image segmentation for Scene 4: (a) original image, (b) mask image, (c) scatter plot of sample patterns, segmentation results for (d) fuzzy ART, (e) proposed T1 fuzzy ART with adaptive vigilance, and (f) proposed IT2 fuzzy ART.

Table 9
Recognition rate (R_Rate) and confusion matrix for Scene 4 (S4), when segmented into Road (R), Forest (F), and Sky (S), for fuzzy ART, T1 fuzzy ART with adaptive vigilance (Proposed T1), and IT2 fuzzy ART (Proposed IT2).

		Recognized class								
		Fuzzy ART			Proposed T1			Proposed IT2		
R_Rate (%)		65.76			78.85			91.30		
		R	F	S	R	F	S	R	F	S
True class	R	4831	975	4512	4020	1870	4428	8207	2111	0
	F	7902	21,131	49	1547	27,157	378	1171	27,911	0
	S	258	0	342	56	179	365	140	57	403

Table 10
Recognition rates for scenes S1, S2, S3, and S4, for fuzzy ART, T1 fuzzy ART with adaptive vigilance (Proposed T1), and IT2 fuzzy ART (Proposed IT2).

	Recognition rate (%)		
	Fuzzy ART	Proposed T1	Proposed IT2
S1	79.96	90.22	96.26
S2	80.09	85.02	92.38
S3	76.12	92.78	96.37
S4	65.76	78.85	91.30

6. Conclusion

ART-based clustering methods offer self-organization and simplicity. However, the fixed value of the vigilance parameter ρ in the algorithm limits the flexibility of the cluster generation, which inhibits the performance of the algorithm when the input pattern space consists of clusters of different volumes. In this paper, we proposed methods for modifying the conventional fuzzy ART algorithm by dynamically adapting and selecting the value of the vigilance parameter ρ for each pattern, as a function of its relative distance from the cluster center. The vigilance parameter in the proposed T1 fuzzy ART method was determined by defining it as a function of the fuzzifier m. To manage the uncertainty associated with m, we extended the algorithm to an IT2 fuzzy ART method using two values of fuzzifier m (i.e., m_1 and m_2).

It is to be noted that the use of our proposed T1 adaptive vigilance function in the fuzzy ART algorithm does not affect its on-line learning behaviour. However, our proposed IT2 fuzzy vigilance function requires the storage of previous input patterns to construct an IT2 FS and perform its type reduction. This additional requirement increases the computational complexity of the algorithm, although significant improvement in the learning capability of the network may be achieved with such an approach. In the various experiments using the three techniques, the IT2 fuzzy approach was found to be the most suitable for image segmentation applications.

It was found that the clustering performance with respect to the recognition rate improved up to 11% using our proposed T1 fuzzy ART, and up to 23% using our proposed IT2 fuzzy ART, for popular data sets. A similar result was also observed in image segmentation problems, where an improvement up to 25% was observed for various example images. We argue that several types of data obtained from the real world can be inherently uncertain, and our proposed methods may improve performance due to the ability of the integrated fuzzy sets in effectively modeling this uncertainty.

We believe that the methods proposed in this paper may be considered to be early attempts for effectively generating T2 fuzzy vigilance functions for pattern sets, and a more attentional vigilance membership function may further improve classification results. However, there are certain areas that may require some improvement.

First, in the proposed T1 fuzzy ART with adaptive vigilance, we require the selection of a proper vigilance function and the parameters of the selected function such that it performs optimally, since the function is critical in deciding the spatial nature of the vigilance value that may affect the clustering process. The geometric shape and parameters of the function are also critical for proper designing of the IT2 fuzzy ART method.

Second, the choice of fuzzifiers m_1 and m_2 for the proposed IT2 fuzzy algorithm is essential in obtaining the FOU. In general, if we use a combination of unsuitable fuzzifiers, IT2 fuzzy ART may yield poor clustering results compared with the conventional fuzzy ART due to inappropriate management of uncertainty of the available information. Therefore, the determination of m_1 and m_2 is an important research area in which several existing methods (e.g., neural networks and genetic algorithms) may be exploited.

Third, general T2 fuzzy MFs, using parameterized functions representing the secondary membership, could also be used to represent the sample data. For example, symmetric and asymmetric Gaussians may be used to model the data to add more flexibility than IT2 fuzzy MFs.

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] Y. Cai, J. Wang, Y. Tang, Y. Yang, An efficient approach for electric load forecasting using distributed ART (adaptive resonance theory) & HS-ARTMAP (hyper-spherical ARTMAP network) neural network, Energy 36 (2) (2011) 1340–1350.
[2] T. Caliński, J. Harabasz, A dendrite method for cluster analysis, Commun. Stat. Theor. Methods 3 (1) (1974) 1–27.
[3] G. Carpenter, Distributed learning, recognition, and prediction by ART and ARTMAP neural networks, Neural Netw. 10 (8) (1997) 1473–1494.
[4] G. Carpenter, S. Grossberg, A massively parallel architecture for a self-organizing neural pattern recognition machine, Comput. Vis. Graph. Image Process. 37 (1) (1987) 54–115.
[5] G. Carpenter, S. Grossberg, J. Reynolds, ARTMAP: supervised real-time learning and classification of nonstationary data by a self-organizing neural network, Neural Netw. 4 (5) (1991) 565–588.
[6] G. Carpenter, S. Grossberg, D. Rosen, Fuzzy ART: fast stable learning and categorization of analog patterns by an adaptive resonance system, Neural Netw. 4 (6) (1991) 759–771.
[7] O. Castillo, J. Castro, P. Melin, A. Rodriguez-Diaz, Application of interval type-2 fuzzy neural networks in non-linear identification and time series prediction, Soft. comput. 18 (6) (2014) 1213–1224.
[8] C. Chang, Z. Ye, M. Zhang, Fuzzy-ART based adaptive digital watermarking scheme, IEEE Trans. Circuits Syst. Video Technol. 15 (1) (2005) 65–81.
[9] B. Choi, F.C.-H. Rhee, Interval type-2 fuzzy membership function generation methods for pattern recognition, Inf. Sci. 179 (13) (2009) 2102–2122.
[10] S. Coupland, R. John, An investigation into alternative methods for the defuzzification of an interval type-2 fuzzy set, in: Proc. 2006 IEEE Int. Conf. on Fuzzy Systems, pp. 1425–1432.
[11] T. Dang, L. Ngo, W. Pedrycz, Interval type-2 fuzzy c-means approach to collaborative clustering, in: Proc. 2015 IEEE Int. Conf. on Fuzzy Systems, pp. 1–7.
[12] H. Frigui, C. Hwang, Fuzzy clustering and aggregation of relational data with instance-level constraints, IEEE Trans. Fuzzy Syst. 16 (6) (2008) 1565–1581.
[13] M. Girolami, Mercer kernel-based clustering in feature space, IEEE Trans. Neural Netw. 13 (3) (2002) 780–784.
[14] H. Hagras, A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots, IEEE Trans. Fuzzy Syst. 12 (4) (2004) 524–539.
[15] N. Hoa, T. Bui, An improved learning rule for fuzzy ART, J. Inf. Sci. Eng. 30 (3) (2015) 713–726.
[16] C. Hwang, F.C.-H. Rhee, An interval type-2 fuzzy c spherical shells algorithm, in: Proc. 2004 IEEE Int. Conf. on Fuzzy Systems, volume 2, pp. 1117–1122.
[17] C. Hwang, F.C.-H. Rhee, Uncertain fuzzy clustering: interval type-2 fuzzy approach to c-means, IEEE Trans. Fuzzy Syst. 15 (1) (2007) 107–120.

[18] P. Innocent, R. John, I. Belton, D. Finlay, Type-2 fuzzy representations of lung scans to predict pulmonary emboli, in: Proc. 2001 Joint 9th IFSA World Congress and 20th NAFIPS Int. Conf. volume 4, pp. 1902–1907.
[19] R. John, Type-2 inferencing and community transport scheduling, in: Proc. 1996 IEEE Int. Conf. on Fuzzy Systems, pp. 1369–1372.
[20] N. Karnik, J. Mendel, Applications of type-2 fuzzy logic systems: handling the uncertainty associated with surveys, in: Proc. 1999 IEEE Int. Conf. on Fuzzy Systems, volume 3, pp. 1546–1551.
[21] N. Karnik, J. Mendel, Centroid of a type-2 fuzzy set, Inf. Sci. 132 (1) (2001) 195–220.
[22] K. Kim, S. Kim, A passport recognition and face verification using enhanced fuzzy ART based RBF network and PCA algorithm, Neurocomputing 71 (16) (2008) 3202–3210.
[23] S. Lee, T. Chen, Using evolutionary computation approach to improve the performance of the fuzzy art for grouping parts, J. Chinese Inst. Ind. Eng. 18 (5) (2001) 55–62.
[24] F. Li, J. Zhan, Fuzzy adapting vigilance parameter of ART-II neural nets, in: Proc. 1994 IEEE Int. Conf. on Neural Networks, volume 3, pp. 1680–1685.
[25] C. Lin, C. Lin, An ART-based fuzzy adaptive learning control network, IEEE Trans. Fuzzy Syst. 5 (4) (1997) 477–496.
[26] O. Linda, M. Manic, Uncertainty-robust design of interval type-2 fuzzy logic controller for delta parallel robot, IEEE Trans. Ind. Inf. 7 (4) (2011) 661–670.
[27] R. Madeo, C. Lima, S. Peres, Gesture unit segmentation using support vector machines: segmenting gestures from rest positions, in: Proceedings of the 28th Annual ACM Symposium on Applied Computing, ACM, 2013, pp. 46–52.
[28] P. Melin, O. Castillo, A review on type-2 fuzzy logic applications in clustering, classification and pattern recognition, Appl. Soft. Comput. 21 (2014) 568–577.
[29] J. Mendel, Uncertain Rule-Based Fuzzy Logic system: Introduction and New Directions, Prentice–Hall PTR, 2001.
[30] J. Mendel, F. Liu, Super-exponential convergence of the karnik–mendel algorithms for computing the centroid of an interval type-2 fuzzy set, IEEE Trans. Fuzzy Syst. 15 (2) (2007) 309–320.
[31] S. Oh, H. Jang, W. Pedrycz, A comparative experimental study of type-1/type-2 fuzzy cascade controller based on genetic algorithms and particle swarm optimization, Expert Syst. Appl. 38 (9) (2011) 11217–11229.
[32] F. Olivas, F. Valdez, O. Castillo, P. Melin, Dynamic parameter adaptation in particle swarm optimization using interval type-2 fuzzy logic, Soft. comput. 20 (3) (2016) 1057–1070.
[33] W. Pedrycz, Proximity-based clustering: a search for structural consistency in data with semantic blocks of features, IEEE Trans. Fuzzy Syst. 21 (5) (2013) 978–982.
[34] W. Pedrycz, H. Izakian, Cluster-centric fuzzy modeling, IEEE Trans. Fuzzy Syst. 22 (6) (2014) 1585–1597.
[35] W. Pedrycz, K. Li, M. Reformat, Evolutionary reduction of fuzzy rule-based models, in: Fifty Years of Fuzzy Logic and its Applications, Springer, 2015, pp. 459–481.
[36] C. Peraza, F. Valdez, O. Castillo, Interval type-2 fuzzy logic for dynamic parameter adaptation in the harmony search algorithm, in: Proc. 2016 IEEE Int. Conf. on Intelligent Systems, 2016, pp. 106–112.
[37] V. Pham, L. Ngo, W. Pedrycz, Interval-valued fuzzy set approach to fuzzy co-clustering for data classification, KnowBased Syst. 107 (2016) 1–13.
[38] A. Ramakrishnan, Segmentation of Hand Gestures Using Motion Capture Data, University of California, Davis, 2011.
[39] F.C.-H. Rhee, Uncertain fuzzy clustering: insights and recommendations, IEEE Comput. Intell. Mag. 2 (1) (2007) 44–56.
[40] S. Roberts, R. Everson, I. Rezek, Maximum certainty data partitioning, Pattern Recognit. 33 (5) (2000) 833–839.
[41] L. Romdhane, H. Bannour, B. Ayeb, IMIOL: a system for indexing images by their semantic content based on possibilistic fuzzy clustering and adaptive resonance theory neural networks learning, Appl. Artif. Intell. 24 (9) (2010) 821–846.
[42] E. Rubio, O. Castillo, P. Melin, Interval type-2 fuzzy possibilistic c-means clustering algorithm, in: Recent Developments and New Direction in Soft–Computing Foundations and Applications, Springer, 2016, pp. 185–194.
[43] P. Simpson, Fuzzy min-max neural networks-part 2: clustering, IEEE Trans. Fuzzy Syst. 1 (1) (1993) 32–45.
[44] Z. Sun, L. Mak, K. Mao, W. Tang, Y. Liu, K. Xian, Z. Wang, Y. Sui, A knowledge-driven ART clustering algorithm, in: Proc. 2014 IEEE Int. Conf. on Software Engineering and Service Science, pp. 645–648.
[45] R. Tibshirani, G. Walther, T. Hastie, Estimating the number of clusters in a data set via the gap statistic, J. Royal Stat. Soc. 63 (2) (2001) 411–423.
[46] S. Tomida, T. Hanai, H. Honda, T. Kobayashi, Gene expression analysis using fuzzy ART, Genome Inf. 12 (2001) 245–246.
[47] E. Tsao, J. Bezdek, N. Pal, Fuzzy kohonen clustering networks, Pattern Recognit. 27 (5) (1994) 757–764.
[48] A. Wilson, A. Bobick, J. Cassell, Recovering the temporal structure of natural gesture, in: Proc. 1996 IEEE Int. Conf. on Automatic Face and Gesture Recognition, pp. 66–71.
[49] K. Žalik, An efficient k-means clustering algorithm, Pattern Recognit. Lett. 29 (9) (2008) 1385–1391.
[50] M. Zarandi, I. Türkşen, O. Kasbi, Type-2 fuzzy modeling for desulphurization of steel process, Expert Syst. Appl. 32 (1) (2007) 157–171.