Knowledge-Based Systems 133 (2017) 234–254

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

# Interpretable interval type-2 fuzzy predicates for data clustering: A new automatic generation method based on self-organizing maps

Diego S. Comas [a,b,*], Juan I. Pastore [a,b], Agustina Bouchet [a,b], Virginia L. Ballarin [b], Gustavo J. Meschino [c]

[a] Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, Argentina
[b] Digital Image Processing Lab, Instituto de Investigaciones Científicas y Tecnológicas en Electrónica (ICyTE), Facultad de Ingeniería, Universidad Nacional de Mar del Plata-CONICET, Juan B. Justo 4302, Mar del Plata (Zip Code B7608FDQ), Argentina
[c] Bioengineering Lab, Instituto de Investigaciones Científicas y Tecnológicas en Electrónica (ICyTE), Facultad de Ingeniería, Universidad Nacional de Mar del Plata-CONICET, Juan B. Justo 4302, Mar del Plata (Zip Code B7608FDQ), Argentina

ARTICLE INFO

*Article history:*
Received 17 February 2017
Revised 6 June 2017
Accepted 13 July 2017
Available online 14 July 2017

*Keywords:*
Fuzzy predicates
Interval type-2 fuzzy logic
Self-organizing maps
Interpretable clustering
Knowledge discovery

ABSTRACT

In previous works, we proposed two methods for data clustering based on automatically discovered fuzzy predicates which were referred to as SOM-based Fuzzy Predicate Clustering (SFPC) [Meschino et al., Neurocomputing, 147, 47–59 (2015)] and Type-2 Data-based Fuzzy Predicate Clustering (T2-DFPC) [Comas et al., Expert Syst. Appl., 68, 136–150 (2017)]. In such methods, fuzzy predicates allow both data clustering and knowledge discovering about the obtained clusters. This last feature constitutes novelty comparing to other existing approaches and it is a major contribution in the data clustering field. Based on these previous methods, in the present paper a new automatic clustering method based on fuzzy predicates is proposed which uses Self-Organizing Maps (SOMs) and is called Type-2 SOM-based Fuzzy Predicate Clustering (T2-SFPC). The new method does not require any prior knowledge about the clustering addressed. First, a random partition is defined on the dataset to be clustered and SOMs are configured and trained using the resulting data subsets. Second, an automatic clustering approach is applied on the SOM codebooks, discovering representative data of the different clusters, which are called cluster prototypes. Third, interval type-2 membership function formed by Gaussian-shape sub-functions and fuzzy predicates are defined, allowing data clustering and its interpretation. The proposed method preserves all the advantages of the previous methods SFPC and T2-DFPC in relation to the knowledge extraction capabilities and their potential application on distributed clustering and parallel computing, but results obtained on several public datasets tested showed more compactness and separation of the clusters defined by the T2-SFPC, outperforming both the previous methods and the several classical clustering approaches tested, considering internal and external validation indices. Additionally, both clustering interpretation and optimization capabilities are improved by the proposed method when compared to the methods SFPC and T2-DFPC.



## 1. Introduction

Data clustering refers to grouping data according to a similar criterion [1] revealing hidden structures in data. It has multiple applications in very different fields such as: data mining, marketing, machine learning, bioinformatics, image segmentation, pattern recognition, among others; and new methods are continually proposed [2–5]. Clustering methods are primarily designed to assign clusters to data, usually not requiring prior information about the expected results except for the number of clusters to be obtained, which is typically required. As a result, the outcome of the most common methods typically consists of a vector containing the corresponding cluster for each datum, including the cluster centroids or prototypes of the type of data corresponding to each of the clusters (called cluster prototypes).

Despite traditional approaches of data clustering are only used to group data; other potential applications can emerge. In fact, clustering can be addressed as a set of data analysis techniques

* Corresponding author at: Digital Image Processing Lab, Instituto de Investigaciones Científicas y Tecnológicas en Electrónica (ICyTE), Facultad de Ingeniería, Universidad Nacional de Mar del Plata-CONICET, Juan B. Justo 4302, Mar del Plata (Zip Code B7608FDQ), Argentina.

*E-mail addresses:* diego.comas@fi.mdp.edu.ar, diegoscomas@gmail.com (D.S. Comas), jpastore@fi.mdp.edu.ar (J.I. Pastore), abouchet@fi.mdp.edu.ar (A. Bouchet), vballari@fi.mdp.edu.ar (V.L. Ballarin), gmeschin@fi.mdp.edu.ar (G.J. Meschino).

http://dx.doi.org/10.1016/j.knosys.2017.07.012
0950-7051/

which discover groups of similar data and their results can be exploited extracting information about them. Such information could be related to what the common properties of the data inside a same cluster are and how these properties differ from a cluster to another [3].

In this regard, Fuzzy Logic (FL), conceived as a natural extension of Boolean logic which introduces degrees of truth between 0 (completely false) and 1 (completely true), is able to model linguistic expressions and concepts, including imprecision and vagueness, being excellent for modeling and implementing human reasoning expressed by linguistic expressions, achieving interpretable clustering.

Typical FL models are based on fuzzy inference systems using IF-THEN rules considering approaches such as Mamdani and Takagi–Sugeno–Kang [6–12] applied in image classification, image segmentation, speech recognition, control, among others. Although widely used, a fuzzy inference system requires defining fuzzification, aggregation and defuzzification operators and its outcome is a continuous variable. Considering data clustering applications, these characteristics of the fuzzy inference systems become difficult to understand the relation between the data and their properties and the system outcome, i.e. the assigned cluster [2,3]. On the other hand, models based on fuzzy predicates extend the Boolean predicates, modeling degrees of truth of predicates with values between 0 and 1 [13]. When applied on data clustering, fuzzy predicate models allow to implement knowledge about the clustering, explaining which values of each feature are related to each of the clusters and modeling these relationships using membership functions and predicates. Such models have been successfully applied in data clustering [2,3,14,15], having the following features:

- Each cluster is explained by a fuzzy predicate interpreted as *"The datum belongs to the cluster k"*, being $k$ a cluster; explaining which values (which characteristics) of each feature are related to each of the clusters.
- In order to assign clusters to data, degrees of truth of the predicates are computed for each datum using the membership functions and the fuzzy operators and in each case the cluster corresponding to the predicate with the maximum degree of truth is assigned to the datum.
- Resulting degrees of truth quantify in what grade each datum meets the characteristics required to belong to a cluster (i.e. how a datum is represented by their prototypes).

In the traditional approach, it is required knowledge of experts to define both the membership functions and the predicates. Once designed, the fuzzy predicates apply that knowledge for assigning clusters to data. Nevertheless, in recent works an approach alternative to the traditional one has been studied for data clustering which is based on the automatic generation of the membership functions and the fuzzy predicates by analyzing the data to be clustered. Such approach has an enormous and immediate advantage in relation with the traditional approach: it not only allows the data clustering, but also provides knowledge about the clustering obtained by interpreting the membership functions and the predicates generated. As a consequence, relevant information about data can be obtained, even when no prior information about the problem addressed is available [3].

In this regard, in the previous works [2,3] we proposed two methods for data clustering through fuzzy predicates in which membership functions and fuzzy predicates are automatically generated from the data to be clustered. In [2], a method based on Self-Organizing Maps (SOMs) (a set of wide known unsupervised and nonparametric neural networks with remarkable abilities for dealing with noise, outliers, and missing values) is proposed called SOM-based Fuzzy Predicate Clustering (SFPC). In the SFPC, a SOM is automatically trained and set and, then, Fuzzy C-Means (FCM) [16] clustering is applied to the codebook of the SOM, extracting cluster prototypes. From these prototypes, membership functions and fuzzy predicates are defined, linguistically explaining the clusters. The method includes a variant where several SOMs are generated from data subsets and predicates obtained from the different SOMs are combined, which could be applied to distributed clustering. Predicates are used to perform the data clustering and some analysis of the interpretation of the membership functions and predicates is given.

In [3], the method called Type-2 Data-based Fuzzy Predicate Clustering (T2-DFPC) is introduced. Unlike the SFPC, interval type-2 FL is used which defines a degree of truth by an interval in [0, 1] called interval of truth values, instead of a number between 0 and 1 as in the case of type-1 FL, which adds additional degrees of freedom considering data clustering. Interval type-2 FL provides more appropriate models than type-1 FL for dealing with vagueness and imprecision about the data characteristics and can reduce the effect on cluster assignments in data affected by noise [3,13,17]. In the T2-DFPC, the cluster prototypes are extracted directly from data without using SOMs, combining FCM with the Bayesian Information Criterion (BIC) [18,19], defining automatically the proper number of clusters in each case. Before the cluster prototype extraction, a random partition is performed on the data, obtaining disjoint subsets. The method is also suitable for distributed clustering. The T2-DFPC includes an analysis of the obtained membership functions and predicates, describing how the knowledge can be extracted. Additionally, it is also proposed a measure of intervals of truth values defining a methodology for interval comparing which allows the cluster assignment when interval type-2 fuzzy predicates are used.

Based on these two previous methods, in the present paper a new clustering method called Type-2 SOM-based Fuzzy Predicate Clustering (T2-SFPC) is proposed, which automatically generates interval type-2 membership functions and fuzzy predicates, allowing data clustering and knowledge discovery. The method proposed uses SOMs in order to obtain cluster prototypes, exploiting their advantages for noise, outliers, and missing values dealing, as in the SFPC; following the methodology used in that method for the automatic configuration and training of the SOMs. However, unlike the previous SFPC, in the method T2-SFPC, $M$ SOMs are automatically configured and trained from $M$ disjoint subsets defined by a random partition on the data, where $M \in \mathbb{N}$ is a method parameter. Once the $M$ SOMs are defined, the clustering approach combining FCM with the BIC is applied not requiring knowing the number of clusters to be obtained. Once cluster prototypes are extracted, interval type-2 membership functions and fuzzy predicates are generated in a different way to that proposed in the T2-DFPC [3]. Specifically, the new proposal includes parametrizable interval type-2 membership functions, i.e. it is possible the optimization of the parameters of the membership functions provided that a specific goal is defined, for instance adopting a clustering quality measure. As a result of the proposed method, one fuzzy predicate is defined for each cluster. The clustering assignment is performed using the methodology introduced in [3] by means of the measure of interval of truth values.

The contribution of the proposed method is a new general methodology for data clustering, which can be applied to most of the clustering problems. The interval type-2 membership functions merge all knowledge extracted of the cluster prototypes from the $M$ SOMs. The method T2-SFPC preserves all the characteristics of the previous methods, mainly those related to the SOM abilities for discovering natural data groupings and, also, to the knowledge discovery capabilities studied in the T2-DFPC. As in the T2-DFPC, linguistic expressions extracted from the predicates can be adapted to match the terminology of the domain experts, not requiring any prior knowledge about the dataset or the clustering problem addressed. Tests performed considering widely different datasets reveal better results from the proposed T2-SFPC than those obtained both from the SFPC, the T2-DFPC and classical clustering methods, meaning that the proposed method is an excellent clustering method choice when both data clustering and knowledge extraction from the clustering results are needed.

The rest of this paper is structured as follow. In Section 2, it is presented an analysis of the main existing papers concerning to SOMs used for data analysis as well as FL applications. Important concepts related to SOMs and interval type-2 FL are presented in Section 3 and, after that, the method proposed called T2-SFPC is explained in detail. In Section 4, experiments performed to the assessment of the proposed method are described and their results are presented, including an example of the interpretation of the obtained clustering in the case of segmentation of brain magnetic resonance images. Finally, in Sections 5 and 6, discussion and conclusions are presented, commenting on the results and the limitations of the method proposed as well as future work.

## 2. Related works

In the present Section, both some existing clustering approaches based on SOMs and some applications of FL concerned to the method proposed are described. Given the wide number of papers related to these issues, the present analysis is intended to cover the most relevant papers on the topic. Nevertheless, further reviews can be consulted in [2,7,11,20]. Descriptions of the methods SFPC and T2-DFPC are omitted, as these were given in the previous Section.

As it has been mentioned, in order to generate membership functions and fuzzy predicates from data, it is necessary to extract representative examples of data which describe each cluster to be obtained. Such examples can be achieved by applying data analysis approaches and using their results as cluster prototypes [2,3].

A very interesting and widely applied data analysis method consists of SOMs, which define a mapping from data to cells in the SOM space. There are several papers proposing different clustering approaches based on SOMs. In general, they include two-level clustering schemes where a SOM is trained with the available data in the first level. The second level clustering is performed using another SOM fixing the number of cells to the number of clusters [21], crisp-clustering methods [2,22–24], or fuzzy clustering methods such as FCM [25]. Grouping prototype vectors contained in a SOM instead of directly grouping data has some advantages. First, some clustering algorithms might have high computational cost even considering small number of data [22]. Second, as the prototypes in a SOM are computed averaging data, they are less sensitive to noise effects. Third, considering problems with few data, a SOM can generalize the data space making easier the discovery of clusters than in the case of a clustering method applied directly on the data. Some SOM-based clustering approaches are summarized below.

A multilayer SOM approach is proposed in [21] by Lampinen and Oja. The authors assure that unsupervised learning techniques such as SOMs can reduce the training required in complex data clustering problems. The method allows to find the most representative prototype vectors for a given dataset and, also, to make a topological mapping from the feature space to the map space, preserving the input topographic space. A two-level SOM-clustering algorithm is studied in Vesanto and Alhoniemi [22], using agglomerative clustering in the second level. Each datum in the original dataset belongs to the same cluster than its nearest prototype. It is indicated that this approach can reduce the computational cost of clustering revealing that, in most of the cases, it is better to group the codebook of a trained SOM than to apply clustering directly on data.

More complex clustering approaches based on SOMs are reported in other papers. In [26], Kiang extends SOMs in order to include a contiguity-constrained clustering method to perform clustering based on the cell map generated by a SOM. The method considers an agglomerative approach to recursively merging groups from a SOM until a desired number of clusters is reached. In [27], Sarlin and Eklund apply a SOM-FCM two-level scheme, performing the clustering on the SOM codebook and assigning a partial membership of each cell to each cluster, obtaining overlapping clusters. This approach allows to analyze the membership degree in the self-discovered clusters using SOMs. In [24], Ortiz et al. propose a SOM-based method for segmentation of brain magnetic resonance images. The proposal includes image acquisition and preprocessing, feature extraction (including texture features), feature selection based on genetic algorithms and image segmentation using SOMs. It is proposed a SOM-clustering approach defining cluster constrains by considering the map space and the relationship with the input space (data space) using an entropy-gradient function computation to group the cells of the SOM codebook. The SOM quality is evaluated considering quantification and topographic errors. In [23], Taşdemir et al. indicate that SOMs find an optimal distribution of the prototype vectors in the data space such that they better approximate the unknown density distribution of the data. As a result, once a SOM is trained, it contains knowledge about the training data providing relevant information about the clusters, which is extracted by means of explanatory visualization or clustering. The authors affirm that typical visualization methods require expert knowledge in order to interpret the information contained in a trained SOM. Therefore, they propose an automated clustering method for SOMs using hierarchical agglomerative clustering and a connectivity matrix.

In all the SOM-based methods previously described, SOMs are used for clustering and analysis of data by means of very different approaches which define clusters on the SOM codebooks. The methods were applied on data corresponding to different problems showing that SOM is a proper approach when data analysis and exploration is required. In the method proposed in the present work (T2-SFPC), this characteristic of SOM is exploited in order to obtain cluster prototypes from the SOM codebooks, i.e. examples of data corresponding to different clusters, but unlike the previous mentioned methods, the extracted cluster prototypes are not used to group the original dataset. Instead, they are used to generate membership functions and fuzzy predicates defining a FL system which allows both clustering of the original dataset and knowledge extraction about the clustering problem in the form of linguistic expressions. In other words, in the method proposed, SOMs are used for data analysis in order to generate an interpretable FL system for clustering.

In relation to FL applied to data clustering and classification, most of existing applications are based on inference systems with IF-THEN rules designed from experts' knowledge, see for instance the works analyzed in the surveys [7] and [11]. However, the generation of models based on FL disregarding experts' knowledge is also addressed by some authors, including methods based on SOMs.

In [10], Mansoori proposes a fuzzy rule-based clustering algorithm in order to perform an unsupervised cluster analysis. The author states that the main limitation of both fuzzy and crisp clustering algorithms is their sensitivity to the number and the initial positions of the clusters. Additionally, he explains that the discovered knowledge from typical approaches is not easily understandable for humans. The proposed algorithm tries to find the potential clusters and to identify them with some interpretable fuzzy rules.

Bodenhofer and Bauer [28] make a general analysis of methods for membership functions interpretation, observing that when membership functions are defined, especially whether it is done by means of an automatic method, it is important to preserve the semantic meaning of the natural language expressions in order to achieve interpretable descriptions. In other words, for example, a membership function associated to the attribute *"low"* should correspond to lower values than a membership function associated to the attribute *"high"*. Moreover, the authors explain that interpretability is the key property of a FL system. As a result, according to the authors, if a FL system is not focused on the interpretability, then it can be replaced by methods which are more abstract and computationally less expensive. In this regard, in [29] Zadeh indicates that the meaning of the natural language expressions is contained both in membership functions and fuzzy predicates and, therefore, their interpretation are strongly related to the knowledge.

SOMs are used by Drobics et al. in [20] in order to extract linguistic and interpretable descriptions from clusters, proposing a three-stage approach which combines SOMs and FCM for extracting cluster prototypes. Once cluster prototypes are defined, the authors propose to analyze the prototypes feature by feature, considering membership functions located in predefined positions. Finally, predicates are defined by means of an optimization algorithm considering the attributes described by the membership functions and different fuzzy operators. This method differs significantly from the method proposed in the present work and, also, from the previous method SFPC presented in [2]. Specifically, both the methods SFPC and the T2-SFPC define the membership functions considering the centroids and the standard deviation of the extracted cluster prototypes, which means that they preserve the relationships between the values of the different features observed in each cluster when fuzzy predicates are defined. In other words, in these methods, the positions, shape and width of the membership functions depend on the properties observed in the cluster prototypes. In this regard, analyzing values of each feature independently of the others as it is proposed by Drobics et al. implies disregarding the original notion of a cluster, i.e. a set of data with similar properties.

## 3. Methods

In this Section, concepts related both with SOMs and interval type-2 FL in data clustering are revised. As both topics are well known, only the most important concepts are presented. Then, the method proposed called Type-2 SOM-based Fuzzy Predicate Clustering (T2-SFPC) is explained in detail.

### *3.1. Self-organizing maps*

In 1982, Kohonen proposed the SOM [30] which consists of a regular grid of cells mapping an input space (data space) to a cell space (output space), preserving the topology of the input space and having remarkable abilities to remove noise, to detect outliers, and to complete missing values in data [2].

Each cell of a SOM is associated to a vector, called prototype vector, which has a size equal to the dimension of the input space. The set of all the prototype vectors is called codebook. Before the training stage, the codebook is typically initialized using linear, random or data-analysis-based initialization [31]. During the training phase, the SOM codebook is adjusted in order to map close data in the input space to close cells in the map. It is expected that in a well-trained SOM, the codebook represents the training dataset, preserving its characteristics, i.e. having similar probabilistic density function. This last feature is used in the method proposed in this work as well it was used in the previous work [2], in order to obtain cluster prototypes from SOMs applying a second level clustering on previously configured and trained SOMs, considering 2-D maps.

How good a SOM codebook represents the training dataset statistics depends both on the map size, topology, neighborhood function and training type, which are parameters that should be selected for each dataset. This selection is generally done heuristically. However, according to existing works [2,32,33], it is possible to obtain an optimal SOM for a given dataset by training several SOMs with different parameters and computing SOM error measures after the training process. In this sense, in the method SFPC presented in [2], it is proposed to consider three different error measures to obtain an optimal SOM for a dataset, which are: quantization error, topographic error, and topographic product. Formal definition of these error measures can be found in [2], but the next conceptual assertions are given here:

- The quantization error allows to know whether the prototype vectors in a trained SOM are close to the training dataset, considering only distance measures on the data space.
- The topographic error is a measure of how good a SOM preserves the topology of the data space. i.e. whether very close prototype vectors of the SOM codebook have been assigned to adjacent cells in the map space.
- The topographic product combines distances both in the data space and in the map space, allowing to assess how good the neighborhood relations in a SOM are preserved.

If the prototype vectors perform an organized projection of the training data according to a similarity criterion and preserving the data topology, then these three errors tend to be minimized [2]. In the present work, in order to guarantee a good representation of the training dataset in the cluster prototypes extracted from SOMs, which are used to generate membership functions and fuzzy predicates, the automatic procedure proposed in [2] for the configuration and training of SOMs is used. Specific application of this procedure in the method proposed in the present work is given in Section 3.3.

### *3.2. Interval type-2 fuzzy predicates in data clustering*

FL was conceived as a natural extension of Boolean logic, defining degrees of truth of logic propositions with values between 0 (false) and 1 (true) [34]. Typically, FL is selected when dealing with linguistic statements and working with concepts described by vague expressions are required [3]. FL has been applied to data clustering and classification applications both using experts' knowledge to generate fuzzy models as well as proposing automatic methods based on data analysis. The papers [2,3,14,15] are examples of such applications and, also, the surveys [7,11].

Traditional notion of FL, called type-1 FL, defines a degree of truth as a real number between 0 (completely false) and 1 (completely true). Values between these two limits indicate that the logic preposition is not completely false or completely true, which is known as gradualism principle of the FL [34]. According to recent works [3,11,35], defining degrees of truth by numbers could not be enough in problems with great imprecision, disagreement between different experts to define degrees of truth resulting in imprecise knowledge, or when data is affected by noise; which may occur in data clustering problems. In contrast to type-1 FL, interval type-2 FL or interval-valued FL which is the same concept allows to define degrees of truth by intervals of truth values which are able to model both variability in data and knowledge and showed better performance than type-1 FL on data clustering and classification [3,7,11].

In particular, in [3] interval type-2 FL was considered to automatically generate interval type-2 fuzzy predicates from data, allowing both data clustering and extracting knowledge about clustering. As it was previously mentioned, on the basis of that approach, interval type-2 FL and fuzzy predicates are used in the method proposed in the present work, but unlike the previous work in the present approach fuzzy predicates are defined using cluster prototypes obtained from SOMs.

In order to introduce the notation used in the rest of this paper, the most relevant definitions related to interval type-2 FL and fuzzy predicates are given below. Further analysis of the concepts treated here can be consulted in [2,3,13]. After introducing the concepts, an example of application of experts' knowledge and fuzzy predicates is analyzed to clarify, considering the tissue detection in brain magnetic resonance images.

The next definitions are adopted [3]:

**Definition 3.2.1.** An interval of truth values is an interval $A = [a_L, a_R]$, with $0 \le a_L \le a_R \le 1$, which defines the degree of truth of a logic expression when interval type-2 FL is used.

**Definition 3.2.2.** A fuzzy predicate $p(\mathbf{x})$, where $\mathbf{x}$ indicates a set of objects or variables, is a declarative sentence which assigns one or more properties to $\mathbf{x}$. The value taken by the fuzzy predicate $p(\mathbf{x})$, noted by $\nu(p(\mathbf{x}))$, is called degree of truth.

**Definition 3.2.3.** An interval type-2 fuzzy predicate $p(\mathbf{x})$ is a fuzzy predicate whose degree of truth $\nu(p(\mathbf{x}))$ is an interval of truth value $A_{p(\mathbf{x})} = [a_{p(\mathbf{x}),L}, a_{p(\mathbf{x}),R}]$, where $a_{p(\mathbf{x}),\,L}$ and $a_{p(\mathbf{x}),\,R}$ are the endpoints of the interval.

**Definition 3.2.4.** An interval type-2 membership function $\bar{\mu}$ defined on an universe $X$, is a function $\bar{\mu} : X \to \chi$, where $\chi$ is the set of all the possible intervals of truth values, i.e. $\chi = \{[a_L, a_R]/a_L \le a_R \wedge a_L, a_R \in [0, 1]\}$. A membership function associates values of a variable with degrees of truth.

From the point of view of fuzzy predicates, an interval type-2 membership function $\bar{\mu}$ defines with what degree of truth different values taken by a variable satisfy an attribute described for $\bar{\mu}$. As a result, a membership function can be interpreted as mathematically equivalent to a fuzzy predicate associated to the same variable and attribute that $\bar{\mu}$.

**Definition 3.2.5.** The functions $\varphi^-_{\bar{\mu}} : X \to [0, 1]$ and $\varphi^+_{\bar{\mu}} : X \to [0, 1]$ are respectively the lower membership function and the upper membership function of $\bar{\mu}$, defined as follow:

$$\varphi^-_{\bar{\mu}}(x) = \min(\bar{\mu}(x)), \ \forall x \in X, \tag{1}$$

$$\varphi^+_{\bar{\mu}}(x) = \max(\bar{\mu}(x)), \ \forall x \in X. \tag{2}$$

**Definition 3.2.6.** The Footprint of Uncertainty (FOU) of $\bar{\mu}$ is the set of all points between the lower membership function and the upper membership function, i.e.:

$$FOU_{\bar{\mu}} = \bigcup_{x \in X} \left\{ \left[ \varphi^-_{\bar{\mu}}(x), \varphi^+_{\bar{\mu}}(x) \right] \right\}. \tag{3}$$

Considering the last definition, the FOU is associated with the vagueness or the variability around of the attribute described by an interval type-2 membership function. In [3] where the method T2-DFPC is presented, it is proposed to analyze the shape and size of the FOUs of the interval type-2 membership functions generated from the method in order to obtain information about the clusters and the attributes discovered. Additionality, in other works, different measures of fuzziness had been defined operating with the FOU [36].

**Definition 3.2.7.** As in general fuzzy predicates can relate one or more variables with properties, fuzzy predicates can be both simple or compound. A simple predicate directly associates a variable with an attribute and its degree of truth is usually obtained through a membership function. On the other hand, a compound fuzzy predicate combines logically two or more simple predicates using conjunctions ($\wedge$), disjunctions ($\vee$) and complements ($\neg$), which in a wide sense are known as fuzzy aggregation operators.

Since in interval type-2 FL the degrees of truth are interval of truth values, compound predicates are evaluated using fuzzy conjunctions $C$: $[0, 1]^n \to [0, 1]$, disjunctions $D$: $[0, 1]^n \to [0, 1]$, and complements $c$: $[0, 1] \to [0, 1]$ applied on the ends of the intervals of truth values [3,13].

In the method T2-SFPC proposed in the present work, compound fuzzy predicates obtained exclusively consider conjunctions of simple predicates. Therefore, the conjunction between interval type-2 fuzzy predicates is formally defined in the next paragraphs.

Let $p(\mathbf{x})$ and $q(\mathbf{y})$ be two interval type-2 fuzzy predicates respectively with degrees of truth $\nu(p(\mathbf{x})) = A_{p(\mathbf{x})} = [a_{p(\mathbf{x}),L}, a_{p(\mathbf{x}),R}]$ and $\nu(q(\mathbf{y})) = A_{q(\mathbf{y})} = [a_{q(\mathbf{y}),L}, a_{q(\mathbf{y}),R}]$. It is important to clarify that as $p(\mathbf{x})$ and $q(\mathbf{y})$ are respectively defined on the variables $\mathbf{x}$ and $\mathbf{y}$ which are not necessary the same, a new predicate obtained combining $p(\mathbf{x})$ and $q(\mathbf{y})$ will be defined on the union of $\mathbf{x}$ and $\mathbf{y}$, which here will be noted as $\mathbf{z}$. The conjunction between $p(\mathbf{x})$ and $q(\mathbf{y})$ is defined as follow:

**Definition 3.2.8.** The conjunction between $p(\mathbf{x})$ and $q(\mathbf{y})$ defines a new fuzzy predicate $r(\mathbf{z}) \equiv p(\mathbf{x}) \wedge q(\mathbf{y})$ whose degree of truth is computed as:

$$\nu(r(\mathbf{z})) = \nu(p(\mathbf{x}) \wedge q(\mathbf{y})) = \left[ C\left(a_{p(\mathbf{x}),L}, a_{q(\mathbf{y}),L}\right), C\left(a_{p(\mathbf{x}),R}, a_{q(\mathbf{y}),R}\right) \right], \tag{4}$$

where $C$: $[0, 1]^2 \to [0, 1]$ is a fuzzy conjunction [13].

In the literature, a wide set of fuzzy aggregation operators have been proposed. Selecting different fuzzy operators should be made according to the properties of each operator and how predicates are interpreted and evaluated by the experts in each application. Considering previous results on data clustering [2,3], in the present work three different fuzzy operators are considered: compensatory logic operations: Geometric Mean Based Compensatory Fuzzy Logic (GMCFL) and Arithmetic Mean Based Compensatory Fuzzy Logic (AMCFL), and standard triangular norms (MIN-MAX). Formal definitions and analysis of these fuzzy operators can be found in [2,13,37,38].

Typically, in data clustering application each cluster is described by a compound fuzzy predicate $p_k(\mathbf{x})$ linguistically interpreted as *"The datum* $\mathbf{x}$ *belongs to cluster k"*. Cluster assignment is performed by determining which cluster has the predicate with the highest degree of truth [2,3]. Therefore, if interval type-2 FL is used, it is required a method for comparing intervals of truth values. In [3] a novel methodology for interval comparing in data clustering using interval type-2 fuzzy predicates was proposed which is based on the concept of measure of interval of truth values. This methodology is used in the present work in order to assign clusters to data, once fuzzy predicates obtained from the T2-SFPC have been evaluated. Both the concept of measure of interval of truth value and the clustering assignment method originally proposed in [3] are now recalled, considering their importance for the clustering method proposed in the present paper:

**Definition 3.2.9.** Let $\chi$ be the set of all the closed intervals contained in [0, 1], i.e. the set of all the possible intervals of truth values. The measure of interval of truth values $f : \chi \to \mathbb{R}^+$ is:

$$f(A) = f([a_L, a_R]) = \frac{a_L + a_R}{2} a_R, \tag{5}$$

where $A = [a_L, a_R]$ is an interval of truth values.

The measure $f$ describes the degree of truth of an interval of truth values with a number, mapping from the interval space to $\mathbb{R}^+$. The measure combines the mean value of the interval with its maximum and allows to induce an order on intervals. The higher the value of $f(A)$, the higher the degree of truth represented by the interval. The next properties are satisfied by $f$:

- If $A = [0, 0]$ (the minimum interval of truth values) then $f(A) = 0$ (the minimum of $f$).
- If $A = [1, 1]$ (the maximum interval of truth values) then $f(A) = 1$ (the maximum of $f$).
- Given two intervals of truth values $A = [a_L, a_R]$ and $B = [b_L, b_R]$ where $a_L = a_R = a$, $b_L = b_R = b$, and $a < b$, then $f([a, a]) = a^2 < f([b, b]) = b^2$, following the order in $\mathbb{R}$ for $a$ and $b$.
- $f$ induces a transitive order on intervals of truth values ($<_f$).

**Definition 3.2.10.** Let $\{p_k(\mathbf{x})\}_{k=1,\ldots,K}$ be a set of $K$ compound fuzzy predicates used for data clustering, each describing one of the clusters $\{1, \ldots, K\}$, with $\mathbf{x} \in [-1, 1]^d$ a generic datum (considering normalized data), where $p_k(\mathbf{x})$ is linguistically interpreted as *"The datum* $\mathbf{x}$ *belongs to cluster* $k$*"*. A datum $\mathbf{x}' \in [-1, 1]^d$ is assigned to one of the $K$ clusters by means of the set of predicates $\{p_k(\mathbf{x})\}_{k=1,\ldots,K}$ as follow:

*Step 1:* Compute $\nu(p_k(\mathbf{x}))$, $k = 1, \ldots, K$, for $\mathbf{x} = \mathbf{x}'$, obtaining the intervals of truth values $\{\nu(p_k(\mathbf{x}'))\}_{k=1,\ldots,K} = \{[a_{p_k(\mathbf{x}'),L}, a_{p_k(\mathbf{x}'),R}]\}_{k=1,\ldots,K}$.

*Step 2:* Apply the measure of intervals of truth values $f$ on the intervals $\{\nu(p_k(\mathbf{x}'))\}_{k=1,\ldots,K} = \{[a_{p_k(\mathbf{x}'),L}, a_{p_k(\mathbf{x}'),R}]\}_{k=1,\ldots,K}$ defined in the previous step, obtaining the values:

$$\left\{f\left(\nu(p_k(\mathbf{x}'))\right)\right\}_{k=1,\ldots,K} = \left\{f\left(\left[a_{p_k(\mathbf{x}'),L}, a_{p_k(\mathbf{x}'),R}\right]\right)\right\}_{k=1,\ldots,K}. \tag{6}$$

*Step 3:* Assign to the datum $\mathbf{x}'$ the cluster corresponding to the highest measure of intervals of truth values $\{f(\nu(p_k(\mathbf{x}')))\}_{k=1,\ldots,K}$, i.e. to assign the cluster $k' \in \{1, \ldots, K\}$ to $\mathbf{x}'$ where:

$$f\left(\nu(p_{k'}(\mathbf{x}'))\right) = \max\left\{f\left(\nu(p_k(\mathbf{x}'))\right)\right\}_{k=1,\ldots,K}. \tag{7}$$

Considering the Definitions 3.2.9 and 3.2.10, the next analysis can be done:

- If a datum $\mathbf{x}$ is close to the properties of a cluster $k$, then the mean value of the interval resulting of evaluate $p_k(\mathbf{x})$ is high. As a result, as the properties of the cluster are better met by the datum, the value taken by $f$ for the interval of truth values defined by $\nu(p_k(\mathbf{x}))$ increases. Therefore, the procedure given in the Definition 3.2.10 assigns to the datum $\mathbf{x}'$ the cluster $k' \in \{1, \ldots, K\}$ whose properties are best met by the datum.
- Considering two intervals of truth values with the same mean value, resulting of the predicate evaluation for a datum $\mathbf{x}'$ in two different clusters, $f$results the highest value of the measure for the interval with the highest maximum value, which means the properties of the corresponding cluster are the best met by the datum. Therefore, in the case of two intervals with same mean value for a same datum in two different clusters, the procedure of the Definition 3.2.10 assigns to the datum the cluster whose properties are the best met.

In the next paragraphs, an example of the use of experts' knowledge and fuzzy predicates for data labeling is analyzed.

Pixel labeling in brain magnetic resonance images is a problem widely studied from very different approaches, including fuzzy predicates generated using experts' knowledge [15]. The problem consists in the segmentation of the magnetic resonance images by grouping its pixels in three clusters, each corresponding to one of the brain tissues: gray matter, white matter, and cerebrospinal fluid. Typically, when pixel labeling is performed from magnetic resonance images, medical experts analyze gray intensities of one or more sequences of magnetic resonance images, such as PD, T1, and T2. Each sequence describes different physical responses of the tissues when magnetic fields are applied on the human body.

In the case of brain magnetic resonance images, typically, experts analyze gray intensities of images corresponding to the sequences PD, T1, and T2 in order to decide which tissue corresponds to each pixel which requires to thoroughly understand how different pixels are represented in the different sequences. This problem is addressed in [15] where authors propose a framework for brain magnetic image segmentation based on type-1 fuzzy predicates and experts' knowledge and several medical experts were consulted for defining both membership functions and fuzzy predicates.

Experts' knowledge is expressed as linguistic descriptions, defining both attributes on the gray intensity, such as *"bright"*, *"dark"*, *"gray"*, and how these are related to determine the tissue corresponding to each pixel of the image. Once relevant attributes are identified, in consultation with the experts a membership function is defined for each attribute, describing with what degree of truth each value of the gray intensity satisfies the attribute. A compound fuzzy predicate describing each tissue is defined. Given a pixel to be assigned to a tissue, all the compound predicates are evaluated considering the gray intensities of the pixel in the sequences PD, T1, and T2 and using the membership functions. The pixel is assigned to the tissue for which the degree of truth of the corresponding predicate is the highest, as detailed in the Definition 3.2.10. In Fig. 1a diagram of the use of experts' knowledge to form fuzzy predicates and membership functions is presented for this concrete application.

### *3.3. Proposed method: Type-2 SOM-based fuzzy predicate clustering (T2-SFPC)*

The method T2-SFPC consists of four stages: A) Random dataset partition, B) Configuration and training of SOMs, C) Extraction of cluster prototypes, and D) Generation of an interval type-2 fuzzy predicate system.

Given a dataset to be clustered, a random data partition is applied at the first stage, obtaining $M$ disjoint subsets ($M \in \mathbb{N}$). Each of these subsets contains some of the samples (data) in the original dataset, i.e. the subsets act as different realizations of the same process which defines the dataset. At stage B, $M$ SOMs are automatically configured and trained, one for each of the subsets previously generated at stage A. At stage C, a second level clustering is performed on the $M$ SOM codebooks obtaining the cluster prototypes, combining crisp FCM clustering with the BIC in order to define the proper number of cluster in each case and following the method proposed in [19], successfully applied in [3]. Finally, at stage D, the cluster prototypes are analyzed generating interval type-2 fuzzy membership functions and fuzzy predicates, explaining the discovered clusters and allowing a clustering on the original dataset. In Fig. 2, a pipeline of the method T2-SFPC is shown.

As it was mentioned in Section 1, the method T2-SFPC is inspired in the methods SFPC and T2-DFPC, proposed respectively in [2] and [3]. The T2-SFPC is a new method for data clustering which automatically generates interval type-2 fuzzy predicates for data clustering, having the next features:

- Subsets of the initial dataset are used to automatically train SOMs and cluster prototypes are extracted from the SOMs in an analogous way to that suggested in the SFPC, but in the T2-SFPC the number of clusters is defined by means of the BIC.
- The interval type-2 fuzzy predicates are generated from cluster prototypes, but unlike the method T2-DFPC the interval type-2 membership functions are parametrizable, i.e. the T2-SFPC allows the application of optimization methods if a specific goal

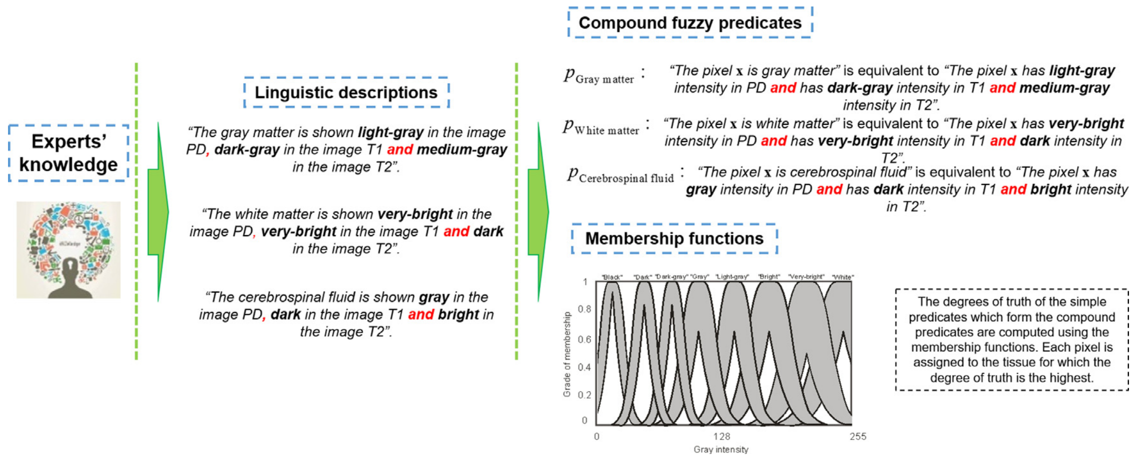


**Fig. 1.** Diagram of the use of experts' knowledge to form fuzzy predicates and membership functions for pixel labelling in brain magnetic resonance images detecting gray matter, white matter and cerebrospinal fluid. Attributes used by the experts are indicated in bold. Linguistic connectives are shown in red font. Using the descriptions, a compound fuzzy predicate is defined describing each tissue. Membership functions are also defined in accordance with experts describing the degree of truth in which each value of gray intensity satisfies each attribute. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

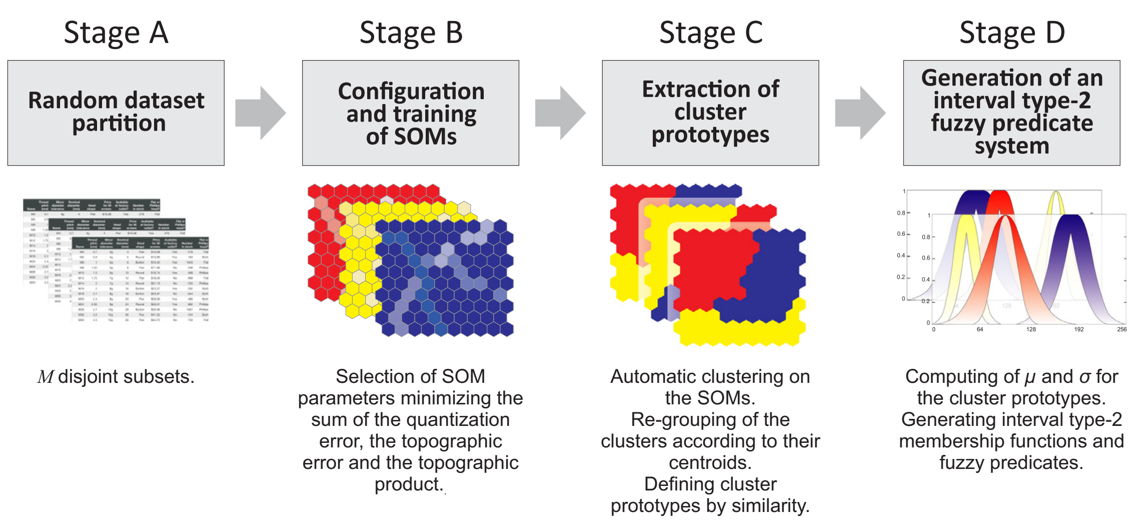


**Fig. 2.** Pipeline of the method T2-SFPC. Stage A: Random dataset partition, obtaining $M$ disjoint subsets. Stage B: Configuration and training of SOMs. Stage C: Extraction of cluster prototypes. Stage D: Generation of an interval type-2 fuzzy predicate system.

is defined, which could be, for instance, a clustering quality measure.

- Each cluster is explained by one fuzzy predicate $p_k(\mathbf{x})$ interpreted as *"The datum* $\mathbf{x}$ *belongs to the cluster k"* and its degree of truth is computed by the logic combination of simple predicates defined by the interval type-2 membership functions generated from the method.
- The cluster assignment is performed following the method described at the end of Section 3.2, which was originally proposed in [3].
- The obtained clustering is linguistically interpretable, i.e. an expert is able to give linguistic meaning to the membership functions and the predicates automatically discovered, preserving all the knowledge extraction characteristics of the method T2-DFPC proposed in [3].

Each of the stages of the method T2-SFPC is described in detail below. Hereinafter, $\mathbf{X} \subset [-1, 1]^d$ represents a normalized dataset to be clustered, where $d$ is the dimension of the data space (the number of features) and $N$ is the number of data in $\mathbf{X}$. A datum in $\mathbf{X}$ is a $d$-uple $\mathbf{x} = (x_1, x_2, \ldots, x_d)$, with $\mathbf{x} \in \mathbf{X}$, $x_i \in [-1, 1]$, $i = 1, \ldots, d$.

### 3.3.1. Stage A: Random dataset partition

Let $\mathbf{X} \subset [-1, 1]^d$ be the initial-dataset which defines the data to be clustered. In the present stage, a random partition is applied on the set $\mathbf{X}$, defining $M$ disjoint subsets of $\mathbf{X}$, noted by $\{\mathrm{P}_m\}_{m=1,\ldots,M}$, where $\mathbf{X} = \mathrm{P}_1 \cup \mathrm{P}_2 \cup \ldots \cup \mathrm{P}_M$ and $M \in \mathbb{N}$. This partition process was also implemented at the first stage of the method T2-DFPC presented in [3].

It is assumed that all data are generated for the same process. Data represent a sample of a problem from the real world. If data are statistically representative, then each partition will be also representative (provided it contains a proper quantity of data), and it will contain similar information, but not exactly the same. In other words, the data contained in each of the subsets $\mathrm{P}_m$, $m = 1, \ldots, M$, are samples randomly selected without reposition (resampling) from the data contained in the original dataset $\mathbf{X}$ and, therefore, their characteristics will differ from a subset to another. This process acts considering diversity in the dataset $\mathbf{X}$. The results obtained from clustering on each of the subsets will operate as different "opinions" of how clustering should be, represented as cluster prototypes.

This random data partition presents two advantages for the method proposed:

- The quantity of data in each of the resulting subsets is $M$ times less than in the whole dataset. As a consequence, the computational cost involved in the analysis of the data in each subset, i.e. in the subsequent stages of the method, is dramatically reduced, and, in addition, the subsets processing could be performed in a parallel system, i.e. all subsets can be processed at the same time.
- This process could be not required if the data are physically separated in distinct locations. This makes the method suitable to be applied in distributed datasets in which in general it is required the processing of large volume of data, as suggested in the previous works [2,3], which constitutes an interesting feature.

The partition size $M$ is a method parameter that should be selected according to the number of data in $\mathbf{X}$. As a requirement, $M$ should be selected considering the quantity of data in each subset enough to represent the original population in $\mathbf{X}$. Considerations referred to the proper selection of the value of $M$ will be given later.

In the next stage, previous to the extraction of the cluster prototypes, the data in the subsets $\{\mathrm{P}_m\}_{m=1,\dots,M}$ are used to train and set $M$ different SOMs, one for each $\mathrm{P}_m$, taking advantages of the known abilities of the SOMs for noise and outliers removing and missing value dealing, previously studied and exploited in [2].

#### 3.3.2. Stage B: Configuration and training of SOMs

In this stage, the data contained in the subsets $\{\mathrm{P}_m\}_{m=1,\dots,M}$ generated at stage A are used in order to obtain $M$ distinct SOMs, each one trained with the data of one of the datasets $\mathrm{P}_m,\ m = 1,\dots,M$. Once the $M$ SOMs are defined, the codebook of a given trained SOM $m$, $m \in \{1, 2, ..., M\}$ will consist of prototype vectors which are representative vectors of the data in the training dataset $\mathrm{P}_m$. As the subsets $\mathrm{P}_m,\ m = 1,\dots,M$ were defined to consider diversity of the data in $\mathbf{X}$, it is expected that each codebook reveals different characteristics about the data in $\mathbf{X}$.

In order to obtain in each SOM a good representation of the training data in the codebook, i.e., obtain codebook's vectors with probability density functions consistent with the training data in $\mathrm{P}_m$; various SOMs are trained for each $\mathrm{P}_m$, using different combinations of quantity of cells and topologies. In this way, an optimal SOM is selected for each $\mathrm{P}_m$ considering the minimization of the sum of the error measures introduced in Section 3.1, i.e. the sum of the quantization error, the topographic error, and the topographic product, as it was considered in [2] to select the optimal SOM for a given training dataset.

The concrete application of such procedure in the method proposed is given below:

*Step 1:* For each subset $\mathrm{P}_m$ resulting of the stage A, SOMs with different combinations of quantity of cells and topologies are trained to select the optimal one.

From $\eta = 5\,N^{0.5}$, an estimated number of cells obtained with this heuristic formula [39], SOMs with $\eta$, $2\eta$, $3\eta$ and $4\eta$ cells are considered for each subset $\mathrm{P}_m$, in each case both with hexagonal and rectangular grid topologies considering 2D maps. As a result, 8 distinct SOMs are trained, 4 for each type of grid topology, all with the same training dataset $\mathrm{P}_m$.

In all cases, Gaussian neighborhood functions and batch training are used.

Even when the SOM is very robust with respect to its initialization, fast convergence is sought. In order to achieve this requirement, all SOMs are initialized by the procedure known as linear initialization, which proposes an ordered initial state of the codebook. In this method, proposed by Kohonen [40,41], the initial codebook is constituted by a regular, two-dimensional sequence of vectors, a mesh, taken along the $d$-dimensional hypercube with limits taken by minimum and maximum values of the training data in $\mathrm{P}_m$. The axis of the mesh are the eigenvectors corresponding to the two largest eigenvalues of the training data (considering a 2D SOM). In the SOM implementation used in this work, the eigenvectors are computed by the Gram–Schmidt procedure [42].

*Step 2:* For each $\mathrm{P}_m$, it is selected the SOM for which the sum of the quantization error, the topographic error, and the topographic product is the lowest.

This procedure is repeated for the $M$ subsets $\{\mathrm{P}_m\}_{m=1,\dots,M}$ obtaining $M$ distinct SOMs, one for each dataset $\mathrm{P}_m,\ m = 1,\dots,M$. The optimal parameters resulting for one of the $M$ SOMs can differ from the optimal parameters resulting for other of the $M$ SOMs, considering they were trained with different datasets. As a result, each subset $\mathrm{P}_m$ is associated to a SOM codebook whose data, which will be noted by $\Phi_m$, are representative of the data in $\mathrm{P}_m$ used as training data of the SOM.

The prototypes vectors in each $\Phi_m$, $m = 1,\dots,M$, are used in the next stage in order to extract the cluster prototypes.

#### 3.3.3. Stage C: Extraction of cluster prototypes

In this stage, in the same way done at stage #2 of the method T2-DFPC proposed in [3], cluster prototypes are extracted for the data in each subset $\mathrm{P}_m$. However, unlike the T2-DFPC, in the present work the cluster prototypes are extracted from the data in the codebooks $\{\Phi_m\}_{m=1,\dots,M}$ corresponding to the $M$ SOMs configured and trained in the previous stage. As a result, cluster prototypes obtained here, which will be used to generate interval type-2 membership functions, are different to the cluster prototypes obtained using the method T2-DFPC in which the prototypes are extracted directly from the subsets $\{\mathrm{P}_m\}_{m=1,\dots,M}$. This is one of the main difference of the method T2-SFPC proposed in the present paper compared to the previous proposal T2-DFPC. As cluster prototypes are different between both methods, clustering results will be also different.

The clustering of the data in each $\Phi_m$, $m = 1,\dots,M$, is performed using FCM with crisp clustering as clustering algorithm [16] and the BIC [18] for the automatic determination of the proper number of clusters, following the method developed in [19], which was also successfully applied in [3]. This automatic clustering approach was not considered in the original SFPC proposed in [2].

As a result, the SOM codebooks $\{\Phi_m\}_{m=1,\dots,M}$ are clustered using the clustering approach based on FCM-BIC. Considering one of the SOMs, whose codebook is the set $\Phi_m$, the resulting clustering could be noted as $\{\Delta_{m,j}\}_{j=1,\dots,\kappa_m}$, where $\Phi_m = \Delta_{m,1} \cup \dots \cup \Delta_{m,j} \cup \dots \cup \Delta_{m,\kappa_m}$ and $\kappa_m \in \mathbb{N}$ is the number of clusters defined on $\Phi_m$, noting that the number of clusters obtained for the different codebooks could differ because the result of the automatic clustering scheme depends on the data in each codebook which come from distinct SOMs. The set of cluster centroids obtained from the codebook clustering is represented by $\{\Upsilon_{m,j}\}_{\substack{m=1,\dots,M\\ j=1,\dots,\kappa_m}}$. The number of clusters for the whole original dataset $\mathbf{X}$ is the maximum of $\{\kappa_m\}_{m=1,\dots,M}$, symbolized by $K$, $K \in \mathbb{N}$.

After clustering on SOM codebooks is performed, a second clustering is run on all the clusters centroids $\{\Upsilon_{m,j}\}_{\substack{m=1,\dots,M\\ j=1,\dots,\kappa_m}}$, using a single FCM with crisp clustering fixing the number of clusters to $K$. As a result, all the centroids are clustered in $K$clusters according to their similarity. The sets of re-clustered centroids are noted by $\{\Omega_{l,k}\}_{\substack{l=1,\dots,L_k\\ k=1,\dots,K}}$, where $L_k$ is the number of clusters defined on the codebooks $\{\Phi_m\}_{m=1,\dots,M}$ using the FCM-BIC clustering whose centroids were now reassigned to the cluster $k$.

In an upper step, based on the results of the clustering of the centroids, the data in each cluster of each codebook are reassigned in order to obtain the cluster prototypes. Let $\Delta_{m,j}$ be the prototype vectors of the SOM $m$ grouped in the cluster $j$ resulting of the

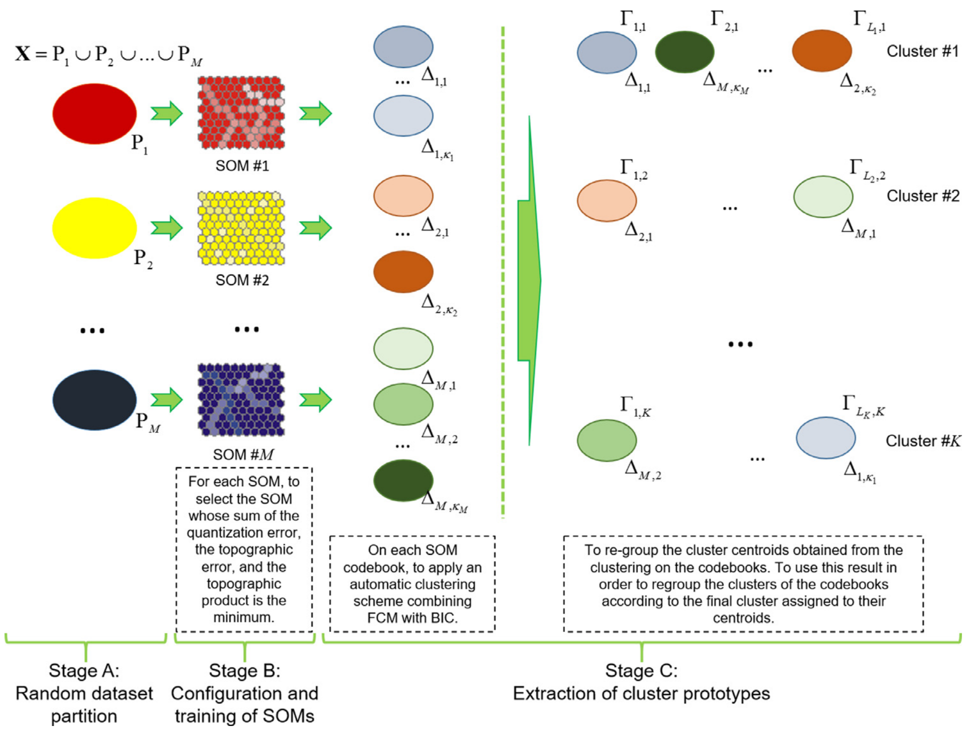


**Fig. 3.** Detailed diagram of the stages A, B, and C of the method T2-SFPC proposed. After a random data partition is performed, $M$ distinct SOMs are automatically configured and trained. Then, cluster prototypes are extracted by means of an automatic FCM-BIC clustering scheme applied on the SOMs codebooks.

application of the FCM-BIC clustering approach and let $Y_{m,j}$ their centroid, where $m \in \{1, ..., M\}$ and $j \in \{1, ..., \kappa_m\}$. If in the second clustering step the centroid $Y_{m,j}$ was assigned to the cluster $k$ and now it is the centroid $\Omega_{l,k}$, then the prototype vectors $\Delta_{m,j}$ are now the set of cluster prototypes $\Gamma_{l,k}$. This procedure is followed for all the sets of grouped codebooks $\{\Delta_{m,j}\}_{j=1,\dots,\kappa_m}$ defining the sets of cluster prototypes $\{\Gamma_{l,k}\}_{\substack{l=1,\dots,L_k\\ k=1,\dots,K}}$.

As a result, the cluster prototypes are extracted from the SOMs configured and trained at stage B. Each set of cluster prototypes $\Gamma_{l,k}$ for a given cluster $k$, describes the characteristics of this cluster according to the prototypes vectors which was assigned to $\Gamma_{l,k}$. Consequently, different descriptions of a same cluster can be obtained from the sets $\Gamma_{l,k}$ varying $l$ in 1, ..., $L_k$ for a fixed $k$. These descriptions act as "different opinions" of the cluster $k$ capturing variability and diversity of the data inside the cluster. In Fig. 3a detailed diagram of the stages A, B, and C of the method T2-SFPC is shown.

In the next stage, the cluster prototypes are used to generate interval type-2 membership functions and fuzzy predicates allowing to perform the clustering of the data in **X** and to interpret its result.

### 3.3.4. Stage D: Generation of an interval type-2 fuzzy predicate system

In this stage, the cluster prototypes $\{\Gamma_{l,k}\}_{\substack{l=1,\dots,L_k\\ k=1,\dots,K}}$ obtained at stage C are analyzed in order to define interval type-2 membership functions and fuzzy predicates, generating an interval type-2 fuzzy predicate system which is able to perform data clustering. Once the predicates are defined, it is possible to apply a clustering of the data in **X**, or new data related to the same process which generated **X**, following the procedure described in Section 3.2. The clustering obtained is linguistically interpretable, preserving the same extraction knowledge capabilities of the T2-DFPC proposed and analyzed in [3].

In previous works [2,3], once cluster prototypes are extracted, type-1 Gaussian-shape membership functions are generated by analyzing the clusters centroids and the standard deviation of the prototypes. In particular, in the method T2-DFPC [3], in a first step, type-1 Gaussian membership functions are defined from the cluster prototypes and, then, these functions are aggregated using fuzzy operators in order to obtain interval type-2 fuzzy membership functions.

Despite of the high accuracy of the clustering results obtained for the method T2-DFPC, only numeric values of the membership functions (samples at fixed distances) are known and stored in vectors during the method implementation, consequently, their functional expressions are not known. Therefore, it is not possible to perform an optimization of the membership function parameters, considering, for instance, a clustering quality measure.

Unlike the procedure proposed in the method T2-DFPC [3], in the present work, interval type-2 membership functions formed by Gaussian-shape sub functions are adopted to be generated from the cluster prototypes in the present stage. Considering a feature $i$ and a cluster $k$, the interval type-2 membership function is noted by $\bar{\mu}_{i,k} : [-1, 1] \to \chi$, where $\chi$ represents the set of all the possible intervals of truth values as it was defined in Section 3.2. In Fig. 4, a graphical representation of the $\bar{\mu}_{i,k}$ adopted is shown and its functional expression is given below:

$$\bar{\mu}_{i,k}(x) = \left[\varphi^-_{\bar{\mu}_{i,k}}(x), \varphi^+_{\bar{\mu}_{i,k}}(x)\right], \tag{8}$$

$$\varphi^-_{\bar{\mu}_{i,k}}(x) = \begin{cases} e^{-(x-\omega_1)^2/2(\varsigma_1{}^2)} & \text{if } x \le \rho \\ e^{-(x-\omega_2)^2/2(\varsigma_2{}^2)} & \text{if } x > \rho \end{cases}, \tag{9}$$

$$\varphi^+_{\bar{\mu}_{i,k}}(x) = \begin{cases} e^{-(x-\omega_3)^2/2(\varsigma_3{}^2)} & \text{if } \ x < \omega_3 \\ e^{-(x-\omega_4)^2/2(\varsigma_4{}^2)} & \text{if } x > \omega_4 \\ 1 & \text{else} \end{cases}, \tag{10}$$

with $x \in [-1, 1]$, $\omega_1$ to $\omega_4$ are the center of Gaussian sub functions which define $\bar{\mu}_{i,k}$ and $\varsigma_1$ to $\varsigma_4$ are the parameters controlling the

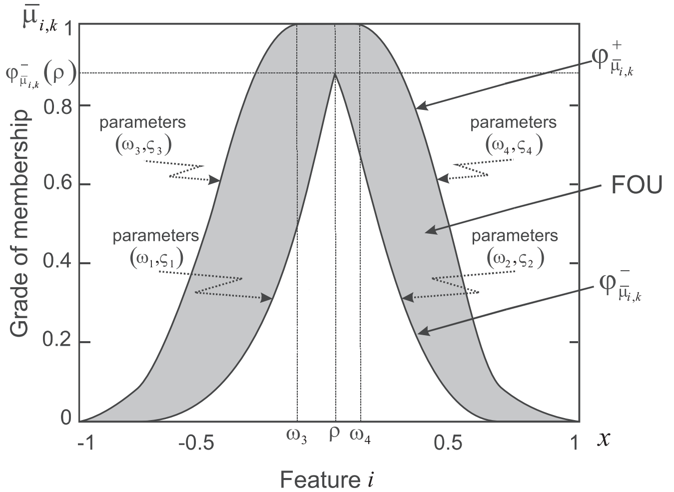


**Fig. 4.** Interval type-2 membership function formed by Gaussian-shape sub functions proposed in the method T2-SFPC for a feature $i$ and a cluster $k$. The parameters $\omega_1$ to $\omega_4$ and $\varsigma_1$ to $\varsigma_4$ are respectively the centers and the width parameters of the Gaussian sub functions.

width of the Gaussian functions. As it is shown in the graph of the Fig. 4, $\rho$ is the value of $x$ for which the value of the lower membership function $\varphi^-_{\bar{\mu}_{i,k}}$ is maximum. The parameters $\omega_1$ to $\omega_4$ and $\varsigma_1$ to $\varsigma_4$ are obtained considering known values for the expected $\bar{\mu}_{i,k}$ and the solution for the parameters of a Gaussian function when values taken by the function are known. In the next paragraphs, this solution is present and, then, the specific procedure proposed for determining the parameters of $\bar{\mu}_{i,k}$ is presented in detail.

The parameters $\omega$ and $\varsigma$ for a Gaussian-shape function $\mu(x) = e^{-(x-\omega)^2/2(\varsigma^2)}$, i.e. the center and the width, can be computed knowing two values taken by the function for two different values of $x$. Let $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$, with $y_1 = \mu(x_1)$ and $y_2 = \mu(x_2)$, two pair of points of the solution expected for $\mu(x)$. The next equation system can be defined:

$$\begin{cases} \mu(x_1) = y_1 = e^{-(x_1-\omega)^2/2(\varsigma^2)} \\ \mu(x_2) = y_2 = e^{-(x_2-\omega)^2/2(\varsigma^2)} \end{cases}. \tag{11}$$

From the previous equation system, it is possible to arrive to the next solutions for $\omega$ and $\varsigma$:

$$\omega_{a,b} = \frac{2x_1 - 2\eta x_2 \pm \sqrt{4\eta(x_1-x_2)^2}}{2-2\eta}, \tag{12}$$

$$\varsigma_{a,b} = \sqrt{\frac{-\left(x_1-\omega_{a,b}\right)^2}{2\ln(y_1)}} = \sqrt{\frac{-\left(x_2-\omega_{a,b}\right)^2}{2\ln(y_2)}}, \tag{13}$$

where $\eta = \frac{\ln(y_1)}{\ln(y_2)}$. The previous equations allow to obtain two solutions for the values of $\omega$ and $\varsigma$, i.e. $\omega = \omega_a \wedge \varsigma = \varsigma_a$ or $\omega = \omega_b \wedge \varsigma = \varsigma_b$, meaning there are two possible Gaussian-shape functions which take the values $y_1$ and $y_2$ for $x = x_1$ and $x = x_2$ respectively. A final solution for $\omega$ and $\varsigma$ can be found considering a third point $P_3 = (x_3, y_3)$ and evaluating $\mu(x_3)$ for the parameters $(\omega_a,\ \varsigma_a)$ and $(\omega_b,\ \varsigma_b)$. Therefore, the final solution corresponds to the pair of parameters for which $\mu(x_3)$ is closest to $y_3$. As a result, this procedure allows to define the parameters of a Gaussian-shape function on the basis of three known points which should be approximated for the expected solution. It should be noted that the final solution will take exactly the values $y_1$ in $x = x_1$ and $y_2$ in $x = x_2$ and a value close to $y_3$ in $x = x_3$.

Based on the considerations given in the previous paragraphs, the interval type-2 membership functions and the fuzzy predicates are generated from the cluster prototypes extracted at stage C following the next four steps:

*Step 1:* Create type-1 membership functions by analyzing the data contained in the cluster prototypes $\{\Gamma_{l,k}\}_{\substack{l=1,\dots,L_k \\ k=1,\dots,K}}$. Taking into account the methods SFPC and T2-DFPC presented in [2] and [3], type-1 Gaussian membership functions are used in the present work. For each feature $i$ and each cluster $k$, the cluster centroids of the prototypes, noted by $\{\Omega_{l,k}\}_{\substack{l=1,\dots,L_k \\ k=1,\dots,K}}$, are assigned to the centers of Gaussian type-1 membership functions $\{\mu_{i,l,k}\}_{\substack{i=1,\dots,d \\ l=1,\dots,L_k \\ k=1,\dots,K}}$, with $\mu_{i,l,k} : [-1,1] \to [0,1]$, and the width parameters are defined by standard deviation of the prototype vectors in each $\{\Gamma_{l,k}\}_{\substack{l=1,\dots,L_k \\ k=1,\dots,K}}$, which are noted by $\{\sigma_{i,l,k}\}_{\substack{i=1,\dots,d \\ l=1,\dots,L_k \\ k=1,\dots,K}}$. As a result, each Gaussian type-1 membership function is defined by $\mu_{i,l,k}(x) = e^{-(x-\omega_{i,l,k})^2/2(\sigma_{i,l,k}{}^2)}$, where $\omega_{i,\ l,\ k}$ is the component $i$ of the centroids $\Omega_{l,\ k}$, $x \in [-1,1]$.

This process is repeated for each feature and cluster. As there are $L_k$ data subsets for each cluster, $L_k$ type-1 Gaussian membership functions are defined for each cluster and feature.

The standard deviation defines the width of each type-1 membership function, describing how the degree of truth decreases when the values of the feature for a specific datum moves away from the cluster centroid. The maximums of the membership functions are located in the positions of the cluster centroids.

It is important to note that each Gaussian membership function describes how the values taken for a feature are related to a cluster, i.e. with what degree of truth a datum is closed to, or meets, the properties observed in the cluster, according to the information obtained from the cluster prototypes. Each Gaussian membership function contributes a description, having $L_k$ membership functions, i.e. descriptions, of each cluster and feature.

*Step 2:* Combine the $L_k$ descriptions for each cluster $k = 1, \dots, K$ and each feature $i = 1, \dots, d$ aggregating the $L_k$ Gaussian type-1 membership functions previously obtained in order to generate an interval type-2 membership function noted by $\hat{\bar{\mu}}_{i,k} : [-1,1] \to \chi$ defining a bounded area, which describes the variability of the prototypes extracted for each cluster and feature. This membership function $\hat{\bar{\mu}}_{i,k}$ is an approximate solution to the expected membership function $\bar{\mu}_{i,k}$ and it is obtained using MIN-MAX fuzzy operators as follow:

$$\varphi^-_{\hat{\bar{\mu}}_{i,k}}(x) = \min\left(\mu_{i,1,k}(x), \dots, \mu_{i,L_k,k}(x)\right), \tag{14}$$

$$\varphi^+_{\hat{\bar{\mu}}_{i,k}}(x) = \max\left(\mu_{i,1,k}(x), \dots, \mu_{i,L_k,k}(x)\right), \tag{15}$$

$\forall x \in [-1,1]$, $i \in \{1, \dots, n\}$, and $k \in \{1, \dots, K\}$; where $\varphi^-_{\hat{\bar{\mu}}_{i,k}} : [-1,1] \to [0,1]$ and $\varphi^+_{\hat{\bar{\mu}}_{i,k}} : [-1,1] \to [0,1]$ are respectively the lower and the upper membership function of $\hat{\bar{\mu}}_{i,k}$.

*Step 3:* Generate for each $\hat{\bar{\mu}}_{i,k}$, $k = 1, \dots, K$, $i = 1, \dots, d$, an interval type-2 membership function $\bar{\mu}_{i,k}$ formed by Gaussian-shape sub functions, as it was previously defined and shown in Fig. 4, computing the parameters $\omega_1$ to $\omega_4$ and $\varsigma_1$ to $\varsigma_4$ as follow:

- For the left part of the lower membership function of $\bar{\mu}_{i,k}$, noted by $\varphi^-_{\bar{\mu}_{i,k}}$, define the pairs of points $P_1 = (x_1, y_1)$, $P_2 = (x_2, y_2)$, and $P_3 = (x_3, y_3)$ as the points where the left part of the approximate lower membership function $\varphi^-_{\hat{\bar{\mu}}_{i,k}}$ take the values closest to 0.05, 1, and 0.5 times $\varphi^-_{\hat{\bar{\mu}}_{i,k}}(\rho)$, being $\rho$ the value for which $\varphi^-_{\hat{\bar{\mu}}_{i,k}}$ is maximum, as it is shown in Fig. 4. These specific values are considered to define the points $P_1$ to $P_3$ in order to cover the known range for the expected solution of the left part of $\varphi^{-}_{\bar{\mu}_{i,k}}$, therefore, 0.05 times $\varphi^{-}_{\bar{\mu}_{i,k}}(\rho)$ is close to the minimum of the expected solution, which asymptotically tends to 0, the value 1 covers the maximum of known values for the function and 0.5 times $\varphi^{-}_{\bar{\mu}_{i,k}}(\rho)$ is used because as it is located in the middle of the range it is a proper point to operate as third point in order to decide between the two possible combinations of values for $\omega$ and $\varsigma$ obtained from the Eqs. (12) and (13). Using these three points $P_1$ to $P_3$, compute the values of the parameters $\omega_1$ and $\varsigma_1$ corresponding to the left part of $\varphi^{-}_{\bar{\mu}_{i,k}}$, following the procedure previously explained to obtain the parameters of a Gaussian-shape function.

- For the right part of the lower membership function of $\bar{\mu}_{i,k}$, noted by $\varphi^{-}_{\bar{\mu}_{i,k}}$, define the pairs of points $P_1 = (x_1, y_1)$, $P_2 = (x_2, y_2)$, and $P_3 = (x_3, y_3)$ as the points where the right part of the approximate lower membership function $\varphi^{-}_{\hat{\bar{\mu}}_{i,k}}$ take the values closest to 0.05, 1, and 0.5 times $\varphi^{-}_{\bar{\mu}_{i,k}}(\rho)$, based on the same observations followed before for the left part of $\varphi^{-}_{\bar{\mu}_{i,k}}$. Considering these three pairs of points $P_1$ to $P_3$, compute the values of the parameters $\omega_2$ and $\varsigma_2$ corresponding to the right part of $\varphi^{-}_{\bar{\mu}_{i,k}}$, following the procedure previously explained.
- For the left part of the upper membership function of $\bar{\mu}_{i,k}$, noted by $\varphi^{+}_{\bar{\mu}_{i,k}}$, define the pairs of points $P_1 = (x_1, y_1)$, $P_2 = (x_2, y_2)$, and $P_3 = (x_3, y_3)$ as the points where the left part of the approximate upper membership function $\varphi^{+}_{\hat{\bar{\mu}}_{i,k}}$ take the values closest to 0.05, 0.95, and 0.5. As in the case of the lower membership function, in the present case these specific values for determining the points $P_1$ to $P_3$ are selected in order to cover the range of the expected solution for the left part of $\varphi^{+}_{\bar{\mu}_{i,k}}$. Specifically, the value 0.05 is close to the minimum of the expected $\varphi^{+}_{\bar{\mu}_{i,k}}$, which asymptotically tends to 0, 0.95 is close to its maximum and 0.5 (corresponding to middle of the range of $\varphi^{+}_{\bar{\mu}_{i,k}}$) is used to define the third point for deciding between the two solutions for the parameter values obtained from the Eqs. (12) and (13). The value 0.95 is adopted instead the value 1 forcing to the final solution not to overcome the approximate upper membership function $\varphi^{+}_{\hat{\bar{\mu}}_{i,k}}$ in the values close to its maximum. In this sense, if the value 1 is adopted instead 0.95, it is possible that in cases of high spread of the function $\varphi^{+}_{\hat{\bar{\mu}}_{i,k}}$ the solution obtained takes higher values than $\varphi^{+}_{\hat{\bar{\mu}}_{i,k}}$ around of its maximum and, as this function represents degrees of truth, it is not desirable. After selecting the points $P_1$ to $P_3$, compute the values of the parameters $\omega_3$ and $\varsigma_3$ corresponding to the left part of $\varphi^{+}_{\bar{\mu}_{i,k}}$ using the procedure previously explained.
- For the right part of the upper membership function of $\bar{\mu}_{i,k}$, noted by $\varphi^{+}_{\bar{\mu}_{i,k}}$, define the pairs of points $P_1 = (x_1, y_1)$, $P_2 = (x_2, y_2)$, and $P_3 = (x_3, y_3)$ as the points where the right part of the approximate upper membership function $\varphi^{+}_{\hat{\bar{\mu}}_{i,k}}$ take the values closest to 0.05, 0.95, and 0.5, following the same considerations given before for the left part of $\varphi^{+}_{\bar{\mu}_{i,k}}$. Using the points $P_1$ to $P_3$, compute the values of the parameters $\omega_4$ and $\varsigma_4$ corresponding to the right part of $\varphi^{+}_{\bar{\mu}_{i,k}}$ using the procedure previously given.

As a result of this procedure, the parameters $\omega_1$ to $\omega_4$ and $\varsigma_1$ to $\varsigma_4$ of the different Gaussian sub functions which form $\bar{\mu}_{i,k}$ are defined, obtaining an interval type-2 membership function $\bar{\mu}_{i,k}$ for each cluster $k = 1, \ldots, K$ and each feature $i = 1, \ldots, d$. In Fig. 5, a diagram of the Step 1 to 3 of the present stage is shown.

*Step 4:* One fuzzy predicate is defined for each cluster $k \in \{1, \ldots, K\}$ ($K$ compound predicates) by logically operating with the degrees of truth of the interval type-2 membership functions $\{\bar{\mu}_{i,k}\}_{\substack{i=1,\ldots,d \\ k=1,\ldots,K}}$ generated in the previous step. Compound fuzzy predicates are defined considering a conjunction between predicates. For each cluster $k \in \{1, \ldots, K\}$, the next compound predicate is generated in order to explain the cluster $k$:

$$p_k(\mathbf{x}) \equiv \bar{\mu}_{1,k}(x_1) \wedge \bar{\mu}_{2,k}(x_2) \wedge, \ldots, \wedge \bar{\mu}_{d,k}(x_d); \; k = 1, 2, \ldots, K. \tag{16}$$

The predicate $p_k(\mathbf{x})$ is linguistically read as *"The datum* $\mathbf{x}$ *belongs to the cluster k"* and $\bar{\mu}_{i,k}(x_i)$ can be linguistically interpreted as *"The value of the feature i in the datum* $\mathbf{x}$ *is close to the prototypes of the cluster k"*. As closer the value of feature $i$ of the datum $\mathbf{x}$ to the value of the maximum of $\bar{\mu}_{i,k}$ as higher the degree of truth of $\bar{\mu}_{i,k}(x_i)$. As $\bar{\mu}_{i,k}(x_i)$ is higher, $p_k(\mathbf{x})$ should also be higher, reflecting the fact that if the datum $\mathbf{x}$ is close to the centroid of the cluster $k$, then its belonging to the cluster $k$ should increase.

The degrees of truth of all the predicates $\{p_k(\mathbf{x})\}_{k=1,\ldots,K}$ can be computed following the procedure given in the Definition 3.2.8, selecting a fuzzy operator. Cluster assignment is performed applying the procedure explained in the Definition 3.2.10, using the concept of measure of interval of truth values [3].

In Fig. 6, an example of the results obtained from the steps of the present stage is shown, supposing a dataset with two features in which two clusters were discovered, i.e. $K = 2$, and three sets of cluster prototypes were extracted for each cluster at stage C, i.e. $L_k = 3$ both for $k = 1$ and for $k = 2$.

## 4. Experiments

In this Section, experiments done in order to assess the method T2-SFPC and the corresponding results are presented and described in detail. At the end, an illustrative example of the interpretation and knowledge extraction from the membership functions and the fuzzy predicates generated with the T2-SFPC is given, considering the segmentation of brain magnetic resonance images.

### *4.1. Assessment of the proposed method*

In order to make clear the methodology followed for testing and its results, both of these are presented in the next subsections.

#### *4.1.1. Methodology*

Clustering assessing is not trivial, mainly regarding the selection of the best indices for analyzing and comparing [43–45]. The performance of validity indices is highly variable, particularly in complex models. The only way to analyze the results of a clustering algorithm is analyzing the context of each case. This choice depends on the objective pursued for the application of the algorithm, which could be data exploration, generalization of a previous labeling, modeling, among others.

When the objective is a model for a labeled dataset, the number of clusters is known and, consequently, it can be considered a method parameter. In such a case, external validation measures [44] can be used in order to assess the clustering results because expected labels are known. On the other hand, if data exploration is pursued, it requires automatically finding compact clusters, having distant centroids (or other criteria). In this case, parameter optimization is usually applied in order to estimate the number of clusters, as it is the case of the BIC used in the T2-SFPC. Since no information about the expected results is available, internal validation measures must be used for assessment purposes [44].

Based on these previous observations, in the present work, the Silhouette index [44] (an internal measure) and the Accuracy index (an external measure) were selected in order to evaluate the performance of the method T2-SFPC and of other methods used for testing.

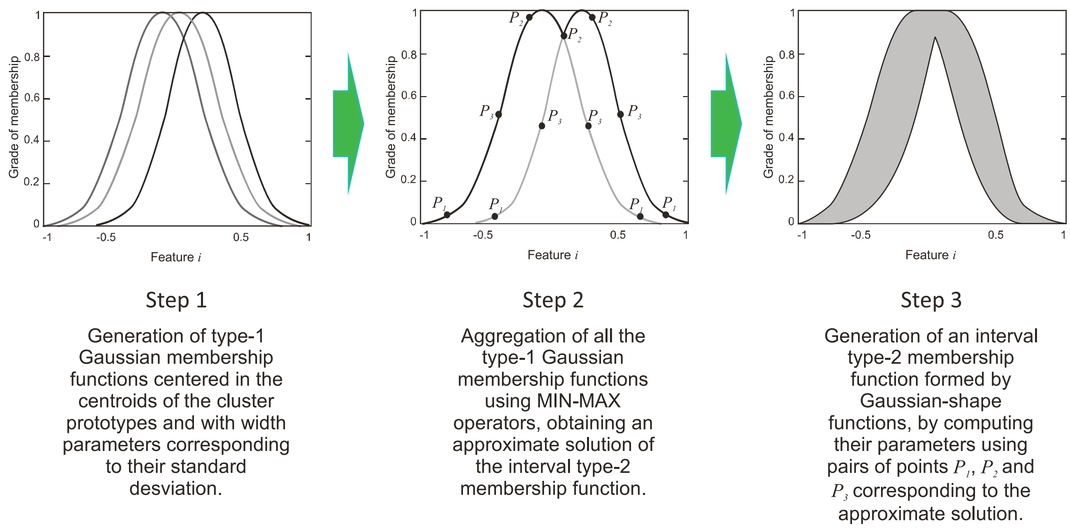


**Fig. 5.** Diagram of the steps 1 to 3 of the stage 4 of the method T2-SFPC. It is shown a case where three sets of cluster prototypes were generated in the previous stage, obtaining three type-1 membership functions. The sets of points $P_1$ to $P_3$ selected to compute the parameters of each of the four parts of the interval type-2 membership function are indicated in the graph above of the step 2 explanation. Four sets of points $P_1$ to $P_3$ are selected corresponding respectively to the left part of the upper membership function, the left part of the lower membership function, the right part of the lower membership function, and the right part of the upper membership function.

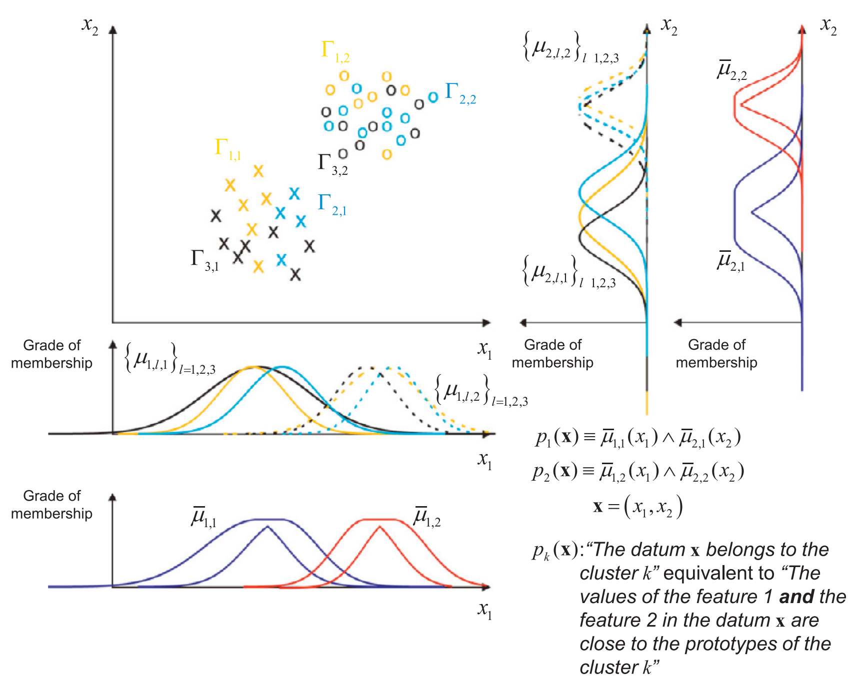


**Fig. 6.** Example of the results of the stage D of the method T2-SFPC, considering a dataset where two clusters were discovered at stage C and three sets of cluster prototypes were extracted. The method outcome consists of the generation of the interval type-2 membership functions $\{\bar{\mu}_{i,k}\}_{i=1,\dots,d;\ k=1,\dots,K}$ ($d$ is the number of features and $K$ the number of discovered clusters) and the fuzzy predicates $\{p_k(\mathbf{x})\}_{k=1,\dots,K}$, which allow both the clustering of the data and the knowledge extraction about the clustering problem.

The Silhouette index is useful for evaluating how compact and how separated are the clusters obtained for a given dataset. Typically, it is used an average of the Silhouette indices over the clusters, which has values between −1 and 1. The higher the Silhouette, the more compact and separated the clusters are [44].

The Accuracy measure allows to know what the proportion of corrected labelled data is, provided that the expected labels (gold standard) for the data are known. Gold standards are ideally produced by human experts or based on previous experimentation, often labeling data at the same moment they are obtained [46]. In the present work, as cluster assignations vary from a clustering to another, previous to the accuracy computing, each defined cluster is assigned to one (and only one) class, considering the majority labels in each cluster. Then, the measure is simply estimated as the ratio between the quantity of data assigned correctly to the quantity of available data [2,3].

In order to be independent of the algorithm initialization, each clustering algorithm was run 20 times and values reported are averages of the Accuracy and Silhouette indices. Standard deviations and statistical tests of significance were also computed and reported.

The datasets selected are:

- Banknote dataset (5 features - 2 selected, 2 clusters, 1372 data) [47].

- Wisconsin breast cancer dataset (32 features - 3 selected, 2 clusters, 569 data) [48].
- Pima Indians diabetes dataset (8 features - 3 selected, 2 clusters, 768 data) [48].
- Iris dataset (3 clusters, 4 features, 150 data) [49].
- Wine dataset (3 clusters, 13 features, 198 data) [50].
- Moon dataset (2 features, 2 clusters, 2000 data), a synthetic dataset presented in [2].
- MRI1 dataset, 4000 pixels randomly selected per cluster from simulated magnetic resonance images (3 clusters, 3 features, 12,000 data) [51]. These data were taken without any noise or distortion and were generated by means of computer simulation.
- MRI2 dataset, 200 pixels randomly selected per cluster from the previous dataset (3 clusters, 3 features, 600 data) [51].
- Seeds dataset (7 features, 3 clusters, 210 data) [52].

In the case of the Wisconsin breast cancer and Pima Indians diabetes datasets, in order to facilitate result comparisons, only the features suggested in [2] were used.

Regarding to the Banknote dataset, the two first features were selected following the considerations given in [3].

The methods used in the tests, including the proposed method, are detailed below and an acronym is defined for each of them:

- **T2-SFPC:** Type-2 SOM-based Fuzzy Predicate Clustering (the method proposed). SOMs' sizes are automatically determined according to the number of training data (as described in Section 3.3).
- **T2-DFPC:** Type-2 Data-based Fuzzy Predicate Clustering. It is the method presented in [3]. In this case, SOMs are not used and cluster prototypes are extracted directly from clustering defined on $M$ disjoint subsets of the original dataset. As it was previously mentioned, FCM and BIC are used is this method as used in the T2-SFPC, but unlike it, values of the interval type-2 membership function are numerically obtained aggregating type-1 Gaussian-shape membership functions.
- **SFPC:** SOM-based Fuzzy Predicate Clustering. It is based on the method presented in [2], but an improvement was added to the original proposal during the test: it was included the FCM-BIC clustering approach in order to automatically defines the proper number of cluster in each case.
- **SOM-FCM:** a SOM was trained with the dataset and an automatic FCM-BIC clustering scheme was applied to the resulting codebook (configuring a basic two-level clustering scheme) [3].
- **K-means** [1]**:** the K-means algorithm combined with the BIC was applied to the dataset, determining the proper number of clusters in each case. Cluster centroids were obtained and data were assigned to the cluster corresponding to the closest centroid, considering Euclidian distance.
- **FCM** [16]**:** FCM algorithm was combined with the BIC and was applied to the dataset considering random initial centroids. Once centroids were obtained, each datum was assigned to the cluster which it belongs with the highest membership value.

Expectation-Maximization [53] was also tried, but it never outperformed other methods along the different datasets tested. Therefore, these results are not included in the present work.

Variants of all the methods were considered, removing the automatic BIC-based clustering schemes and using traditional cluster techniques, requiring knowing the expected number of clusters for each tested dataset. These clustering algorithms without automatic clustering were respectively called: T2-SFPC-*wac*, SFPC-*wac*, T2-DFPC-*wac*, SOM-FCM-*wac*, K-means-*wac*, and FCM-*wac*. By considering the original algorithms and these variants without automatic clustering, the performance of the automatic clustering scheme based on the BIC was analyzed. It should be noted that the variant called SFPC-*wac* is in fact the original method called SFPC proposed in [2].

In order to test changes of the performance of the FL-based methods when the size of the random partition ($M$) changes, i.e. to analyze the sensibility of the methods in relation with the numbers of subsets considered, all the methods were run with $M$ varying from 2 to 10. In addition, for the type-1 FL methods SFPC and SFPC-*wac*, the value $M = 1$ was also tested. Different results and performance comparisons are given in the next sub-sections.

In the case of the SFPC and SFPC-*wac* method, when $M > 1$ was considered, fuzzy predicates were evaluated using the algorithm option suggested in [2] for these methods, referred as Option 3 (Op. 3) in the referred work.

In the next sections, results of the tests performed following the previous considerations are presented and described. First, accuracies results are given. Then, Silhouette indices are detailed.

*4.1.2. Results: Accuracy*

Clustering accuracies obtained for the tested datasets are shown in Fig. 7. Accuracy is represented by boxplots. A graph is shown for each dataset. In the case of the methods based on fuzzy predicates, three grouped whiskers are shown for each method, since results using different FL operators were evaluated and compared, considering the standard triangular norms (MIN-MAX operators) and the compensatory FL operators based on GMCFL and AMCFL. Results reported correspond to the highest average value of Accuracy for each method, considering the results obtained for the tested values of the parameter M, which is indicated in each case. Also, the result for SFPC and SFPC-*wac* for $M = 1$ is included. A horizontal line is shown, representing the best of the medians of the results achieved by the method proposed (T2-SFPC) or its variant without automatic clustering scheme (T2-SFPC-*wac*), the best of both.

A detailed analysis of the clustering accuracies results is done in the next paragraphs for the different datasets. When two different accuracy results $acc_1$ and $acc_2$ are compared, the percentage difference between the accuracies given as $difference\,\% = \frac{acc_1 - acc_2}{acc_2} \times 100\%$ is used.

Additionally, in order to compare the results of the proposed method (T2-SFPC) against the test methods based on fuzzy predicates previously proposed in [2] and [3] (SFPC and T2-DFPC), Table 1 is presented. In this table, mean value and standard deviation of the accuracy obtained for the methods T2-SFPC, T2-DFPC, and SFPC are shown, including percentage differences between T2-DFPC and SFPC against the proposed method. Values of statistical significance tests for the comparisons are reported. In all cases, it was selected and informed the results of the algorithms considering the best result in mean value for the methods with and without FCM-BIC clustering approach. The values of $M$ corresponding to the best accuracies are indicated. In the right column of the table, results of SFPC with $M = 1$ is reported. The result of the best of all these methods in each dataset is indicated in bold font.

As it can be noted, the Banknote dataset (Fig. 7a) could be successfully clustered by almost all the methods, considering accuracy results. However, the proposed approach was the best one, improving the other methods by 0.5% or more, with statistical significance, except in the case of the SFPC ($M = 5$) where the difference observed was not statistically significant. Additionally, there was not statistically significant difference between the performance of the proposed method and its variant without automatic clustering scheme (T2-SFPC-*wac*).

Classical clustering methods had good results in the Wisconsin breast cancer dataset (Fig. 7b). The best of them was the K-means-*wac*, which overcame by 0.4% the proposed method ($p < 0.001$). Nevertheless, even changing the logic operators applied, accuracy results of the T2-SFPC were always higher than 0.94%, showing very low variance (standard deviation $= 0.002$) along the different

**Table 1**
Accuracy results of the methods based on fuzzy predicates: T2-SFPC (proposed method), T2-DFPC and SFPC (with $M=1$ and $M>1$). The best result between variants of the methods with and without automatic determination of the number of clusters is reported. When variant without automatic clustering is the best, this is indicated in brackets as *wac*. The value of the partition size corresponding to the highest accuracy ($M$) is indicated. Comparisons against the proposed method are included. The best result for each dataset is indicated in bold font.

| Dataset | T2-SFPC | T2-DFPC | | SFPC | | SFPC ($M=1$) | |
|---|---|---|---|---|---|---|---|
| | | Obtained result | Difference with T2-SFPC | Obtained result | Difference with T2-SFPC | Obtained result | Difference with T2-SFPC |
| Banknote | **$0.871 \pm 0.007$ ($M=5$)** | $0.866 \pm 0.004$ ($M=5$) | 0.627% ($p<0.001$) | $0.870 \pm 0.009$ ($M=5$) | 0.197% ($p=0.714$) | $0.857 \pm 0.000$ | 1.658% ($p<0.001$) |
| Wisconsin breast cancer | $0.955 \pm 0.002$ (*wac*) ($M=2$) | $0.952 \pm 0.001$ (*wac*) ($M=2$) | 0.304% ($p<0.001$) | $0.944 \pm 0.019$ ($M=2$) | 1.248% ($p=0.006$) | **$0.958 \pm 0.001$** | **−0.257% ($p<0.001$)** |
| Pima Indians diabetes | **$0.714 \pm 0.013$ ($M=4$)** | $0.708 \pm 0.020$ ($M=4$) | 0.708% ($p=0.489$) | $0.712 \pm 0.015$ ($M=3$) | 0.247% ($p=0.724$) | $0.702 \pm 0.005$ | 1.678% ($p<0.001$) |
| Iris | $0.922 \pm 0.022$ ($M=3$) | $0.929 \pm 0.009$ (*wac*) ($M=2$) | −0.789% ($p=0.179$) | $0.921 \pm 0.028$ ($M=2$) | 0.145% ($p=0.875$) | **$0.933 \pm 0.029$** | **−1.214% ($p=0.269$)** |
| Wine | **$0.964 \pm 0.007$ (*wac*) ($M=2$)** | $0.953 \pm 0.011$ (*wac*) ($M=2$) | 1.149% ($p<0.001$) | $0.925 \pm 0.046$ (*wac*) ($M=3$) | 4.189% ($p<0.001$) | $0.910 \pm 0.000$ | 5.926% ($p<0.001$) |
| Moon | $0.951 \pm 0.012$ ($M=3$) | $0.958 \pm 0.011$ ($M=2$) | −0.697% ($p=0.020$) | **$0.959 \pm 0.012$ ($M=3$)** | **−0.806% ($p=0.037$)** | $0.946 \pm 0.011$ | 0.531% ($p=0.062$) |
| MRI1 | $0.971 \pm 0.001$ (*wac*) ($M=3$) | $0.960 \pm 0.001$ ($M=3$) | 1.183% ($p<0.001$) | $0.963 \pm 0.008$ (*wac*) ($M=3$) | 0.853% ($p=0.006$) | **$0.972 \pm 0.000$** | **−0.113% ($p=0.004$)** |
| MRI2 | $0.979 \pm 0.002$ (*wac*) ($M=3$) | **$0.981 \pm 0.002$ (*wac*) ($M=2$)** | **−0.248% ($p=0.031$)** | $0.969 \pm 0.007$ (*wac*) ($M=3$) | 1.070% ($p=0.006$) | $0.954 \pm 0.000$ (*wac*) | 0.954% ($p<0.001$) |
| Seeds | $0.949 \pm 0.006$ ($M=2$) | $0.949 \pm 0.005$ ($M=2$) | −0.025% ($p=0.957$) | $0.932 \pm 0.014$ ($M=2$) | 1.866% ($p<0.001$) | **$0.952 \pm 0.000$** | **−0.350% ($p=0.027$)** |

executions. Considering the methods based on fuzzy predicates, only the SFPC lightly improved the result of the T2-SFPC by 0.257% ($p<0.001$ and $M=1$).

For the Pima Indians diabetes dataset (Fig. 7c), the best result corresponded to the proposed method with 0.714 of accuracy (standard deviation $=0.013$). The second-best method with statistically significant difference was the SFPC-*wac* with $M=1$, which obtained an Accuracy 1.678% lower than the T2-SFPC ($p<0.001$). Differences both with T2-DFPC and SFPC ($M=3$) did not report statistical significance.

For the classic Iris dataset (Fig. 7d), the proposed method was one of the best methods, considering that differences reported against the other methods based on fuzzy predicates were in all cases not statistically significant. The method T2-SFPC showed Accuracy of 0.922% with low variance (standard deviation $=0.022$).

Observing the results for the Wine dataset (Fig. 7e), results suggest that the proposed method without automatic clustering (T2-SFPC-*wac*) was the best, overcoming the rest of the methods by more than 1.149% with statistical significance. As it is noted, for this dataset the best results corresponded to the methods based on fuzzy predicates.

In the case of the Moon dataset (Fig. 7f), obtained results show that the methods based on fuzzy predicates could improve the classic methods by differences higher than 6% with statistical significance. The best result for the T2-SFPC was 0.951 (standard deviation $=0.012$), overcome by the T2-DFPC by 0.697% ($p=0.020$) and by the SFPC by 0.806% ($p=0.037$, $M=3$), i.e. both differences were statistical significant. In addition, SFPC with $M=1$ was overcome by the proposed method by 0.531% ($p=0.062$).

Analyzing the results for the MRI1 dataset (Fig. 7g), it is observed that the best methods were the classical FCM in its two variants, i.e. FCM and FCM-*wac*, improving the performance of the method T2-SFPC by 0.793% with statistical significance, which obtained an accuracy of 0.971 (standard deviation $=0.001$). The proposed method was the second best of the methods based on fuzzy predicates, only lightly overcome by the SFPC ($M=1$) by 0.113% ($p=0.004$).

Considering the MRI2 dataset (Fig. 7h), the proposed method was the second-best method, presenting an accuracy of 0.979 (standard deviation $=0.002$), only overcome by the T2-DFPC-*wac* by 0.248% with statistical significance.

Finally, in the case of the Seeds dataset (Fig. 7i), it is observed that the method proposed was lightly overcome by the SFPC ($M=1$) by 0.350% with statistical significance. The difference observed against the method T2-DFPC was not statistically significant. The rest of the tested methods were overcome by the proposed method in all cases with statistical significance.

It should be noted, observing the graphs shown in Fig. 7, that for the methods based on fuzzy predicates the fuzzy operators with the highest accuracy were the same for a given dataset between the different methods, i.e. when MIN-MAX were the best fuzzy operators in a dataset, they were the best in the rest of the methods based on fuzzy predicates for that dataset. The same occurs in the cases of the operators based on GMCFL and AMCFL. In this sense, this observation indicates that the best fuzzy operator strongly depends on each clustering problem.

Considering Table 1, the proposed method was overcome with statistical significance by the SFPC with $M=1$ in the Wisconsin breast cancer, MRI1, and Seed dataset, but these differences were lower than 0.35%. In addition, the T2-DFPC overcame the proposed method in the cases of the Moon and MRI2 dataset by 0.697% and 0.248% respectively, and the SFPC with $M>1$ only overcame the proposed method in the case of the Moon dataset by 0.806%. No significant differences were observed between the proposed method and the SFPC ($M>1$) for the datasets Banknote, Pima Indians diabetes, and Iris. In relation with the method T2-DFPC, for the datasets Pima Indians diabetes and Iris the resulting differences were not significant. In the rest of the cases, comparisons indicated a higher Accuracy of the method proposed than those of the methods T2-DFPC and SFPC.

Summarizing, considering accuracy values, even when the proposed method was outperformed in some cases, it was always one of the best, considering the experiments done testing very different datasets. When the proposed method was neither in the first

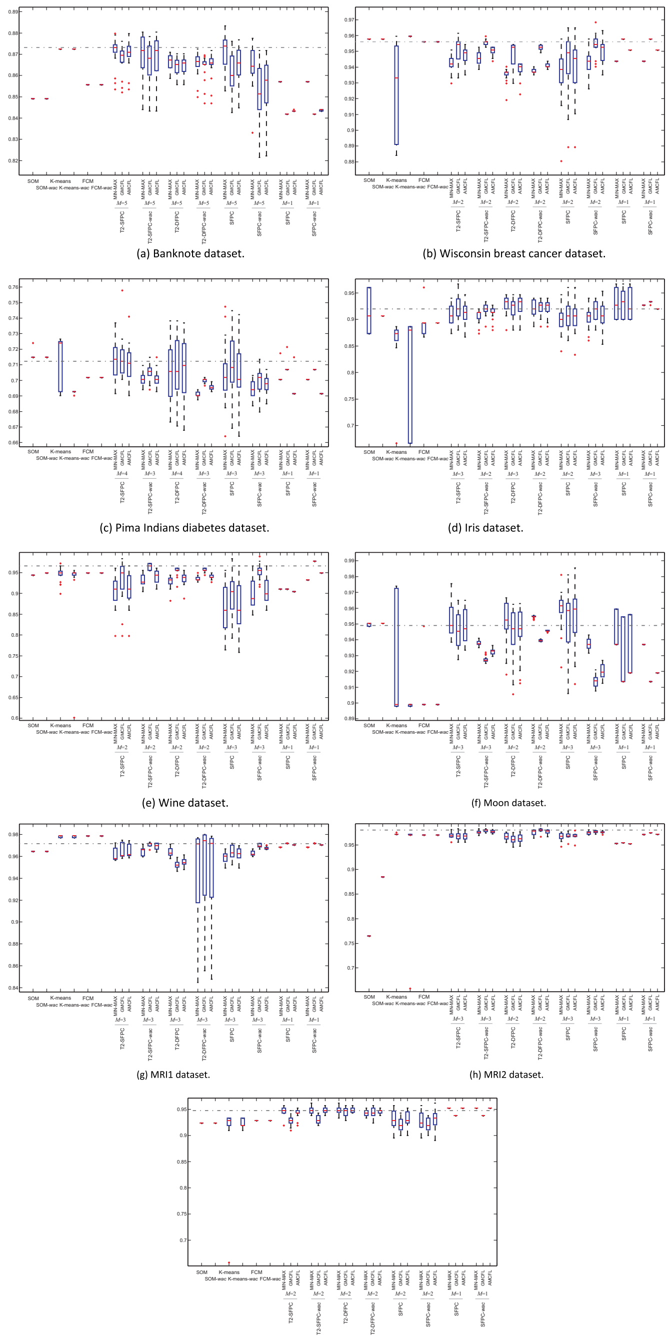


**Fig. 7.** Clustering accuracies for the test datasets. Boxplots in groups of three indicate respectively the logic operators MIN-MAX, GMCFL, and AMCFL. The partition size ($M$) for the highest Accuracy is shown in the corresponding cases. The horizontal dotted line indicates the best accuracy value obtained for T2-SFPC or T2-SFPC-*wac* (the median of the accuracies over all the tests run on each dataset). (a) Banknote. (b) Wisconsin breast cancer. (c) Pima Indians diabetes. (d) Iris. (e) Wine. (f) Moon. (g) MRI1. (h) MRI2. (i) Seeds.

nor in the second place, the accuracy was only a little lower. In all cases, the differences never exceed 0.806%. This evidence makes the approach reliable. Even in the cases where the T2-SFPC was not the best; results showed that the approaches based on fuzzy predicates are always in the top of the list with small differences in the performance. It is remarkable that except for the MRI1 dataset, methods based on fuzzy predicates showed to be the best or the second-best method.

Concerning to the sensibility of the methods T2-SFPC, T2-DFPC, SFPC, T2-SFPC-*wac*, T2-DFPC-*wac*, and SFPC-*wac* in relation with the numbers of subsets considered, i.e. the value adopted by $M$, in most of the cases the results showed no statistically significant differences compared to the best accuracies values, considering values of $M$ close to those of the optimal results. However, for values of $M$ higher than 6, the smaller datasets Wisconsin breast cancer, Pima Indians diabetes, Iris, Wine, MRI2, and Seeds reported a decrease of the Accuracy compared to the optimal case, with statistical significance. Additionally, for most of the datasets the best value of $M$ was the same for the different methods and when it was not equal, it differed by 1. Therefore, according to the accuracy values, the methods are not highly sensitive to the value of $M$ and it can be adjusted taking in to account the size of the datasets and, when it is known, the number of expected clusters.

### 4.1.3. Results: Silhouette index

Regarding to the Silhouette index, the obtained results are summarized in Table 2. In this table, for each dataset, the best result between T2-SFPC and T2-SFPC-*wac* is indicated in the first column. In the second column, the best result for the test methods are reported. In all cases, the name of the best method and its parameters are informed. Differences between Silhouette indices are reported when these resulted statistically significant. It is remarkable that in most of the cases, the proposed method T2-SFPC or its variant without automatic clustering (T2-SFPC-*wac*) obtained the highest Silhouette index. Differences observed exceeded 6.7% when they were statistically significant, indicating that the clusters generated by the method proposed are more compact and separated than those obtained by means of the test methods. In the cases of Pima Indians diabetes, MRI1, and Seeds datasets, differences resulted not statistically significant.

Observing the value of $M$ for the highest value of Silhouette index of the T2-SFPC along the different datasets, the best parameter value for Accuracy and Silhouette indices are the same or are close, except for Banknote and Wisconsin breast cancer. Additionality, in these two datasets Silhouette index reported to be more sensitive to the value of $M$ than in the other datasets in which the variation of the Silhouette index when the value of $M$ moved from the optimal value was similar to the observed for the Accuracy index. In this sense, the proposed method resulted lightly sensitive to the value of $M$ considering the Silhouette index.

## 4.2. Interpretable clustering from the method T2-SFPC: an illustrative example using brain magnetic resonance images

In this section, an illustrative example of the use of the method T2-SFPC proposed is studied, considering the segmentation of brain magnetic resonance images. As it was mentioned in the example given at the end of Section 3.2, the problem consists of providing a label for each pixel of the images detecting gray matter, white matter, and cerebrospinal fluid. Considering this kind of problem, the method T2-SFPC not only is able to assign tissues to pixels, but also it allows to extract linguistic descriptions which agree with those used by experts when they describe how the different tissues are detected. Therefore, when expert knowledge is not available for a problem, it is possible to discover knowledge about the clustering through the method T2-SFPC and this knowledge can be expressed using linguistic descriptions.

In Fig. 8, in order to provide visual examples of the images studied in the present section and of the results obtained with the method T2-SFPC, three of the slices contained in the database of simulated brain magnetic resonance images [51] are shown, including images corresponding to the sequences PD, T1, and T2, the gold-standard segmentation and the segmentation generated

**Table 2**
Comparisons for Silhouette indices between the proposed method and the best of the test methods. Differences are informed when resulted statis'tically significant. Partition sizes ($M$) are included in the corresponding cases. In all cases, the best result of the methods with and without automatic determination of the number of clusters is informed.

| Dataset | Best Silhouette for the proposed method | Best Silhouette for the test methods | Difference (when $p < 0.05$) |
|---|---|---|---|
| Banknote | T2-SFPC-*wac* ($M=2$) ($-0.102 \pm 0.038$) | T2-DFPC-*wac* ($M=2$) ($-0.118 \pm 0.035$) | 0.016 (13.5%) |
| Wisconsin breast cancer | T2-SFPC ($M=4$) ($-0.004 \pm 0.102$) | K-means-*wac* ($-0.057 \pm 0.048$) | 0.053 (92.9%) |
| Pima Indians diabetes | T2-SFPC ($M=4$) ($0.026 \pm 0.063$) | SFPC ($M=3$) ($0.021 \pm 0.086$) | – |
| Iris | T2-SFPC-*wac* ($M=4$) ($0.167 \pm 0.056$) | FCM ($0.091 \pm 0.071$) | 0.076 (83.5%) |
| Wine | T2-SFPC-*wac* ($M=2$) ($0.095 \pm 0.030$) | FCM-wac ($0.089 \pm 0.027$) | 0.006 (6.7%) |
| Moon | T2-SFPC ($M=4$) ($0.110 \pm 0.050$) | K-means ($-0.013 \pm 0.089$) | 0.123 (> 100%) |
| MRI1 | T2-SFPC ($M=3$) ($0.231 \pm 0.017$) | SOM-*wac* ($0.246 \pm 0.037$) | – |
| MRI2 | T2-SFPC-*wac* ($M=2$) ($0.236 \pm 0.025$) | K-means ($0.209 \pm 0.037$) | 0.027 (12.9%) |
| Seeds | T2-SFPC ($M=2$) ($0.122 \pm 0.023$) | K-means-*wac* ($0.126 \pm 0.022$) | – |

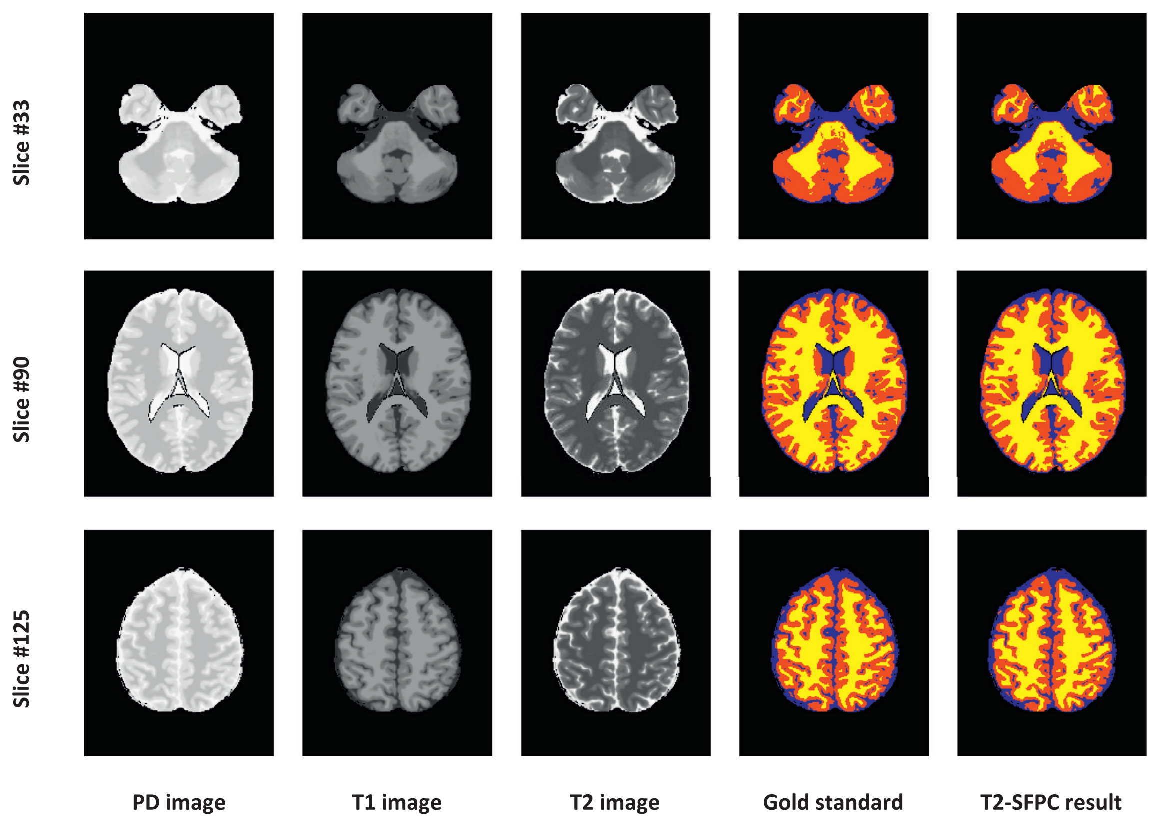


**Fig. 8.** Images corresponding to three representative slices in the database of simulated brain magnetic resonance images [51]. For each slice, images are presented in the next order: images corresponding to the sequences PD, T1, and T2, expected segmentation (gold standard) and resulting segmentation using the method T2-SFPC proposed. In the segmented images, colors indicate different brain tissues: gray matter (red), white matter (yellow), and cerebrospinal fluid (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

from the method T2-SFPC. In order to obtain the image segmentation, first of all the method T2-SFPC was applied on the dataset MRI2. Then, the interval type-2 fuzzy predicate system generated was used for labeling each of the pixels corresponding to the selected slices. The tissues were assigned to the clusters considering which tissue was the most frequent in each of the discovered clusters when the MRI2 was applied on the fuzzy predicate system. Each pixel of the slices was represented by a datum to be clustered which was formed by the gray intensities of the pixel in the images corresponding respectively to the sequences PD, T1, and T2.

In Fig. 9, the interval type-2 membership functions generated automatically by means of the method T2-SFPC are shown. In this case, the random partition size was set to $M=5$. Gray intensity values are normalized, corresponding the value $-1$ to black (the lowest gray level intensity) and $+1$ to white (the highest gray level intensity).

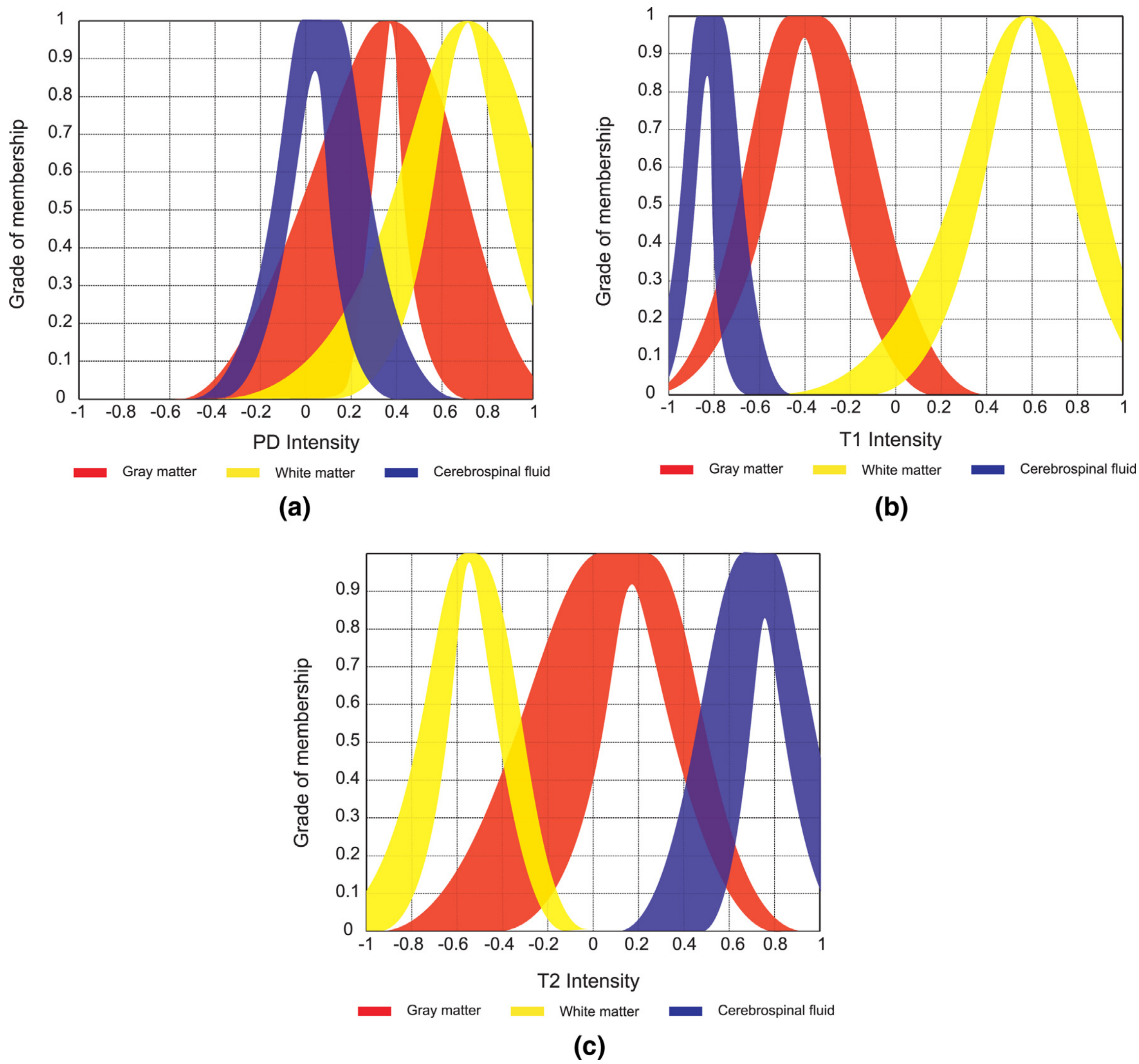


**Fig. 9.** Interval type-2 membership functions obtained for the MRI2 dataset using the method T2-SFPC and a random partition size $M = 5$. Different colors represent different brain tissues: gray matter (red), white matter (yellow), and cerebrospinal fluid (blue). (a) Membership functions for the feature "PD Intensity". (b) Membership functions for the feature "T1 Intensity". (c) Membership functions for the feature "T2 Intensity". (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

By analyzing the membership functions in Fig. 9, it is possible to identify relationships among values of the features and the tissues. Let consider, for instance, the cerebrospinal fluid whose membership functions are shown in blue in Fig. 9. For this label, the membership function for the feature "PD Intensity" (Fig. 9a) is centered in medium values of the normalized feature values, approximately located between −0.2 and 0.4 in the normalized gray scale. This group of values can be identified with the attribute *"Gray"* considering they are in the middle region of the gray scale. For the same tissue, the values of "T1 Intensity" (Fig. 9b) are situated around low values of gray intensity, specifically between −1 and −0.6. Therefore, the cerebrospinal fluid is related to *"Dark"* values of "T1 Intensity". Finally, the membership function corresponding to cerebrospinal fluid for the "T2 Intensity" (Fig. 9c) is located at the right of the scale, between 0.4 and 1, which means the cerebrospinal fluid is associated to the attribute *"Bright"* for the feature "T2 Intensity".

The analysis of the previous paragraph can be written as a fuzzy predicate describing the relationships observed, i.e. the attributes required for a pixel belonging to cerebrospinal fluid: $p_1(\mathbf{x})$: *"The pixel* **x** *belongs to cerebrospinal fluid"* which is equivalent to the compound predicate *"In the pixel* **x** *the PD intensity is Gray, the T1 intensity is Dark and the T2 intensity is Bright"*. The degrees of truth of the simple predicates *"In the pixel* **x** *the PD intensity is Gray"*, *"In the pixel* **x** *the T1 intensity is Dark"*, and *"In the pixel* **x** *the T2 intensity is Bright"* are defined by the interval type-2 membership functions obtained for the cerebrospinal fluid in the respective features, i.e. the membership functions shown in blue in Fig. 8. It should be noted that $p_1(\mathbf{x})$ allows to evaluate the degree of truth in which a pixel **x** belong to cerebrospinal fluid and it is computed by means of the simple predicates extracted from the membership functions in Fig. 9.

Attributes and predicates for the tissues gray matter and white matter can be defined extending the analysis given before. As a result, the next fuzzy predicates are defined:

- For gray matter: $p_2(\mathbf{x})$:*"The pixel* **x** *belongs to gray matter"* which is equivalent to *"In the pixel* **x** *the PD intensity is Light-Gray, the T1 intensity is Dark-Gray and the T2 intensity is Medium-Gray"*.
- For white matter: $p_3(\mathbf{x})$: *"The pixel* **x** *belongs to white matter"* equivalent to *"In the pixel* **x** *the PD intensity is Very-Bright, the T1 intensity is Very-Bright and the T2 intensity is Dark"*.

In addition, a visual analysis can be conducted by visual inspection of the images of Fig. 8, focusing on pixels labelled and comparing these pixels with the same space locations in the images PD, T1, and T2. As a result, it is possible to confirm the discovered relationships between values of the features and the tissues. For instance, the cerebrospinal fluid (shown in blue in the segmented images) appears as bright in PD, dark in T1 and bright in T2 provided that a person give some visual analysis on the images. The same procedure can be applied to the gray matter and the white matter. Therefore, the membership functions generated by the method T2-SFPC also assist to the visual interpretation of the segmentation.

In relation to number of SOMs trained in the method T2-SFPC, i.e. the value of the $M$ parameter, it is important to note that the method always generates only one interval type-2 membership function for each feature and each discovered cluster, merging all the knowledge about each cluster discovered from the $M$ SOMs. Consequently, the membership function analysis is made in a simpler way than in the case of the method SFPC whether more than one SOM is used in the SFPC ($M > 1$) which, according to the results described in Section 4.1, is often required in order to obtain good clustering results.

As it was mentioned, the method T2-SFPC preserves all the knowledge discovery capabilities of the previous method called T2-DFPC, which were extensively discussed in the stage #4 of this method in [3]. In this regard, it is possible to extract valuable information related to vagueness or variability around of the clusters by analyzing both the area of the resulting FOUs and the width of the membership functions. Considering the method T2-SFPC, a large FOU means a large variation in the prototypes contained in the SOMs, which were trained using the subsets obtained during the random partition in the stage A. In addition, from the point of view of a cluster associated to an interval membership function, a large FOU is related to large vagueness (variability or disagreement) about the degree of truth of the correspondence between values of a feature and the cluster or about the attribute described by the interval membership function.

Applying this analysis to the membership functions obtained for the MRI2 dataset (shown in Fig. 8), the next observations can be given:

- For the cerebrospinal fluid (shown in blue in Fig. 9), the values of “PD Intensity” and “T2 Intensity” are associated to membership functions with relative large FOU, meaning large vagueness for this tissue in these features. In the “T1 Intensity”, the membership function indicates smaller vagueness than for both previous features.
- For the gray matter (shown in red in Fig. 9), the obtained FOUs indicate large vagueness, being larger than those existing about the rest of the tissues. The big width of the membership function for the “T2 Intensity” indicates a high variability around the attribute *“Medium-Gray”* associated to it.
- For the white matter (shown in yellow in Fig. 9), membership functions have small FOUs for all the features, meaning there exist a small vagueness around this cluster and small variability between the cluster prototypes obtained from the $M$ SOMs.
- For the feature “PD Intensity”, membership functions are strongly overlapped which means there exist high fuzziness of the associated attributes.

## 5. Discussion

As it was mentioned, a major contribution of this approach is the interpretability of the clusters. Once the clusters have been found, the membership functions can be analyzed to obtain useful linguistic interpretation of the groups. Specific terminology coming from the field of the data can be used, giving the possibility of getting some new knowledge, as it was analyzed in Section 4.2.

Analyzing the numerical results obtained during the method assessment, performance achieved by the method T2-SFPC proposed in the present work is in the top of the methods tested, being in most of the cases the method with the best performance considering both Accuracy and Silhouette indices. It is important to note that even when Accuracy is high internal measures like Silhouette could be low, being this situation a consequence of the clusters themselves. Some labeled datasets present groups that do not fulfill the requirements of compactness or distant centroids. However, analyzing the results for the Silhouette index the proposed method is outstandingly better than the test methods.

Specifically, considering Accuracy indices, previous methods T2-DFPC and SFPC outperformed the T2-SFPC in some cases, but these differences keep lower than 0.806%. The method based on fuzzy predicates overcame the traditional clustering techniques having the best and the second-best accuracy, except in two of the tested datasets. On the other hand, the obtained Silhouette indices indicate that the method proposed had the best performance in all cases not being overcome by the test methods for any dataset. Therefore, for the tested datasets the proposed method defined more compact and distant clusters than all the test methods. Consequently, considering the results of both validation indices, even when the method proposed was outperformed in accuracy in some cases, its performance is highly acceptable, which evidences that it constitutes a reliable general clustering method.

In relation with the stages composed the method T2-SFPC proposed and the previous methods SFPC and T2-DFPC, major differences are related to: the kind of FL used (comparing with SFPC) and the way in which cluster prototypes are obtained from data (comparing with T2-DFPC).

In this regard, analyzing the stages A and B of the proposed method (explained in Sections 3.3.1 and 3.3.2), SOM-clustering scheme discovers knowledge related to clustering allowing to extract the cluster prototypes. SOMs are automatically configured and trained in the same way it was realized in the previous method SFPC, but unlike it, in the T2-SFPC the suitable number of clusters for a dataset is found using a FCM-BIC scheme and interval type-2 membership functions are defined from the obtained prototypes. As a result, unlike the method SFPC, the proposed one do not require knowing the number of clusters to generate.

In addition, the aggregation process suggested in the present work merges the contribution of each subset of cluster prototypes defining a single interval type-2 membership function for each feature and each cluster. In this sense, the interval type-2 membership function generated acts capturing all the partial descriptions of the clusters modelled by the type-1 membership functions in a first step and merging them in a single model, which is more easily interpretable than those generated by the SFPC with $M > 1$ where several overlapped type-1 membership functions describe the characteristics of a given cluster. Considering that the method SFPC with $M > 1$ only overcame the proposed method in one case, it is possible to conclude that the aggregation process proposed combines successfully the results coming from the different subsets of cluster prototypes as different “opinions” of experts.

Additionally, the computational effort involved in computing interval type-2 fuzzy predicates (introduced in Section 3.2) is similar to that required by the type-1 fuzzy predicates generated by the method SFPC when $M = 1$, and it is smaller than the effort required by the same method with $M > 1$. Due to this and considering the less complexity of the interpretation of the clustering obtained with the method T2-SFPC compared to the SFPC with $M > 1$, the interval type-2 fuzzy predicates are an easy and good performance approach for data clustering.

On the other hand, the method proposed in this work exploits the known advantages of the SOMs for dealing with noise, outliers, and missing values. Comparing with the T2-DFPC, the main differences are: the use of SOMs in order to extract the cluster prototypes and the definition of parametrizable interval type-2 membership functions formed by Gaussian-shape sub-functions, allowing the optimization of their parameters whether a specific goal is defined. The method proposed preserves all the knowledge extraction characteristics of the T2-DFPC, which were analyzed in Section 4.2 considering a real application case on brain magnetic resonance images.

Regarding to the clustering approach based on the BIC, the methods with and without automatic determination of the number of clusters reported similar performance without statistically significant difference in most of the cases. Even when in some of the tested datasets the results using the automatic clustering scheme showed higher variance than variants without it (those methods named with suffix "*-wac*"), in other cases the situation was the opposite. Therefore, according with the observed results, the BIC combined with classical clustering techniques such as FCM and K-means constitutes an appropriate choice for determining the proper number of clusters when no information about the clustering problem is available.

In relation with the selection of the value of the parameter $M$, i.e. the size of the initial random data partition at the stage A, the results of the tests indicate that both the proposed method T2-SFPC and the previous methods T2-DFPC and SFPC are not strongly sensitive to the value of $M$ considering both validation indices analyzed. In this sense, the value of $M$ should be selected according to the number of data in the dataset and the number of clusters expected, when it is available, considering that the proposed method requires training a SOM for each of the subsets defined during the random partition at the stage A. In other words, the more complex the problem addressed is, the lower the proper value of $M$ is, considering a complex clustering problem one in which there are a few data (a few examples) for each of the expected clusters or one in which the clusters that can be detected are very close. In this regard, according to the tests performed by analyzing the stability of the T2-SFPC respect to the value of $M$, to obtain acceptable clustering performance it was required at least 10 data per expected cluster in each of the subsets generated at the stage A. Based on this empirical observation, it is possible to use this limit as a criterion to define the maximum value of $M$ for a given dataset.

On the other hand, if the number of expected clusters is unknown, as it is the case of data exploration based on clustering, the value of $M$ can be adjusted observing the performance of the clustering resulting once the fuzzy system was generated for distinct values of $M$. In such an approach, the value of $M$ can be increased from 2 and for each an internal validation measure can be used to assess the final clustering, obtaining the proper value for $M$. The SOM error measures computed at the stage B of the T2-SFPC in combination with internal validation measures applied on the clustering defined on the SOMs' codebooks at the stage C (the cluster prototypes) can also be analyzed varying $M$ for the determination of a proper value of $M$ if no information about the clustering is available.

In addition to the previous observations related to the value of $M$, the random partition at the stage A of the proposed method can reduce the computational effort required during the method application. In this regard, there are some prominent issues to consider:

- The random data partition reduces the number of training data of the SOMs. Considering an initial dataset **X**, when the size of the random partition $M$ increases, computational effort of the SOM training decreases. As a result, even when the higher value of $M$ the higher number of SOMs to be trained, the computational effort required to train the $M$ SOMs using the subsets obtained from the random data partition is reduced when the value of $M$ is increased. This remarkable feature makes that the computational effort required using the SFPC with $M = 1$ is always bigger than those required by the proposed method, in which in all cases $M$ is higher than 1.
- The method proposed can be applied using parallel computing, which can substantially reduce the computational effort required. In this case, once initial random partition is done, each of the obtained subsets is processed separately, training a SOM for each of these and making a clustering of each SOM codebook. The rest of the method, i.e. the stages C and D should be computed in a central processor.
- The method can be applied in distributed clustering application, where typically the processing of a big number of data is required [3]. In such case, $M$ nodes collect data and define $M$ subsets. Therefore, the initial random partition is not required. Each node analyzes its data, training a SOM and applying the automatic clustering scheme on the SOM codebook. After that, parameters of the different results of the clustering codebooks are shared between the nodes in order to define cluster prototypes and type-1 Gaussian-shape membership functions. Interval type-2 membership functions and fuzzy predicates are generated in a central node and their parameters are shared with the rest of the nodes. As a result, each node is able to group its data using interval type-2 fuzzy predicates, which collect all the information about the clustering problem available in the rest of the nodes.

For all said, it is possible to conclude that the method proposed allows:

- To be applied in cases with distributed datasets (even in physical different places).
- To improve the traditional clustering algorithm performance.
- To simplify the way in which linguistic knowledge is extracted from data, considering this cannot be done using classical clustering algorithms. This is because of the use of fuzzy predicates, simplified in the proposed method using interval type-2 FL.
- To obtain better accuracies than those obtained by the methods that do not use SOMs.
- To obtain better cluster quality, i.e. more compact and more separated clusters than those obtained by the tested methods, which include both traditional clustering techniques and approaches based on fuzzy predicates previously proposed.
- To make data clustering without any requirements of prior knowledge about the problem, even the number of clusters, which is useful when data exploration is required.
- To discover knowledge expressed by linguistic descriptions, which can be interpreted and modified by experts. Additionally, information about vagueness and variability of the attributes described by the membership functions is available.
- To numerically optimize the membership functions generated if a specific goal is defined, which can improve even more the performance of the proposed method overcoming to both the previous methods SFPC an T2-DFPC.

## 6. Conclusion

In this paper, it is proposed a new SOM-based method for the automatic generation of a clustering system based on interval type-2 fuzzy predicates, called Type-2 SOM-based Fuzzy Predicate Clustering (T2-SFPC), which is based on two previous methods based on fuzzy predicates: the SOM-based Fuzzy Predicate Clustering (SFPC) and the Type-2 Data-based Fuzzy Predicate Clustering (T2-DFPC). The method proposed exploits all the advantages of the SOMs for dealing with noise, outliers, and missing values. The fuzzy predicates are automatically-generated from cluster prototypes extracted from SOMs previously configured and trained and no prior knowledge about the clustering problem is required. Interval type-2 FL is used to quantify the degree of truth of the fuzzy predicates, allowing to model the variability and vagueness in the cluster prototypes.

The method proposed preserves all the advantages of the previous methods SFPC and T2-DFPC concerned to: their capabilities of knowledge extraction from the fuzzy predicates defined, the application of the method in distributed clustering problem and their potential implementation in parallel computing reducing the computational efforts involved. However, the method proposed has important advantages respect to the previous ones. Comparing to the SFPC, the T2-SFPC does not require knowing the number of expected clusters and, considering the SFPC with random data partition, the interpretation of the membership functions and fuzzy predicates obtained from the method T2-SFPC is always easier than when SFPC is used. Related to the method T2-DFPC, the proposed method defines parametrizable interval type-2 membership functions, allowing the optimization of the parameters of the membership functions if a specific goal is defined.

The method proposed was tested on different public datasets and compared with classical clustering approaches and the previous methods T2-DFPC and SFPC. Results showed that the method proposed was consistently one of the best, considering the experiments tested in very different datasets. According to the obtained results for the Silhouette index, the proposed method outperformed all the test methods, being the best in compactness and separation of the clusters found.

For all said, the method T2-SFPC proposed constitutes an unsupervised and general clustering method based on fuzzy predicates, which outperformed the previous methods T2-DFPC and SFPC preserving their main advantages related to interpretable clustering.

As future work, it is proposed to implement a performance analysis of all the previous methods and the proposed one considering datasets with big number of data. Also, a deep noise robustness analysis will be done. In addition, new analysis on the knowledge extraction capabilities of the methods based on self-discovered fuzzy predicates will be done trying to discover new relationships between the shape and size of the membership functions and the characteristics of the data belonging to different clusters.

## Acknowledgement

Authors acknowledge support from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) from Argentina.

## References

[1] A.K. Jain, P.J. Flynn, Data clustering: a review, ACM Comput. Surv 31 (1999) 264–323.
[2] G.J. Meschino, D.S. Comas, V.L. Ballarin, A.G. Scandurra, L.I. Passoni, Automatic design of interpretable fuzzy predicate systems for clustering using self-organizing maps, Neurocomputing 147 (2015), doi:10.1016/j.neucom.2014.02.059.
[3] D.S. Comas, G.J. Meschino, A. Nowé, V.L. Ballarin, Discovering knowledge from data clustering using automatically-defined interval type-2 fuzzy predicates, Expert Syst. Appl. 68 (2017) 136–150, doi:10.1016/j.eswa.2016.10.018.
[4] R.B. Dubey, M. Hanmandlu, S.K. Gupta, S.K. Gupta, The brain MR Image segmentation techniques and use of diagnostic packages, Acad. Radiol. 17 (2010) 658–671.
[5] J. Zexuan, L. Jinyao, C. Guo, S. Quansen, C. Qiang, Robust spatially constrained fuzzy c-means algorithm for brain MR image segmentation, Pattern Recognit. 47 (2014) 2454–2466.
[6] A. Celikyilmaz, I.B. Turksen, Enhanced fuzzy system models with improved fuzzy clustering algorithm, IEEE Trans. Fuzzy Syst. 16 (2008) 779–794.
[7] D.S. Comas, G.J. Meschino, J.I. Pastore, V.L. Ballarin, A survey of medical images and signal processing problems solved successfully by the application of Type-2 Fuzzy Logic, J. Phys. Conf. Ser. 332 (2011), doi:10.1088/1742-6596/332/1/012030.
[8] Z. Deng, Y. Jiang, F.-L. Chung, H. Ishibuchi, S. Wang, Knowledge-leverage-based fuzzy system and its modeling, IEEE Trans. Fuzzy Syst. 21 (2013) 597–609.
[9] C.-F. Juang, S.-H. Chiu, S.-W. Chang, A self-organizing ts-type fuzzy network with support vector learning and its application to classification problems, IEEE Trans. Fuzzy Syst. 15 (2007) 998–1008.
[10] E.G. Mansoori, FRBC: a fuzzy rule-based clustering algorithm, IEEE Trans. Fuzzy Syst. 19 (2011) 960–971.
[11] P. Melin, O. Castillo, A review on the applications of type-2 fuzzy logic in classification and pattern recognition, Expert Syst. Appl. 40 (2013) 5413–5423.
[12] J. Zeng, Z.-Q. Liu, Type-2 fuzzy hidden Markov models and their application to speech recognition, IEEE Trans. Fuzzy Syst. 14 (2006) 454–467.
[13] D.S. Comas, J.I. Pastore, A. Bouchet, V.L. Ballarin, G.J. Meschino, Type-2 fuzzy logic in decision support systems, in: R.A. Espin Andrade, R.Bello Pérez, Á. Cobo, J.Marx Gómez, A.Racet Valdés (Eds.), Soft Comput. Bus. Intell., Springer Berlin Heidelberg, Heidelberg, Germany, 2014, pp. 267–280, doi:10.1007/978-3-642-53737-0_18.
[14] G.J. Meschino, D.S. Comas, V.L. Ballarin, A.G. Scandurra, L.I. Passoni, Using SOM as a tool for automated design of clustering systems based on fuzzy predicates, in: P. Estévez, J. Príncipe, P. Zegers (Eds.), Adv. Self-Organizing Maps, Springer Berlin Heidelberg, Heidelberg, Germany, 2013, pp. 85–94, doi:10.1007/978-3-642-35230-0_9.
[15] G.J. Meschino, R.E. Andrade, V.L. Ballarin, A framework for tissue discrimination in Magnetic Resonance brain images based on predicates analysis and compensatory fuzzy logic, Int. J. Intell. Comput. Med. Sci. Image Process. IC-MED 2 (2008) 209–224.
[16] E.H. Ruspini, A new approach to clustering, Inf. Control. 15 (1969) 22–32.
[17] D.S. Comas, G.J. Meschino, M. Brun, V.L. Ballarin, Label-based Type-2 fuzzy predicate classification applied to the design of morphological W-operators for image processing, in: First Lat. Am. Congr. Comput. Intell., San Carlos de Bariloche, Argentina, 2014, pp. 55–60.
[18] C. Fraley, A.E. Raftery, How many clusters? which clustering method? answers via model-based cluster analysis, Comput. J. 41 (1998) 578–588.
[19] Q. Zhao, V. Hautamaki, P. Fränti, Knee point detection in BIC for detecting the number of clusters, Adv. Concepts Intell. Vis. Syst. 5259 (2008) 664–673.
[20] M. Drobics, U. Bodenhofer, W. Winiwarter, Mining clusters and corresponding interpretable descriptions - a three-stage approach, Expert Syst. J. Knowl. Eng. Neural Netw. 19 (2002) 224–234.
[21] J. Lampinen, E. Oja, Clustering properties of hierarchical self-organizing maps, J. Math. Imag. Vis. 2 (1992) 261–272.
[22] J. Vesanto, E. Alhoniemi, Clustering of the self-organizing map, IEEE Trans. Neural Netw. 11 (2000) 586–600.
[23] K. Tasdemir, P. Milenov, B. Tapsall, Topology-based hierarchical clustering of self-organizing maps, IEEE Trans. Neural Netw. 22 (2011) 474–485, doi:10.1109/TNN.2011.2107527.
[24] A. Ortiz, J.M. Gorriz, J. Ramirez, D. Salas-Gonzalez, Improving MR brain image segmentation using self-organising maps and entropy-gradient clustering, Inf. Sci. (Ny). 262 (2014) 117–136, doi:10.1016/j.ins.2013.10.002.
[25] P. Sarlin, Financial performance analysis of european banks using a fuzzified self-organizing map, in: Proc. 15th Int. Conf. Knowledge-Based Intell. Inf. Eng. Syst., Kaiserslautern, Germany, 2011, pp. 186–195.
[26] M.Y. Kiang, Extending the Kohonen self-organizing map networks for clustering analysis, Comput. Stat. Data Anal. 38 (2001) 161–180.
[27] P. Sarlin, T. Eklund, Fuzzy clustering of the self-organizing map: some applications on financial time series, 8th Int. Work. Self-Organizing Maps, Espoo, Finland, Springer, 2011.
[28] U. Bodenhofer, P. Bauer, Towards an axiomatic treatment of "interpretability,", in: Proc. 6th Int. Conf. Soft Comput. (IIZUKA 2000), Iizuka, Japan, 2000, pp. 334–339.
[29] L.A. Zadeh, Knowledge representation in fuzzy Logic, IEEE Trans. Knowl. Data Eng. 1 (1989) 89–100.
[30] T. Kohonen, Self organized formation of topological correct feature maps, Biol Cybern. 43 (1982) 59–96.
[31] M. Attik, L. Bougrain, F. Alexandre, in: Self-organizing Map Initialization, Springer Berlin Heidelberg, 2005, pp. 357–362, doi:10.1007/11550822_56.
[32] H.-U. Bauer, K.R. Pawelzik, Quantifying the neighborhood preservation of self--organizing feature maps, IEEE Trans. Neural Netw. 3 (1992) 570–579.
[33] G. Pölzlbauer, Survey and comparison of quality measures for self-organizing maps, in: Proc. Fifth Work. Data Anal., 2004, pp. 67–82.
[34] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, Inf. Sci. 8 (1975) 199–249.
[35] J.M. Mendel, Type-2 fuzzy sets and systems: an overview, IEEE Comput. Intell. Mag. 2 (2007) 20–29.
[36] D. Wu, J.M. Mendel, Uncertainty measures for interval type-2 fuzzy sets, Inf. Sci. (Ny). 177 (2007) 5378–5393.
[37] D. Dubois, H. Prade, A review of fuzzy set aggregation connectives, Inf. Sci. (Ny). 36 (1985) 85–121.
[38] A. Bouchet, J.I. Pastore, R.E. Andrade, M. Brun, V. Ballarin, Arithmetic mean based compensatory fuzzy logic, Int. J. Comput. Intell. Appl. 10 (2011) 231–243.
[39] J. Vesanto, Data Exploration Process Based on the Self−Organizing Map, 2002.
[40] T. Kohonen, Self-Organizing Maps, Springer, 1997.
[41] T. Kohonen, MATLAB Implementations and Applications of the Self-Organizing Map, Unigrafia Oy, Helsinki, Finland, 2014.
[42] E.W. Cheney, D.R. Kincaid, Linear Algebra : Theory and Applications, Jones & Bartlett Learning, 2012.
[43] E.R. Dougherty, M. Brun, A probabilistic theory of clustering, Pattern Recognit. 37 (2004) 917–925, doi:10.1016/j.patcog.2003.10.003.

[44] M. Brun, C. Sima, J. Hua, J. Lowey, B. Carroll, E. Suh, E.R. Dougherty, Model-based evaluation of clustering validation measures, Pattern Recognit. 40 (2007) 807–824, doi:10.1016/j.patcog.2006.06.026.
[45] M. Halkidi, Y. Batistakis, M. Vazirgiannis, On clustering validation techniques, J. Intell. Inf. Syst. 17 (2001) 107–145.
[46] E. Atashpaz-Gargari, U.M. Braga-Neto, E.R. Dougherty, Relationship between the accuracy of classifier error estimation and complexity of decision boundary, Pattern Recognit. 46 (2013) 1315–1322.
[47] K. Bache, M. Lichman, 2013. UCI Machine Learning Repository http://archive.ics.uci.edu/ml.
[48] S.N. Ghazavi, T.W. Liao, Medical data mining by fuzzy modeling with selected features, Artif. Intell. Med. 43 (2008) 195–206, doi:10.1016/j.artmed.2008.04.004.
[49] R.A. Fisher, The use of multiple measurements in taxonomic problems, Ann. Eugen. 7 (1936) 179–188.
[50] P. Cortez, A. Cerdeira, F. Almeida, T. Matos, J. Reis, Modeling wine preferences by data mining from physicochemical properties, Decis. Support Syst. 47 (2009) 547–553.
[51] R.K.-S. Kwan, A.C. Evans, G.B. Pike, MRI simulation-based evaluation of image-processing and classification methods, IEEE Trans. Med. Imag. 18 (1999) 1085–1097.
[52] M. Charytanowicz, J. Niewczas, P. Kulczycki, P.A. Kowalski, S. Lukasik, S. Zak, A complete gradient clustering algorithm for features analysis of X-ray images, in: E. Piętka, J. Kawa (Eds.), Inf. Technol. Biomed., Springer Berlin Heidelberg, 2010, pp. 15–24.
[53] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.