Expert Systems With Applications 68 (2017) 136–150

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Discovering knowledge from data clustering using automatically-defined interval type-2 fuzzy predicates

Diego S. Comas [a,b,*], Gustavo J. Meschino [c], Ann Nowé [d], Virginia L. Ballarin [b]

[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, Argentina, Juan B. Justo 4302, Mar del Plata, Argentina (Zip Code B7608FDQ)
[c] Bioengineering Lab, Instituto de Investigaciones Científicas y Tecnológicas en Electrónica (ICyTE), Facultad de Ingeniería, Universidad Nacional de Mar del Plata-CONICET, Argentina, Juan B. Justo 4302, Mar del Plata, Argentina (Zip Code B7608FDQ)
[d] Computational Modeling Lab, Computer Science Department, Vrije Universiteit Brussel, Belgium, Pleinlaan 2, Brussels, Belgium (Zip Code B1050)

ARTICLE INFO

Article history:
Received 28 August 2015
Revised 10 October 2016
Accepted 10 October 2016
Available online 12 October 2016

Keywords:
Fuzzy predicates
Interval type-2 fuzzy logic
Clustering
Knowledge-discovering
Vagueness

ABSTRACT

In data clustering fuzzy predicates act as cluster descriptors providing linguistically expressed knowledge which indicates how features are related to each cluster. Fuzzy predicates directly and automatically obtained from data enable discovering knowledge inside clusters, even when there is no prior-information about the clustering problem. In this work a new method for automatic discovering of interval type-2 fuzzy predicates in data clustering is proposed, called Type-2 Data-based Fuzzy Predicate Clustering (T2-DFPC). In a first stage, a data analysis is performed by making a random partition of the original data and running a clustering scheme that automatically determines the suitable number of clusters. From this stage, interval type-2 fuzzy predicates are discovered. Results obtained on very different clustering datasets show that the T2-DFPC method was consistently one of the best in terms of accuracy. The method preserves all known advantages of the interval type-2 FL to deal with problems with vagueness, quantifying the degree of truth of the fuzzy predicates and modelling the variability of the data inside the clusters. The proposed method is a fast, useful, general, and unsupervised approach for interpretable data clustering, being the knowledge-extracting capabilities one of the main contributions. Linguistic expressions can be easily adapted to match the terminology used in the field the data are related to. The predicates are able to generalize the knowledge for new cases (new data), as an intelligent system. This new approach might be surprisingly useful in contexts where, besides the clustering partition, summary information from data is of interest.

1. Introduction

Data clustering is a set of unsupervised techniques designed to group data, discovering unrevealed structures inside them. Clustering is currently applied in many fields like data mining, business and marketing, machine learning, pattern recognition, image segmentation, information retrieval, bioinformatics, among others (Dubey, Hanmandlu, Gupta, & Gupta, 2010; Hsu, Lin, & Tai, 2011; Meschino, Comas, Ballarin, Scandurra, & Passoni, 2015). New clustering algorithms are continuously proposed trying to solve different aspects.

There are plenty of situations where data is required to be analyzed so as to obtain relevant information from them. Groups of similar data could be one of the possible sources of this information provided that some automatic description of the clusters is given as an output of the algorithm used. Most of known approaches only give a partition of data as a result of applying the algorithm or additionally they discover centroids or prototype data. However, once this output is obtained, the challenge is how to extract information about the clusters, trying to answer questions that may arise, like: What do the data in a particular cluster have in common? How was this partition achieved? Why are these data in the same cluster? How do the values of variables or features differ into different clusters?

Based on Fuzzy Logic (FL) theory, proposed by Zadeh, (1965, 1975), data clustering using fuzzy predicates provides an interesting clustering approach where knowledge about the clustering can be linguistically expressed explaining how data features behave into each cluster, i.e. predicates act as cluster descriptors.

* Corresponding authors: Juan B. Justo 4302, Mar del Plata, Argentina (Zip Code B7608FDQ), Fax: +54-223-481-0046.
E-mail addresses: diego.comas@fi.mdp.edu.ar, diegoscomas@gmail.com (D.S. Comas), gmeschin@fi.mdp.edu.ar (G.J. Meschino), ann.nowe@vub.ac.be (A. Nowé), vballari@fi.mdp.edu.ar (V.L. Ballarin).

http://dx.doi.org/10.1016/j.eswa.2016.10.018
0957-4174/

FL defines degrees of truth of predicates with values between 0 (false) and 1 (true). Fuzzy predicates describing different clusters are evaluated for each datum. Cluster assigned to each datum is typically that whose predicate obtained the highest degree of truth during the evaluation.

Typical applications of fuzzy predicates on data clustering requires knowledge about how features and clusters are related in order to define predicates and membership functions. However, in the last year methods for automatic generation of fuzzy predicates had been proposed (Comas et al., 2012; Comas, Meschino, Brun, & Ballarin, 2014a; Drobics, Bodenhofer, & Winiwarter, 2002; Meschino et al., 2015; Meschino, Comas, Ballarin, Scandurra, & Passoni, 2013). If fuzzy predicates could be directly and automatically obtained from the analysis of the information contained in data, then these predicates would enable not only to perform the data clustering but also and as a distinctive feature to discover knowledge about the clusters hidden in the data, by interpreting the fuzzy predicates obtained.

More common applications of FL in data clustering are based on type-1 FL. In this kind of FL, degrees of truth are limited to single values in the [0, 1] interval. Defining degrees of truth by single values may not be adequate in cases of great imprecision such as data affected by noise or disagreement between different experts when expert knowledge is used, among others (Comas, Pastore, Bouchet, Ballarin, & Meschino, 2014b; Melin & Castillo, 2013; Mendel, 2007). Defining degrees of truth using intervals, interval type-2 FL enables to deal with vagueness and imprecision by modelling the variability in data and knowledge using intervals of truth values (Comas, Meschino, Pastore, & Ballarin, 2011; Comas et al., 2014b). The interval type-2 FL has obtained good results in decision-making applications, including data clustering (Comas et al., 2011; Comas et al., 2012; Melin & Castillo, 2013).

In this paper, a new general method called Type-2 Data-based Fuzzy Predicate Clustering (T2-DFPC) is proposed: it is an interval type-2-FL-based method for the automatic discovering of fuzzy predicates in data clustering. Fuzzy predicates and interval type-2 membership functions are automatically found from data, allowing both data clustering and knowledge discovering in each cluster. The method retains all known advantages of the interval type-2 FL to deal with problems with vagueness and imprecision, quantifying the degree of truth of the fuzzy predicates and modelling the variability in the automatically-extracted-knowledge from data.

First, data are analyzed considering a random partition of the original data (or considering a dataset formed by different subsets that could be even physically separated) running a clustering scheme combining Fuzzy C-Means (FCM) clustering (Ruspini, 1969) with Bayesian Information Criterion (BIC) (Fraley & Raftery, 1998; Zhao, Hautamaki, & Fränti, 2008). This clustering scheme automatically defines the most suitable number of clusters for each dataset. Then, fuzzy predicates and interval type-2 membership functions are discovered. Finally, the predicates can be analyzed by interpreting the knowledge given by the membership functions.

The method proposed defines one fuzzy predicate for each discovered cluster which can be interpreted as *"The datum d belongs to cluster k"*. Given a datum *d* to assign to a cluster, degrees of truth of all fuzzy predicates are computed. As interval type-2 FL is used, the degree of truth are intervals of truth values. In order to compare the resulting intervals of truth values (which is not a trivial task), determining whose with highest degree of truth and performing the cluster assignment, it is proposed here using the concept of "measure of interval of membership values", originally defined in (Comas et al., 2014b) for decision support systems, extending this measure for clustering problem. A detailed analysis of such an extension and its application is given.

The obtained predicates from the T2-DFPC makes a different clustering than other classical algorithms. Linguistic expressions can be adapted to match the terminology of the domain experts. Knowledge discovered from the fuzzy predicates does not require any prior expert information (though it can be used later to improve the system). Results obtained on different dataset using the T2-DFPC method show high accuracy, reaching the performance of other algorithms. As interval type-2 membership functions enable to merge the variability of data of each cluster into a single membership function, a unique fuzzy predicate is used to "explain" each discovered cluster. Therefore, fuzzy predicates are easier interpretable.

This new approach might be surprisingly useful in contexts where, besides the clustering partition, summary information from data is of interest, which is typically the case in a decision support system.

This paper is structured as follows. In Section 2, some important concepts concerning to interval type-2 FL and fuzzy predicates are discussed. Next, in Section 3, the method proposed called T2-DFPC is explained in detail as well as the interval comparing method proposed for the cluster assignment. The clustering results and the knowledge discovering from clustering are presented in Section 4. A discussion of the results and the main method's advantages and limitations is given in Section 5, comparing with existing approaches and current applications. Finally, in Section 6 conclusions and future research directions are mentioned.

2. Interval type-2 FL

This Section is intended to discuss and review basics on interval type-2 FL and fuzzy predicates which are necessary in order to introduce the notation used in the rest of the paper.

FL was defined by Zadeh, (1965, 1975) as a natural extension of Boolean Logic, conceived to deal with linguistic expressions and to work with concepts described by vague or imprecise expressions. FL provides an effective conceptual framework for dealing with the problem of knowledge representation in environments of uncertainty and imprecision as is the case of the human reasoning (Zadeh, 1989). In consequence, FL turns out to be an excellent tool to model and to implement the human reasoning that is typically expressed using linguistic expressions (Comas et al., 2014b; Comas et al., 2012).

FL enables to define the degree of truth of a logic proposition with values between 0 (false) and 1 (true). One limitation of the original notion of FL introduced by Zadeh, called type-1 FL, is that their degrees of truth are limited to single values in the [0, 1] interval. Defining degrees of truth by single values may not be adequate in cases with large imprecision like in the cases of data affected by noise, disagreement between different experts to define the degree of truth of a proposition resulting in imprecise knowledge, and other instances (Comas et al., 2014b; Melin & Castillo, 2013; Mendel, 2007). Interval type-2 FL enables to deal with vagueness and imprecision by modelling the variability in data and knowledge using intervals of values of truth.

Most existing schemes of FL both in data clustering and data classification are based on Fuzzy Inference Systems (FIS). These fuzzy systems use IF-THEN rules with Mamdani or Takagi-Sugeno-Kang approaches (Celikyilmaz & Turksen, 2008; Comas et al., 2011; Deng, Jiang, Chung, Ishibuchi, & Wang, 2013; Juang, Chiu, & Chang, 2007; Mansoori, 2011; Melin & Castillo, 2013) in applications related to: image classification or segmentation (John & Innocent, 1998; Tizhoosh, 2005; Zarandi, Zarinbal, & Izadi, 2011), speech recognition (Zeng & Liu, 2006), control (Lee, 1990), among others. Some detailed reviews can be consulted in (Comas et al., 2011; Melin & Castillo, 2013). A FIS requires defining aggregation and defuzzification operators which means its outcome depends on a pragmatic combination of operators and in general it is a real number (not a label), turning difficult to understand how exactly data

are assigned to labels (Meschino et al., 2015). In addition, using interval type-2 FL with IF-THEN rules has an important limitation related to the computational cost demanded by the defuzzification operation, mainly considering the large number of data in typical clustering and classification problems.

Unlike FIS based on IF-THEN rules, fuzzy predicates are a natural extension of Boolean predicates. In particular, in data clustering, predicates determine the degree of truth in which the data correspond to different clusters. Each cluster is described by one compound fuzzy predicate as *"The datum d belongs to cluster k"* that "explain" the cluster using simple predicates relating characteristics of the values that describes the datum with the belonging of the datum to the cluster. For each datum, the degree of truth of its belonging to each cluster is obtained by applying fuzzy aggregation operators on degrees of truth. Once predicates are evaluated for a datum, a cluster is assigned taking the cluster represented by the fuzzy predicate with the highest degree of truth. In previous works, fuzzy predicates have shown good performance in data clustering using type-1 FL (Meschino et al., 2015), not demanding high computational effort as FIS and requiring computational costs similar to classical clustering algorithms.

For these reasons, in the present paper it is proposed the use of interval type-2 FL in order to model degrees of truth of fuzzy predicates, taking into account the variability in the knowledge that will be extracted from data. Due to the adopted approach, as the degrees of truth of the fuzzy predicates are intervals of truth values, it is necessary comparing intervals of truth values in order to determine which predicate has the highest degree of truth. Such a comparison is done using a methodology that will be proposed in the second part of the next Section.

In the rest of the present Section, the concepts related to interval type-2 FL and fuzzy predicates are formally defined in order to introduce the notation and to clarify.

In traditional approaches based on fuzzy predicates, a predicate is defined as the part of a sentence that tells something about an object which is been described and it is mathematically modelled as a function on a universe of discourse X. In such approaches, predicates become propositions where specific object are considered. However, predicate definition varies from theory to theory. Due to this, in the current work it is adopted the approach given in (Meschino et al., 2015), where the term fuzzy predicate is used to refer to a sentence that assigns properties to objects and their values will be named degrees of truth. Additionally, the terms predicate and proposition are used as synonymous unless it appears necessary to distinguish between them.

The next definitions are adopted:

Definition 1. A type-1 fuzzy predicate p is a linguistic expression (a proposition) with a degree of truth $\nu(p) \in [0, 1]$ (Comas et al., 2014b). The fuzzy predicate concept applies the principle of "gradualism" of FL, which considers that a statement can be both true and false having some degree of truth (or falsehood) assigned (Zadeh, 1965).

Definition 2. An interval type-2 fuzzy predicate p is a fuzzy predicate whose degree of truth is an interval of truth values $\nu(p) = [a, b],\ a, b \in [0, 1]$ (Comas et al., 2014b).

Definition 3. A type-1 membership function over an universe X is a function $\mu: X \to [0, 1]$ (Comas et al., 2014b; Zadeh, 1965).

The membership function associates variable values with degrees of truth. From the point of view of fuzzy predicates $\mu(x)$ with $x \in X$, a membership function μ defines with what degree of truth the value x satisfies the attribute described for the linguistic variable associated to μ.

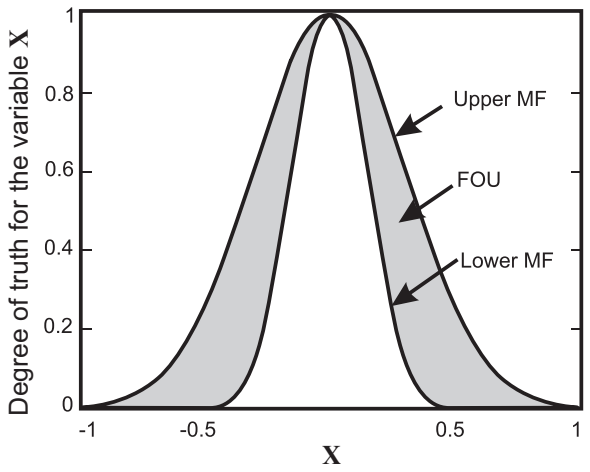

Fig. 1. Interval type-2 membership function for a variable **X** and a generic attribute. Degrees of truth for **X** are bounded for the lower membership function (Lower MF) and the upper membership function (Upper MF). The shaded area represents the FOU.

Definition 4. An interval type-2 membership function $\bar{\mu}$ over an universe X is a function $\bar{\mu}: X \to A$, where $A = \{[a, b]/a \leq b \wedge a, b \in [0, 1]\}$ (Comas et al., 2014b). Values taken by an interval type-2 membership function are interval of truth values.

Definition 5. The functions $\varphi^-_{\bar{\mu}}: X \to [0, 1]$ and $\varphi^+_{\bar{\mu}}: X \to [0, 1]$ represent respectively the lower membership function and the upper membership function of an interval type-2 membership function and they are defined as (Comas et al., 2014b):

$$\begin{aligned} \varphi^-_{\bar{\mu}}(x) &= \min\left(\bar{\mu}(x)\right),\ \forall x \in X \\ \varphi^+_{\bar{\mu}}(x) &= \max\left(\bar{\mu}(x)\right),\ \forall x \in X \end{aligned} \tag{1}$$

Definition 6. The Footprint of Uncertainty (FOU) of $\bar{\mu}$ is the set of all points between the lower and the upper membership function of $\bar{\mu}$, i.e. (Comas et al., 2014b):

$$FOU_{\bar{\mu}} = \bigcup_{x \in X} \left\{\left[\varphi^-_{\bar{\mu}}(x), \varphi^+_{\bar{\mu}}(x)\right]\right\}. \tag{2}$$

In Fig. 1 an interval type-2 membership function for a normalized variable **X** and a generic attribute is shown. The degree of truth defined by $\bar{\mu}$, called interval of truth values or interval of membership values, is bounded for the lower membership function (Lower MF) and the upper membership function (Upper MF). The shaded area represents the FOU. The greater vagueness (variability) is surrounding **X** bigger is the FOU.

Fuzzy predicates can be simple or compound (Meschino et al., 2015):

Definition 7. A simple predicate p_s is a fuzzy predicate whose degree of truth is generally obtained from a membership function, which can be defined by automatic methods or by expert's knowledge.

Definition 8. Two fuzzy predicates p and q are equivalent, this is $p \equiv q$, if and only if $\nu(p) = \nu(q)$.

Definition 9. A compound predicate p_c is a fuzzy predicate equivalent to a logic combination of simple predicates or others compound predicates. The logic combination uses logical operators such as "and" ($\wedge$), "or" ($\vee$), "not" ($\neg$), "implication" ($\Rightarrow$), and "double-implication" ($\Leftrightarrow$).

To define the degree of truth of compound fuzzy predicates, functions that extend Boolean logical operators to operate with FL

are used. The degree of truth of predicates involving the operators "and", "or", and "not" are obtained respectively by means of the fuzzy conjunction (C), disjunction (D), and complement (c) (Comas et al., 2014b). Logical connectives of "implication" and "double-implication" are evaluated using combinations of basic operators.

In the case of interval type-2 FL, the compound fuzzy predicates are evaluated by means of the Zadeh's Extension Principle (Comas et al., 2014b; Zadeh, 1975). Let us consider p and q to be two fuzzy predicates with degrees of truth given for the intervals of truth values $\nu(p) = [\varphi_p^-, \varphi_p^+]$ and $\nu(q) = [\varphi_q^-, \varphi_q^+]$ respectively. According to the Zadeh's Extension Principle the degree of truth of the compound predicate is obtained as follow (Comas et al., 2014b):

Definition 10. The degree of truth of the conjunction between p and q, noted by $p \wedge q$, is a new interval of truth values defined by:

$$\nu(\ p \wedge q) = \left[C\left(\varphi_p^-, \varphi_q^-\right), C\left(\varphi_p^+, \varphi_q^+\right)\right], \tag{3}$$

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

Definition 11. The degree of truth of the disjunction between p and q, noted by $p \vee q$, is a new interval of truth values defined by:

$$\nu(\ p \vee q) = \left[D\left(\varphi_p^-, \varphi_q^-\right), D\left(\varphi_p^+, \varphi_q^+\right)\right], \tag{4}$$

where D: $[0, 1]^2 \rightarrow [0, 1]$ is a fuzzy disjunction.

Definition 12. The degree of truth of the complement of p, noted by $\neg p$, is a new interval of truth values defined by:

$$\nu(\ \neg p) = \left[c\left(\varphi_p^+\right), c\left(\varphi_p^-\right)\right], \tag{5}$$

where c: $[0, 1] \rightarrow [0, 1]$ is a fuzzy complement.

In this work, based on previous successful results (Comas et al., 2014b; Meschino et al., 2013, 2015), three kind of fuzzy operators are considered: compensatory logic operations based on Geometric Mean Based Compensatory Logic (GMCL) and Arithmetic Mean Based Compensatory Logic (AMCL) (Bouchet, Pastore, Andrade, Brun, & Ballarin, 2011) and standard triangular norms (Max-Min) (Dubois & Prade, 1980).

As it is well known, the operators of GMCL and AMCL are sensitive to the whole set of operands being based on geometric and arithmetic means. In contrast, in the widely used operations Max-Min only one of the operands (the maximum or the minimum) dominates the result ignoring the values of the remaining operands.

Considering compensatory operators, the value of conjunctions and disjunctions can be influenced by and therefore "compensated" for the value of any of the degrees of truth considered in the operation. An increase (or a decrease) in the degree of truth of a conjunction or a disjunction as results of changes in the degree of truth of one component can be compensated by an increase (or a decrease) of the degree of truth of another component. This characteristic makes compensatory FL especially suited for selection problems; yet it is also convenient for ranking, appraising, and classificatory purposes (Meschino et al., 2015).

3. Proposed methods

In order to make more understandable the methods proposed, the present Section is divided in two sub-sections. First, the method called Type-2 Data-based Fuzzy Predicate Clustering (T2-DFPC) is explained in detail. Then, interval comparing method proposed in the present work to assign cluster to data is presented.

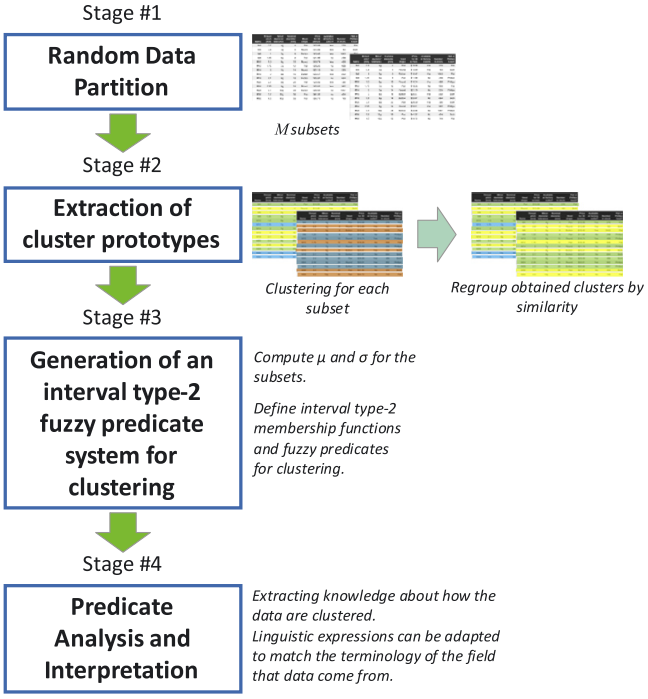

Fig. 2. Processing pipeline for the T2-DFPC method. Stage #1: Random data partition, generating M disjoint subsets. Stage #2: Extraction of cluster prototypes. Stage #3: Generation of an interval type-2 fuzzy predicate system for clustering. Stage #4: Predicate analysis and interpretation.

3.1. Method Type-2 Data-based fuzzy predicate clustering (T2-DFPC)

The method proposed called Type-2 Data-based Fuzzy Predicate Clustering (T2-DFPC) is described in detail in this Section. It automatically defines fuzzy predicates and interval type-2 membership functions from data which enable: (a) to perform the clustering of the data and (b) to explain how the clustering is performed by extracting linguistic knowledge about the clustering problem. Each cluster can be explained by only one fuzzy predicate in a simple way. An expert in the field where the data come from could be able to give linguistic meaning to the membership functions and the predicates automatically discovered. The method proposed can be split into four stages: *1) Random data partition, 2) Extraction of cluster prototypes, 3) Generation of an interval type-2 fuzzy predicate system for clustering*, and *4) Predicate analysis and interpretation*.

Using the data to be clustered, a fuzzy clustering system based on fuzzy predicates and interval type-2 FL is automatically designed. In the first stage, a random partition on data is performed. Data contained in each subset are analyzed in stage #2, applying an automatic crisp clustering scheme on each subset, which automatically defines the suitable number of clusters. The clustering scheme combines the BIC (Fraley & Raftery, 1998) with the FCM algorithm (Ruspini, 1969) following the method developed in (Zhao et al., 2008). The obtained clustering enables to analyze the variability of the data into each cluster and to extract cluster prototypes, i.e. representative data of each discovered cluster. By analyzing the obtained cluster prototypes in stage #3, interval type-2 membership functions and fuzzy predicates describing the data clusters are defined. Finally, in the stage #4, the obtained fuzzy predicates are analyzed and knowledge about the clustering is obtained giving a linguistic meaning related to how the clustering is made.

Fig. 2 shows the processing pipeline for the T2-DFPC method. A detailed description of the steps of each method stage is made in the remainder of this Section.

3.1.1. Stage #1: random data partition

Let $X \subset [-1,1]^n$ be the dataset to be clustered, where N is the number of data in X and n is the dimension of the space of data. The first stage of the method consists of making a random partition on X generating M disjoint subsets of X, namely $\{P_m\}_{m=1,...,M}$, $X = P_1 \cup P_2 ... \cup P_M$. The partition size M is a method parameter. It is desirable that each subset is balanced; i.e. each P_m has approximately the same number of data points in each cluster.

Each subset P_m represents different instances of the same clustering problem. Therefore, it is expected that clustering the data in the P_m, $m = 1, ..., M$, will result in clusters on similar but not necessarily the same regions of the data space. Therefore, the clustering performed on the different subsets could reveal different characteristics of the data space regarding to the distribution of the clusters. Consequently, the clustering on each P_m works as a different "opinion" of how clustering should look like, capturing the variability of the data inside of the clusters. This consideration is used to make the interval type-2 membership functions capturing these aspects of the clustering problem.

3.1.2. Stage #2: extraction of cluster prototypes

In this stage, automatic clustering of each P_m, $m = 1, ..., M$, obtained in the stage #1 is performed. It is used FCM as clustering algorithm (Ruspini, 1969) combining with the BIC (Fraley & Raftery, 1998) for the automatic determination of the number of clusters, following the method developed in (Zhao et al., 2008), requiring no parameter related to the number of cluster.

The number of clusters obtained in each P_m could be different considering that the number of cluster obtained with the automatic clustering scheme depends on the data in each P_m. As a result, for a specific P_m the clustering result can be expressed as $\{\Delta_{m,j}\}_{j=1,...,\kappa_m}$ where $P_m = \Delta_{m,1} \cup ... \Delta_{m,j} ... \cup \Delta_{m,\kappa_m}$ and κ_m is the number of cluster obtained for the subset P_m.

After running the clustering algorithm on the $\{P_m\}_{m=1,...,M}$ subsets, a second clustering on the obtained cluster centroids is performed, regrouping the original clusters of the partitions by similarity. Let $\{Y_{m,j}\}_{\substack{m=1,...,M \\ j=1,...,\kappa_m}}$ be the set of cluster centroids obtained from the clustering of the $\{P_m\}_{m=1,...,M}$, where κ_m is the number of clusters discovered in the subset m, $m = 1, \ldots, M$. In this step, K clusters are defined for the whole set of centroids; where K is defined as the maximum of the number of clusters obtained by clustering the subsets P_m, $m = 1, ..., M$, i.e. K is the maximum of $\{\kappa_m\}_{m=1,...,M}$. In order to regroups the centroids, FCM with crisp clustering is used, grouping the centroids $\{Y_{m,j}\}_{\substack{m=1,...,M \\ j=1,...,\kappa_m}}$ into K clusters.

As a result, groups of the clusters of the subsets P_m, $m = 1, ..., M$ are obtained, which are represented by their centroids. The set of centroids is denoted as $\{\Omega_{l,k}\}_{\substack{l=1,...,L_k \\ k=1,...,K}}$, where L_k is the number of the original clusters of the subsets that were reassigned to the cluster k. The data, now reassigned to the cluster k, will be noted as $\Gamma_{l,k}$, with $l = 1, ..., L_k$ and $k = 1, ..., K$. These data describe different characteristics of a same cluster on the original dataset X. Considering the data in $\Gamma_{l,k}$ for a fixed k, different descriptions or "opinions" for the cluster k can be extracted. These "opinions" act capturing the variability of the data inside the cluster discovered and will be used in the next stage to define interval type-2 fuzzy predicates explaining the clusters. In Fig. 3a detailed diagram of the stages #1 and #2 of the method is shown.

3.1.3. Stage #3: generation of an interval type-2 fuzzy predicate system for clustering

The data contained in $\{\Gamma_{l,k}\}_{\substack{l=1,...,L_k \\ k=1,...,K}}$ are analyzed in this stage defining interval type-2 membership functions and fuzzy predicates that enable to group the original dataset as a whole and

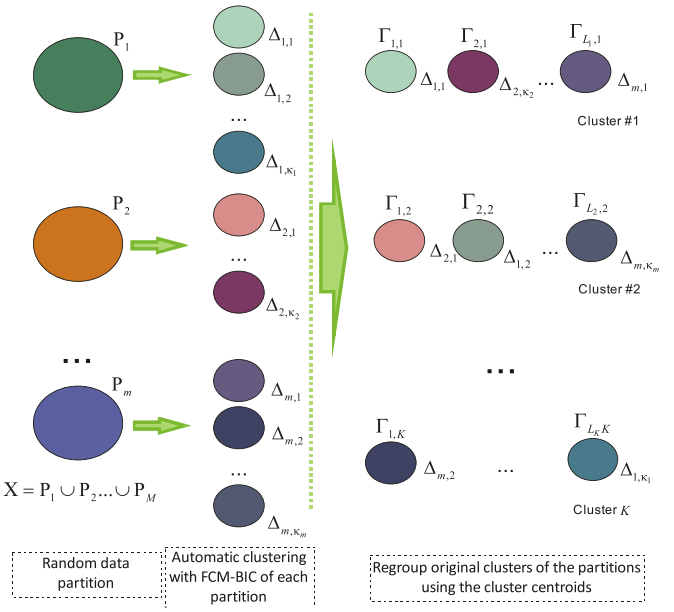

Fig. 3. Detailed diagram of the stages #1 and #2 of the proposed method.

to interpret how the clustering is performed. The final clustering result is obtained evaluating the fuzzy predicates defined in this stage, using the method detailed in the Section 3.2 to compare the resulting intervals of truth values.

The present stage consists of the next steps:

(1) Type-1 membership functions are generated analyzing the data contained in $\{\Gamma_{l,k}\}_{\substack{l=1,...,L_k \\ k=1,...,K}}$. Gaussian membership functions are used, although other shapes could also be considered. For each feature and each cluster, the centers of the type-1 Gaussian membership functions are the cluster centroids contained in the set of centroids $\{\Omega_{l,k}\}_{\substack{l=1,...,L_k \\ k=1,...,K}}$ obtained previously. The widths of the Gaussian membership functions are computed as the standard deviations of the data contained in $\{\Gamma_{l,k}\}_{\substack{l=1,...,L_k \\ k=1,...,K}}$.

As a consequence, a type-1 Gaussian membership function is defined for each feature i and each cluster k. As there are L_k data subsets for each cluster, L_k type-1 Gaussian membership functions are made for each combination of cluster and feature.

The centers of these Gaussian functions are the cluster centers obtained in the previous step; namely $\{\omega_{i,l,k}\}_{\substack{i=1,...,n \\ l=1,...,L_k \\ k=1,...,K}}$, where $\omega_{i,l,k} \in \Omega_{l,k}$. Then, the standard deviations of the data in each cluster, noted by $\{\sigma_{i,l,k}\}_{\substack{i=1,...,n \\ l=1,...,L_k \\ k=1,...,K}}$, for each feature and each subset in $\{\Gamma_{l,k}\}_{\substack{l=1,...,L_k \\ k=1,...,K}}$ are computed.

Considering that the type-1 membership functions created; noted by $\{\mu_{i,l,k}\}_{\substack{i=1,...,n \\ l=1,...,L_k \\ k=1,...,K}}$, where $\mu_{i,l,k} : [-1,1] \to [0,1]$; indicate the correspondence between feature values and clusters, the standard deviation controls the width (sigma) of the Gaussian functions. The standard deviation acts as a parameter controlling how the degree of truth of the membership functions decreases when the values of the feature for a specific datum moves away from the cluster center. That is, for feature values close to the cluster center, the corre-

spondence between the values and the cluster is high and it decreases as values move away from the centroid.

(2) For each cluster $k \in \{1, ..., K\}$ and each feature $i \in \{1, ..., n\}$, a unique interval type-2 membership function noted by $\bar{\mu}_{i,k}$ is now defined, combining the previous type-1 membership functions using fuzzy operators. The lower membership function and the upper membership function of $\bar{\mu}_{i,k}$ are computed respectively as:

$$\begin{aligned} \varphi^{-}_{\bar{\mu}_{i,k}}(x) &= C\left(\mu_{i,1,k}(x), ..., \mu_{i,L_k,k}(x)\right) \\ \varphi^{+}_{\bar{\mu}_{i,k}}(x) &= D\left(\mu_{i,1,k}(x), ..., \mu_{i,L_k,k}(x)\right) \end{aligned}, \quad (6)$$

$\forall x \in [-1,1]$, $i \in \{1, ..., n\}$, and $k \in \{1, ..., K\}$; where C is the conjunction operator of the AMCL (Bouchet et al., 2011) and D is the s-norm called Algebraic sum (Comas et al., 2014b) computed by pairs over the degrees of truth $\mu_{i,l,k}(x)$, $l = 1, ..., L_k$.

The interval type-2 membership functions are defined merging with fuzzy operators all partial descriptions obtained from the M subsets which are contained in the type-1 membership functions $\mu_{i,\,l,\,k}$, with $l = 1, ..., L_k$ for a feature i and a cluster k. The descriptions say what features values have more correspondence with the characteristics collected by the cluster.

(3) Finally, one fuzzy predicate is made for each cluster $k \in \{1, ..., K\}$ (K compound predicates) by logically operating with the degrees of truth defined by the interval type-2 membership functions made in the step 2. For each cluster, the next fuzzy predicate is defined:

$$\begin{aligned} &p_k(d) \equiv \bar{\mu}_{1,k}(d_1) \wedge \bar{\mu}_{2,k}(d_2) \wedge, \ldots, \wedge \bar{\mu}_{n,k}(d_n), \\ &k = 1, 2, \ldots, K \end{aligned}. \quad (7)$$

where d is a generic datum to be assigned to the cluster k.

The fuzzy predicate $p_k(d)$ enables to weigh with what degree of truth the values of datum d correspondence to (are close to) the characteristics detected around the cluster k.

The degrees of truth of all predicates $\{p_k\}_{k=1,...,K}$ are computed using the interval type-2 membership functions and fuzzy operators. Cluster assignment is performed obtaining the predicate with the highest degree of truth using the method detailed in Section 3.2.

3.1.4. Stage #4: predicate analysis and interpretation

The fuzzy predicates defined in the stage #3 can be easily interpreted by analyzing the position of the respective membership functions in the feature range.

First, the predicate $p_k(d)$ can be linguistically read as *"The datum d belongs to cluster k"* and $\bar{\mu}_{i,k}(d_i)$ can be linguistically interpreted as *"The feature i in the datum d corresponds to cluster k"* or *"The feature i in the datum d is near to the centroid of cluster k"*. The nearer the value of feature i in the datum d to the value of the maximum of $\bar{\mu}_{i,k}$, the higher the degree of truth of $\bar{\mu}_{i,k}(d_i)$. As $\bar{\mu}_{i,k}(d_i)$ is higher, $p_k(d)$ should also be higher, reflecting the fact that if the datum d is near the centroid of the cluster k, then the datum d belongs to cluster k.

The generic linguistic interpretations given before can be replaced by linguistic attributes analyzing the positions of the membership functions $\bar{\mu}_{i,k}(d_i)$ related to the range of the feature. For instance, let us consider the interval type-2 membership functions shown in Fig. 4, which could be obtained using the T2-DFPC method for a generic dataset. According to Fig. 4, three clusters have been discovered represented with different colors. Blue color represents the *cluster #1*. Red color represents the *cluster #2*. Yellow color represents the *cluster #3*.

By simple inspection it is easy to see that the *cluster #1* (blue) is related to low values of the *feature #1* and high values of *feature #2*. In the same way, *cluster #2* (red) is related to medium values of the *feature #1* and low values of *feature #2*. Finally, *cluster #3* (yellow) is related to high values of the *feature #1* and medium values of *feature #2*.

Under these considerations for this example the generic fuzzy predicates obtained from the stage #3 of T2-DFPC can be rewritten as follows:

- $p_1(d)$: "The datum d belongs to cluster #1″ is equivalent to "The feature #1 in the datum d is low and the feature #2 in the datum d is high".
- $p_2(d)$: "The datum d belongs to cluster #2″ is equivalent to "The feature #1 in the datum d is medium and the feature #2 in the datum d is low".
- $p_3(d)$: "The datum d belongs to cluster #3″ is equivalent to "The feature #1 in the datum d is high and the feature #2 in the datum d is medium".

The *"feature i"* denominations can be replaced for the name of the features. More specific interpretation can be given for experts in the field where the data come from. In the Section 4.2 an example of the interpretation of the fuzzy predicates obtained for an actual application case is presented.

Further information can be analyzed using the obtained membership functions and predicates. More specifically, analyzing both the area of the resulting FOU and the width of the membership functions it is possible to evaluate the vagueness or variability around of the clusters discovered. As a result, the same analysis can be done on the features and attributes associated to each cluster.

In this regard, large FOU implies large variations between the data contained in the subsets obtained from the partition in the stage #1. As this partitioning process is random, it is expected that each subset generated is statistically representative of the original dataset X. By consequence, each subset contains different instances of the same clustering problem and, therefore, the clustering on the subsets can reveal different "opinions" of how clustering of the whole dataset X should look like. These different "opinions" are partially modelled using type-1 membership functions in a first step and then these are merged into interval type-2 membership functions using fuzzy operators in the stage #3.

Due to the previous considerations, membership functions with large FOU (for a given cluster) indicate that there is a large vagueness (i.e. variability or disagreement) about the degree of truth of the correspondence between feature values and the cluster. Also, large FOU are associated with great variation between the different "opinions" modelled by the type-1 membership functions.

From the feature point of view, it suggests that there is a great vagueness about the attribute modelled by the membership function as well as about the importance of this to assign a data point to the corresponding cluster. On the other hand, wide membership functions are related to a large variability inside a particular cluster. Consequently, both aspects provide information on different aspects of the clustering problem.

This kind of analysis can be used on datasets in different application domain, once the T2-DFPC is applied. Besides the linguistic interpretation of the fuzzy predicates and depending on each particular case, both the width and the shape of the membership functions can reveal important characteristics of problem addressed.

3.2. Interval comparing: method proposed for data clustering using interval Type-2 FL

The fuzzy predicates $\{p_k\}_{k=1,...,K}$ obtained from the T2-DFPC method enables to perform the clustering of the original dataset X. The degrees of truth, which are intervals of truth values, are

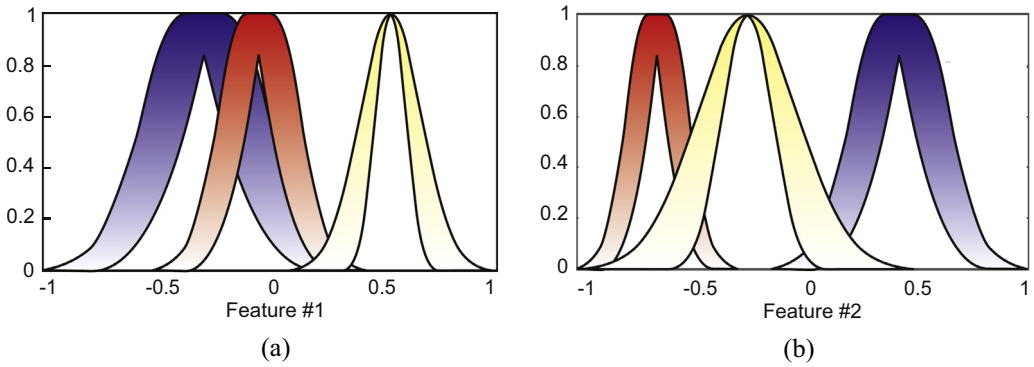

Fig. 4. Example of obtained membership functions from T2-DFPC for a generic dataset. a) Membership functions obtained for the *"Feature #1"*. b) Membership functions obtained for the *"Feature #2"*. Three clusters were discovered. Blue color corresponds to *cluster #1*. Red color corresponds to *cluster #2*. Yellow color corresponds to *cluster #3*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

computed using fuzzy operators according to the definition given in Section 2.

As it was explained, the cluster assignment requires interval comparing. Such a comparison is required to achieve a ranking between intervals in all the possible cases, i.e. both non-overlapping, overlapping and one interval included into the other one. After interval comparing, the predicate with the highest degree of truth for each datum can be obtained, determining the cluster assignment.

Most of existing methods for interval comparing are focused in preference selection and do not completely cover the case where one interval is included in the other, specifically, intervals with same mean value. Such methods include the works (Ishibuchi & Tanaka, 1990; Kundu, 1997; Moore & Lodwick, 2003; Sengupta & Pal, 2000), among others.

In the other hand, in (Comas et al., 2014b) the concept of "measure of interval of membership values" is introduced enabling to describe degrees of truth represented by intervals of truth values using real numbers and to define a ranking for all possible cases. Such an approach was proposed for decision support systems. On the basis of that procedure, in the present section an extension is proposed in order to solve the cluster assignment problem. The new proposal adds value to the original proposal, not only because it includes a new field for its application, but also because it incorporates an analysis of necessary properties for a ranking which will be used in clustering. This new approach consists of the definition of a new function called "measure of interval of truth values" and a detailed description of this approach is given below.

Let us consider an interval of truth value $v(p(d)) = [\varphi^-_{p(d)}, \varphi^+_{p(d)}]$ obtained from evaluating a fuzzy predicate p in a datum (a value) d. Let χ 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 degree of truth represented by the interval $[\varphi^-_{p(d)}, \varphi^+_{p(d)}]$ can be described using the function $f : \chi \to \mathbb{R}^+$ given by:

$$f\left(\left[\varphi^-_{p(d)}, \varphi^+_{p(d)}\right]\right) = \frac{\varphi^-_{p(d)} + \varphi^+_{p(d)}}{2}\varphi^+_{p(d)}, \tag{8}$$

where the function f is called measure of interval of truth values. This measure defines a real number for all the possible interval of truth values, including the cases of non-overlapping, overlapping and one interval included in the other. The higher the value of f the higher the degree of truth.

As it can be seen, f combines the mean value of the interval with its maximum. The following analysis can be done:

- Considering clustering applications, interval type-2 membership functions have maximum values close to those that better meet the properties of the clusters. In the case of the T2-DFPC method, the maximum is close to the centroid of the discovered cluster. In this sense, given a datum and two clusters, the datum will be closer to the maximum of the membership functions that describes the cluster for which the datum better meets its properties. Therefore, the mean value of the interval of truth values resulting of evaluating the predicate for that cluster will be higher giving as a consequence a higher value of f.
- Given two intervals of truth values with same mean value, resulting of the predicate evaluation for two different clusters and a datum, that interval with higher maximum value indicates a higher joint fulfillment of the cluster properties. For this reason, the maximum value of the interval is included in f.

Considering the previous analysis, $f$can be used for inducing a ranking between the intervals of truth values resulting of evaluate the predicates $\{p_k\}_{k=1,\dots,K}$ in a datum $d$and, in consequence, this ranking can be used for assign a cluster to the datum. In addition, the next properties are satisfied by f:

- For the interval $[0, 0] = 0$ (the minimum interval of truth values) $f([0, 0]) = 0$.
- For the interval $[1, 1] = 1$ (the maximum interval of truth values) $f([1, 1]) = 0.5$.
- Given two intervals $[a, a] = a$, $[b, b] = b$ where $a < b$, then $f([a, a]) = a^2 < f([b, b]) = b^2$, following in these case the ranking that can be obtained using type-1 FL.
- The ranking of intervals of truth values induced by f is transitive.

On the basis of all the previous observations, the method proposed for data clustering using interval type-2 FL and fuzzy predicates is given below.

Let us consider a fuzzy system for data clustering formed by a set of K compound fuzzy predicates $\{p_k\}_{k=1,\dots,K}$ and interval type-2 membership functions $\{\bar{\mu}_{i,k}\}_{\substack{i=1,2,\dots,n \\ k=1,2,\dots,K}}$, which could be obtained using the T2-DFPC method or others methods. Each p_k is interpreted as $p_k(d)$: *"The datum d belongs to cluster k,"* describing one of the clusters. Given a datum $d \in [-1, 1]^n$ to be assigned to a cluster, where n is the dimension of the normalized data space, the clustering steps are:

1) Compute the degrees of truth of the K compound fuzzy predicates $\{p_k\}_{k=1,\dots,K}$ for the datum d, using the membership functions and fuzzy operators, resulting in the degrees of

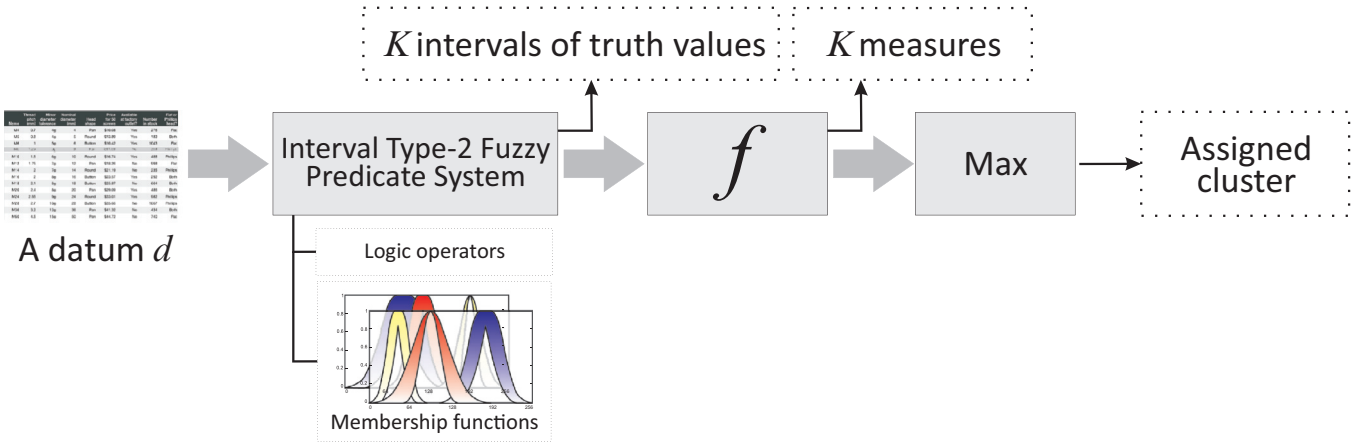

Fig. 5. Proposed method for data clustering using interval type-2 FL and fuzzy predicates. Given a datum d to be assigned to a cluster, all the fuzzy predicates are evaluated obtaining K intervals of truth values (one per cluster). Then, the cluster is assigned considering the predicate with the highest measure of interval of truth values.

truth $\{\nu(p_k(d))\}_{k=1,...,K}$; where $\nu(p_k(d)) = [\varphi^-_{p_k(d)}, \varphi^+_{p_k(d)}]$ is an interval of truth values.

2) Apply the measure of intervals of truth values f to the intervals $\{\nu(p_k(d))\}_{k=1,...,K}$, obtaining the values:

$$\{f(\nu(p_k(d)))\}_{k=1,...,K} = \left\{f\left(\left[\varphi^-_{p_k(d)}, \varphi^+_{p_k(d)}\right]\right)\right\}_{k=1,...,K}. \quad (9)$$

3) Assign to the datum d the cluster corresponding to the fuzzy predicate with the highest value of measure of intervals of truth values, i.e., it is assigned the cluster $k' \in \{1, ..., K\}$ where $p_{k'}$ is such that:

$$f(\nu(p_{k'}(d))) = \max\{f(\nu(p_k(d)))\}_{k=1,...,K}. \quad (10)$$

The fuzzy operators used are typically defined and adjusted according to data clustering problem. Results would be different depending on the fuzzy operators. This methodology can be used for any data clustering problem where interval type-2 FL and fuzzy predicates are used. The methodology proposed is summarized in Fig. 5.

4. Results

In this Section, results of the method T2-DFPC proposed are presented. First, clustering assessment is presented, comparing results of the method proposed with those of existing clustering methods. At the end of this Section, in order to analyze the predicate interpretation and knowledge discovering suggested in the stage #4 of the method, the procedure is explained in detail for a real dataset.

4.1. Clustering assessment

Clustering assessment is not a trivial task, requiring the proper selection of validation indices according to each particular case (Halkidi, Batistakis, & Vazirgiannis, 2001). The selection depends on the aim of the clustering application, i.e. if the clustering is used for data exploration, generalization of a previous labeling, modeling, among others. If the pursued objective is the data exploration, it is necessary automatically finding compact clusters and parameter optimization is used, for instance including the BIC. In such cases, internal validation indices can be used for validation purposes. Conversely, if the generalization of previous labeling is needed, then the number of clusters is known and external validation indices such accuracy can be used in order to assess the clustering obtained.

On the basis of these considerations and taking in to account the possible applications of the method proposed, the clustering assessment was performed using the accuracy measure, estimating the clustering quality obtained with the T2-DFPC method applied to the next nine public datasets previously labelled:

- Wine dataset (3 clusters, 13 features, 198 data) (Cortez, Cerdeira, Almeida, Matos, & Reis, 2009).
- Iris dataset (3 clusters, 4 features, 150 data) (Fisher, 1936).
- MRI1 dataset, 4000 pixels randomly selected per cluster from simulated magnetic resonance images (3 clusters, 3 features, 12,000 data) (Kwan, Evans, & Pike, 1999). These data were taken without any noise or distortion, and they come from computer simulations.
- MRI2 dataset, 200 pixels randomly selected per cluster from the previous dataset (3 clusters, 3 features, 600 data) (Kwan et al., 1999).
- Wisconsin breast cancer dataset (32 features - 3 selected, 2 clusters, 569 data) (Ghazavi & Liao, 2008).
- Pima Indians diabetes dataset (8 features - 3 selected, 2 clusters, 768 data) (Ghazavi & Liao, 2008).
- Moon dataset (2 features, 2 clusters, 2000 data). This is a synthetic dataset whose characteristics will be presented in the next subsection.
- Seeds dataset (7 features, 3 clusters, 210 data) (Charytanowicz et al., 2010).
- Banknote dataset (5 features - 2 selected, 2 clusters, 1372 data) (Bache & Lichman, 2013).

Previous to the accuracy computing, each discovered cluster was assigned to one and only one real label in the gold standard result considering the majority labels in each cluster. To be independent of the algorithm initialization, each clustering algorithm was run 20 times, reporting accuracy estimations corresponding to accuracy averages. The analysis included standards deviation estimation and statistical tests of significance.

In addition, internal validation indices such as Dunn and Silhouette indices were also estimated, revealing none significant differences between the T2-DFPC and the test clustering methods.

Graphs are used to help the comparisons. The results showed that the proposed method (T2-DFPC) outperformed or at least achieved the results of the others, but unlike the algorithms not based on fuzzy predicates, in the proposed approach the clustering could be interpreted.

For the Wisconsin breast cancer and Pima Indians diabetes datasets, in order to facilitate result comparisons, only some features were used as it is suggested in (Meschino et al., 2015).

In the case of the Banknote dataset only the two first features were used, considering the best clustering accuracy obtained for

the T2-DFPC using full search selection. Then, all clustering algorithms were run with the same selected features.

The clustering methods used in the tests were:

- **T2-DFPC**: Type-2 Data-based Fuzzy Predicate Clustering (the method proposed). It used an automatic clustering scheme combining FCM and the BIC. Information to generate interval type-2 membership functions was taken directly from data clustering results.
- **DFPC**: Data-based Fuzzy Predicate Clustering. This is similar to the method proposed but type-1 FL was used. No data partition was made. In consequence, the whole dataset was used to make type-1 membership functions following similar considerations to the T2-DFPC.
- **SOM-FCM** (Meschino et al., 2015): a Self-Organizing Maps (SOM) was trained with the dataset and an automatic FCM-BIC clustering scheme was applied to the its codebook (configuring a two-level clustering scheme).
- **K-means** (Jain & Flynn, 1999): the K-means algorithm combining with the BIC was applied to the dataset determining the proper number of clusters in each case.
- **FCM** (Ruspini, 1969): FCM algorithm was combining with the BIC and was applied to the datasets considering random initial centroids. As a result, K centroids were obtained and data were assigned to the cluster with the highest membership value.
- **EM**, Expectation-Maximization (Bishop, 2006): it is a very known method used to fit a Gaussian mixture models to a dataset. Probability density functions characterizing the dataset were obtained. Then probability of belonging to each cluster was determined, assigning the most likely. The number of cluster was a method parameter.

Additionally, variants of the methods: T2-DFPC, DFPC, SOM-FCM, K-means, and FCM were considered too; by replacing the clustering scheme with the BIC by only the clustering algorithm. In this case it was needed using the number of cluster K in each dataset as method parameter. In the next analysis, these clustering algorithm variants were respectively called: T2-DFPC-*wac*, DFPC-*wac*, SOM-FCM-*wac*, K-means-*wac*, and FCM-*wac*.

Clustering accuracies obtained for the different datasets are shown in Fig. 6. The accuracy is represented by vertical bars. In the cases of methods that use fuzzy predicates, three bars are shown corresponding to different FL operators: standard triangular norms (Max-Min operators), and compensatory operators (GMCL and AMCL). A horizontal line shows the best result achieved by the method proposed. A detailed analysis of the clustering results is done in the next paragraphs for the different datasets. When two different accuracy results acc_1 and acc_2 are compared, a percentage difference of the accuracies is used computed as $difference\,\% = \frac{acc_1 - acc_2}{acc_2} \times 100\%$. Also, results of statistical tests of significance for the differences obtained are reported.

For the Wine dataset (Fig. 6a) the accuracy obtained with the proposed method was only improved by 0.5% ($p < 0.05$) for the DFPC method, a fuzzy predicate method using type-1 FL. T2-DFPC outperformed by 0.1% ($p < 0.05$) to the SOM-FCM-*wac*, the best of the classical clustering method for this dataset.

In the Iris dataset (Fig. 6b) the proposed method outperformed all classical clustering methods by more than 1.4% ($p < 0.05$). However, T2-DFPC accuracy was outperformed by 1.5% for the DFPC-wac ($p < 0.05$).

In the cases of the MRI1 and MRI2 datasets (Fig. 6c and d) the T2-DFPC was one of the best, but it was outperformed for some of the testing methods. However, this differences were not significant according to the values obtained for the significance test. In both datasets, the variant without automatic clustering T2-DFPC-*wac* outperformed the proposed method by around of 1.5%.

In the Wisconsin breast cancer (Fig. 6e), the obtained result indicated a good clustering performance of the proposed methods. Some classical clustering algorithms outperformed the accuracy of the T2-DFPC by 1.0% ($p < 0.05$).

In the dataset Pima Indians diabetes (Fig. 6f), the obtained accuracy was 0.712, overcome by 0.5% by the SOM-FCM algorithm, but that difference was not significant. Analyzing the clustering techniques based on fuzzy predicates, the proposed method had the best accuracy, outperforming the DFPC by 1.7% ($p < 0.05$).

For Moon dataset (Fig. 6g), the proposed method was the best clustering method. The method overcomes by 3.9% the classical clustering ($p < 0.05$) methods, having equal performance that DFPC.

The results of Seeds dataset set (Fig. 6h) reflected an improvement of 0.8% ($p < 0.05$) of the method proposed over the classical clustering methods, while the performances of the fuzzy predicates clustering methods were very similar without significant differences.

Finally, for the Banknote dataset (Fig. 6i) the clustering performance obtained with the T2-DFPC was improved by 0.7% by the K-means method ($p < 0.05$) and by 2.1% by the T2-DFPC-*wac* ($p < 0.05$).

Summarizing, even when the method proposed was overcome in some cases; it was always one of the best, considering the experiments done testing very different datasets. When method proposed was neither in the first nor in the second place, the accuracy was only a little lower. This preliminary evidence makes the approach reliable for data clustering. However, a major contribution of the proposed method is the interpretability of the clusters included in the stage #4 of the method not available in the others test methods.

Comparing with type-1 FL, in the cases where the method proposed had not a better performance than the based on type-1 FL, differences were small, never exceeding 1.7%.

On the other hand, the clustering algorithms using the BIC showed in general good performance in comparison with the respective clustering algorithms without automatic clustering (the -*wac* denominations), showing only minor differences in the accuracy values.

4.2. Predicate interpretation example: how the "knowledge discovering" could be performed

The knowledge about the clustering problem discovered by the T2-DFPC method could be easily extracted analyzing the interval type-2 membership functions and the fuzzy predicates as was explained in the stage #4 of the method.

In order to clarify, in the present Section the interpretation of the predicates discovered for the wide known Iris dataset is given. This dataset consists of 150 samples of iris genus and four features per sample: Sepal Length, Sepal Width, Petal Length, and Petal Width. The problem consists in assign each datum to one of three possible clusters: Iris Setosa, Iris Versicolor, and Iris Virginica. In Fig. 7 the membership functions obtained for this dataset are shown. Different colors indicate different clusters: Iris Setosa (red), Iris Versicolor (blue), and Iris Virginica (yellow).

First, let us consider the cluster Iris Setosa, shown in red color in the Fig. 7. The membership functions were associated to different attributes for different features. This association was performed by analyzing the position of each membership function in the range of each feature as well as the feature meaning. As a result, the Iris Setosa cluster could be related to: *"small"* values of *"Sepal Length"*, *"slightly large"* values of *"Sepal Width"*, *"very small"* values of *"Petal Length"*, and *"very small"* values of *"Petal Width"*.

Using the procedure suggested in stage #4 of the method T2-DFPC, the fuzzy predicate associated with the Iris Setosa cluster could be rewritten as: $p_{Setosa}(d)$: *"The datum d belongs to Iris*

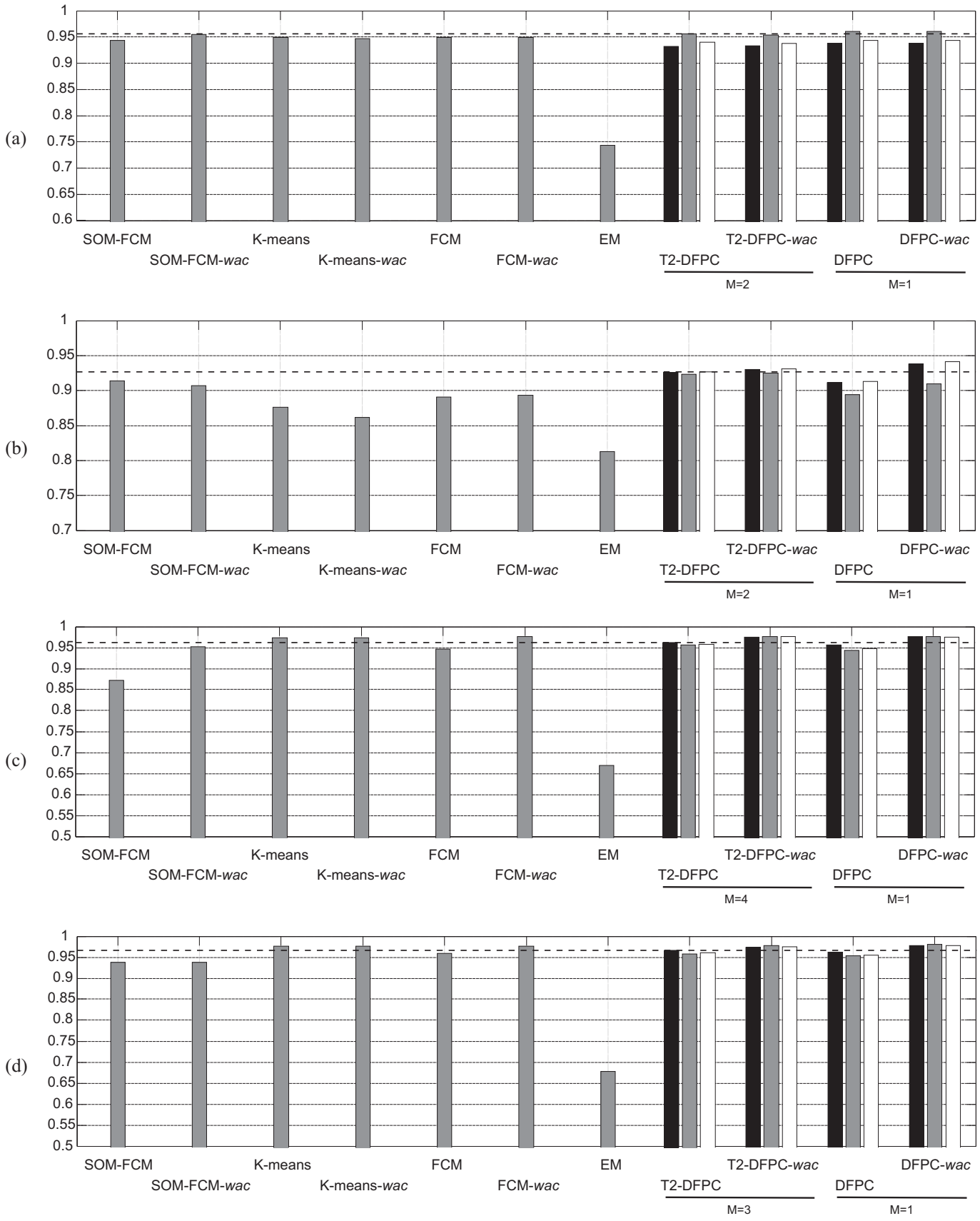

Fig. 6. Clustering accuracies obtained for the test datasets. Bars indicate different logic operators used (black: Max-Min; gray: GMCL; white: AMCL). The partition size used (M) is indicated for each case. The horizontal dotted line indicates the best accuracy value obtained for the method proposed (T2-DFPC). (a) Wine. (b) Iris. (c) MRI1. (d) MRI2. (e) Wisconsin breast cancer. (f) Pima Indians diabetes. (g) Moon. (h) Seeds. (i) Banknote.

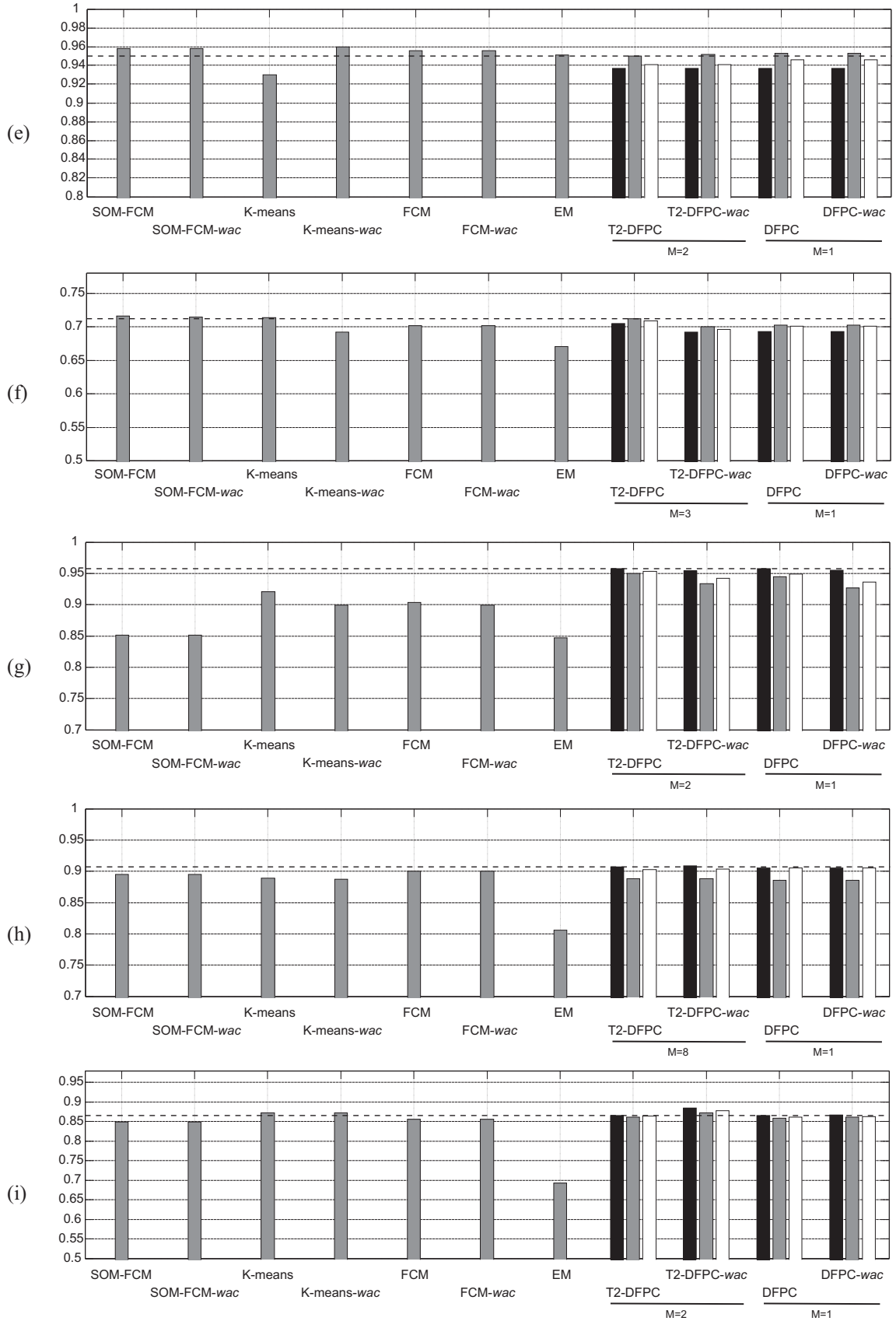

Fig. 6. Continued

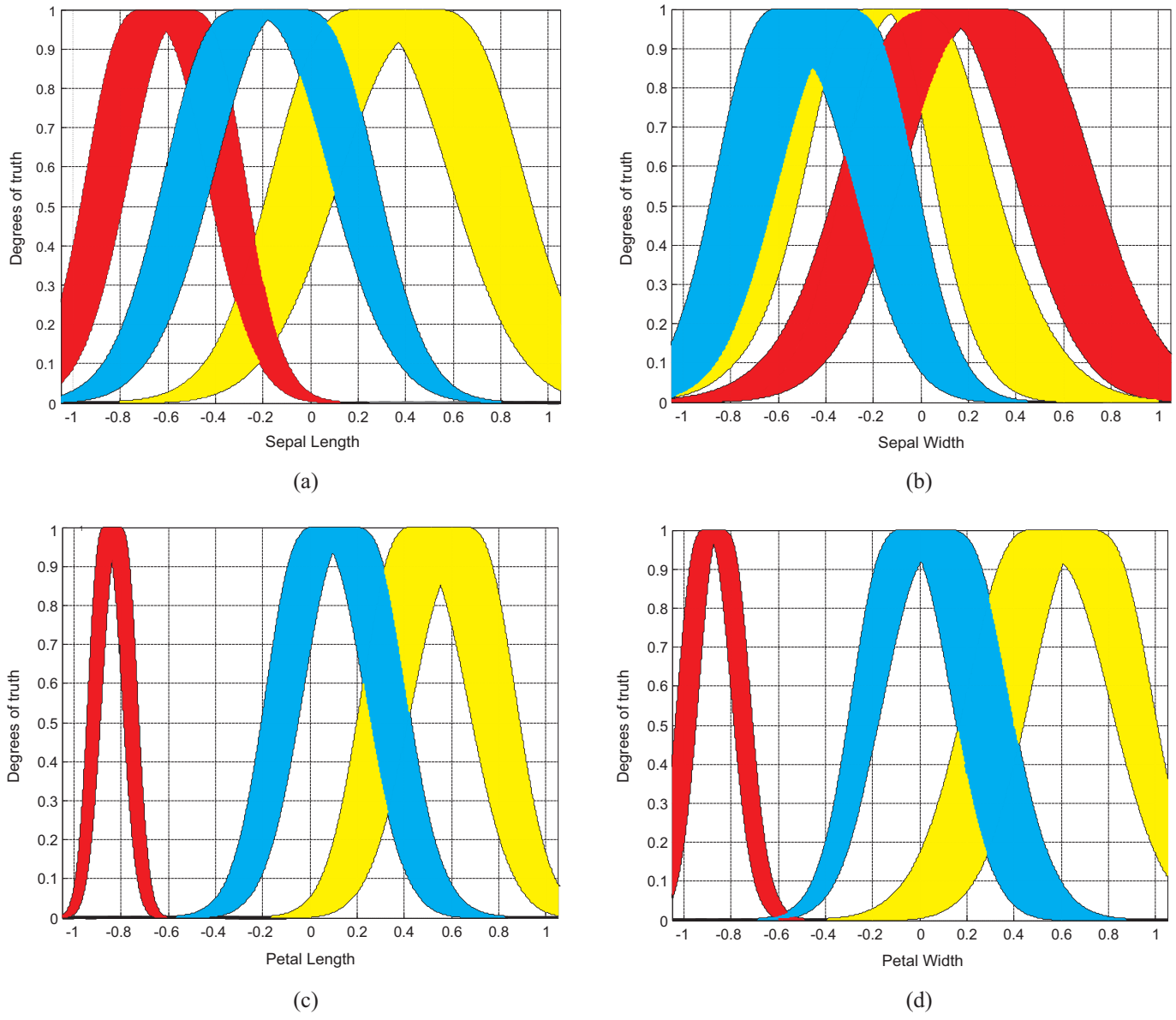

Fig. 7. Interval type-2 membership functions obtained for the Iris dataset. Different colors indicate different clusters: Iris Setosa (red), Iris Versicolor (blue), and Iris Virginica (yellow). (a) Interval type-2 membership functions for the feature *"Sepal Length"*. (b) Interval type-2 membership functions for the feature *"Sepal Width"*. (c) Interval type-2 membership functions for the feature *"Petal Length"*. (d) Interval type-2 membership functions for the feature *"Petal Width"*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Setosa" is equivalent to *"The Sepal Length in the datum d is small and the Sepal Width in the datum d is slightly large and the Petal Length in the datum d is very small and the Petal Width in the datum d is very small"*. Consequently, it is easy to see how the cluster could be described using linguistic expressions extracted by analyzing the membership functions and the fuzzy predicates obtained from the T2-DFPC method. Following the same procedure, the rest of the clusters can be described as follows:

- $p_{Versicolor}(d)$: "The datum d belongs to Iris Versicolor" is equivalent to "The Sepal Length in the datum d is medium and the Sepal Width in the datum d is small and the Petal Length in the datum d is medium and the Petal Width in the datum d is medium".
- $p_{Virginica}(d)$: "The datum d belongs to Iris Virginica" is equivalent to "The Sepal Length in the datum d is large and the Sepal Width in the datum d is medium and the Petal Length in the datum d is large and the Petal Width in the datum d is large".

As it was previously mentioned, a major advantage of using the T2-DFPC approach in comparison with classical clustering algorithm is that it allows to evaluate the vagueness or variability of the fuzzy concept obtained through clustering by analyzing the FOU of the membership functions generated.

Following the observations included in the last part of the stage #4 method explanation and taking into account the membership functions obtained for the Iris dataset (Fig. 7), the following analysis can be made.

For the cluster Iris Setosa, shown in red in the Fig. 7, the membership functions related to the features *"Sepal Length"* and *"Sepal Width"* have relative large FOU, meaning large vagueness for these features around the cluster, i.e. the cluster in these feature is not compact. On the contrary, there is a small vagueness for the *"Petal Length"* and *"Petal Width"* in the cluster. As a result, it is possible

to say that the first two features have lower importance than the last two to assign data to Iris Setosa.

In the case of Iris Versicolor, shown in blue in the Fig. 7, the obtained membership functions were relatively thin except for the feature *"Sepal Width"*. Therefore, this cluster can be associated with low vagueness except for the feature *"Sepal Width"*.

The cluster Iris Virginica, shown in yellow in the Fig. 7, were associated to relatively large FOU for all the features, indicating a high variability in the cluster prototypes obtained and high vagueness around the attributes described by the membership functions.

In addition, for the *"Sepal Width"* (Fig. 7b) the membership functions obtained for the different cluster are highly overlapped, which indicates a high fuzziness of the attributes discovered.

It is possible to extract further information about the correlation between features using the predicates and the membership functions. For example, let us consider the attributes and the membership functions related to the features *"Petal Length"* and *"Petal Width"* (Fig. 7c and d). For each cluster, the feature attributes considered are the same as well as the shape and the position of the membership functions are very similar. In consequence, these observations allow concluding that the features *"Petal Length"* and *"Petal Width"* are strongly correlated in this dataset.

5. Discussion

Analyzing the T2-SFPC method, in stages #1 and #2 (described in Section 3.1), the clustering scheme FCM-BIC discovered knowledge related to clustering. The number of cluster is automatically defined, extracting cluster prototypes. Based on the clustering performances obtained with and without automatic clustering, comparing the results of T2-SFPC with those of T2-SFPC-*wac*, BIC combining with classical clustering techniques like FCM is a good choice for determining the proper number of clusters. Variants of the test clustering algorithms with BIC showed good accuracy as well, enabling to perform clustering without need knowing the number of clusters.

Regarding the clustering assessment results, the obtained accuracy indicates that performance achieved using the method proposed is similar to the results obtained by the test methods. In some cases, the proposed approach outperformed the results of the other methods. In others cases, the test methods showed better performance than the T2-DFPC method. But, a major advantage of the method proposed is that the clustering is interpretable, enabling the knowledge discovery. Besides, in the 100% of the datasets tested the performance of the approach proposed is highly acceptable, which evidences that it constitutes a reliable method.

In relation to the differences observed between the results of models based on fuzzy predicates (T2-DFPC and DFPC) against to traditional clustering, inherent theoretical approaches and computing procedures are different. Fuzzy-predicate-based approaches compute the belonging to a cluster weighting the fulfillment of attributes related to the cluster. This belonging descends when data move away of the maximum of the membership functions, which do not necessarily correspond to the cluster centroid as they are understood in the classical clustering theory. This occurs in the methods T2-DFPC and DFPC. In addition, in the fuzzy predicates partial belonging described by membership functions are aggregated using fuzzy operators. On the contrary, traditional clustering approach are based on distances in the data space.

In general, accuracy depends on how good membership functions capture cluster properties, which in the method proposed depends on quality of the cluster prototypes obtained in the stage #2. It is expected that resulting cluster prototypes describe the properties of clusters depending on the number of data in the dataset, the quality of the features and how the shape of the original clusters matches the theoretical hypothesis of the FCM-BIC approach, i.e. if the clusters correspond to Gaussian-spherical-distributions. Tests carried out during the method development showed a high dependence of the number of data.

Observing accuracy obtained for the test methods based on type-1 FL, results showed similar performance against the methods based on interval type-2 FL, indicating small differences. The essential characteristic of our proposal is combining the results coming from the different subsets of a given dataset as they were different "opinions" of experts. This unique characteristic can give to the proposed method advantages respect to methods based on type-1 FL, especially considering vagueness in data, for instance, caused by noise. Tests based on data affected by noise are currently being run.

Performance obtained comparing intervals by means of the method proposed suggests that the method based on the measure of intervals of truth values can be used as a general methodology to apply interval type-2 FL in models based on fuzzy predicates, taking the advantages of this kind of FL.

In addition, the computational cost involved in computing interval type-2 fuzzy predicates and interval comparing using the proposed methodology is comparable to the required for the type-1 fuzzy predicates scheme. Therefore, that proposal for comparing interval could be applied in pattern recognition problems replacing the models based on FIS, especially considering the computational cost involved in the defuzzification operations.

The only parameter to be considered for the application of the T2-DFPC is the size of the initial random partition (M), except in cases where the dataset is already physically partitioned. Such parameter should be set according to the number of data to be clustered, taking into account that each subset generated should be statistically representative of the initial dataset X. As an additional characteristic, the initial data partition can reduce the computational cost of the clustering stages used for the extraction of clusters prototypes using parallel computing.

In this regard, the T2-DFPC method could be implemented in distributed clustering approach, where in general a large number of data are processed. In such application, M nodes act collecting data, defining M data subsets and replacing the initial random partition of the stage #1 of the method. The stages #2 and #3 of the T2-DFPC method are applied to the subset in each of the M nodes, generating clusters prototypes. Then, each node can compute mean values and standard deviations for its prototypes defining type-1 membership functions following the steps of the stage #4 of the T2-DFPC. Finally, parameters of the Gaussian type-1 membership function are shared for the different nodes, which can generate interval type-2 membership functions and fuzzy predicates. As a result, it is possible to generate a fuzzy system enabling the clustering of the data in each node, where the node just has to evaluate the fuzzy predicates. Main advantage of such an approach is that the fuzzy clustering system includes information of all the nodes.

For all said, the T2-DFPC method can be applied to most data clustering applications. In the next, a brief analysis is given, comparing existing clustering schemes and their application against our method.

Some interesting proposal aims to identify particular cases in data as a consequence of a fuzzy clustering and a posterior supervised classification system (Singh, Verma, & Thoke, 2016). In our approach, outliers are non-significant because they would have low influence in membership functions determination.

Our approach could be suitable to be applied in business scenarios where big data are considered, probably coming from different data sources, and patterns of behavior are trying to be discovered (Cerquitelli, Servetti, & Masala, 2016; Ordoñez et al., 2017). Given the raising quantity of data, methods need to be more robust to noise and they are required to be able to merge results coming

from subsets of the data. Both needs are covered in the proposed method as well because interval type-2 membership functions are generated combining cluster prototypes coming from different data subsets.

Medical problems could find in the T2-DFPC approach a preliminary knowledge from a priori cases, giving a decision support system based on data and maybe adding some expert optimization. As an example, patients with successful and unsuccessful surgeries could be identified in different clusters and additional information about the different cases could be discovered (Swenson, Bastian, & Nembhard, 2016). Other interesting application consist on apply a sophisticated ensemble-based clustering algorithm in order to discover patterns in cancer data (Qi et al., 2013). In this case, further information about the outcome is mandatory to understand the problem.

This paper gives a contribution for the continuous developing subject of data comprehension. Similar works has been published for classification and clustering tasks, aiming to explain the results obtained (Piltaver, Luštrek, Gams, & Martinčić-Ipšić, 2016).

Some previous works suggested fuzzy predicates as a technique applied to magnetic resonance images, considering different image types acquired at the same time (Meschino et al., 2015). This approach could be highly enriched using interval type-2 membership functions to integrate information coming from, for example, different imaging centers in a unique fuzzy system. Once the clustering system learned from data, it is able to suggest a cluster for new data. If clusters were previously analyzed and identified, new data will have a characterization by similarity.

Summarizing, the proposed methodology which includes the T2-DFPC method for the generation of interval type-2 membership functions and fuzzy predicates and the measure of interval of truth values for interval comparing has the next features:

- As a result of the T2-DFPC, each cluster is explained by only one fuzzy predicate as $p_k(d)$: *"The datum d belongs to cluster k,"* whose degree of truth will be computed by the logic combination of simple predicates and interval type-2 membership functions generated automatically from the data.
- The cluster assignment is performed by evaluating all fuzzy predicates for each datum and determining which predicate has the highest measure of interval of truth values, as it is detailed in the Section 3.2.
- The method captures the knowledge contained in data and stores it in interval type-2 membership function and fuzzy predicates.
- The discovered knowledge can be interpreted and also can be modified by experts. Each cluster can be explained by only one predicate in a simpler way than in others algorithms proposed previously (Meschino et al., 2015). An expert user is able to give linguistic meaning to the membership functions and the predicates that were automatically discovered. In addition, information about vagueness and variability inside a cluster as well as fuzziness of the attributes modelled by the membership function can be obtained.
- The generated predicates are able to generalize the knowledge for new cases (new data) by assigning them to clusters, as an intelligent system. In this sense, the clustering approach could be defined as a prototypes-based clustering.
- In order to generate the interval type-2 membership function a considerable amount of data cases are needed, considering that a partition is required and the number of data in each subset must be representative. In consequence, the method could not be correctly applied on clustering problems with small number of data.

6. Conclusion

In this paper it is proposed a new data clustering method called Type-2 Data-based Fuzzy Predicate Clustering (T2-DFPC) where fuzzy predicates automatically-discovered from data are used to perform the clustering. Interval type-2 FL is used to quantify the degree of truth of the fuzzy predicates modelling the variability in the knowledge extracted from data. The proposal includes a methodology for interval comparing, enabling the use of interval type-2 FL in data clustering based on predicates.

The fuzzy predicates act as cluster descriptors which consider how the data behave into each cluster and enabling to discover knowledge about it.

As the interval type-2 membership functions merge all data variability into each cluster in a unique membership function, only one compound fuzzy predicate is defined for each of the discovered clusters. As a result, the knowledge interpretation is easier compared to other existing interpretable clustering algorithms.

The proposed method is a fast, useful, general, and unsupervised approach for interpretable data clustering. The knowledge extracting capabilities is one of the more important contributions. Linguistic expressions obtained from predicates can be easily adapted to match the terminology of the field the data relate to.

The T2-DFPC method was applied to different datasets and results obtained were compared against classical clustering algorithms and a clustering method similar to the proposed one that uses type-1 FL. The T2-DFPC method was consistently one of the best in terms of accuracy, considering experiments on very different datasets. Moreover, the knowledge extraction was very easy, which was shown on the Iris dataset.

These results show the T2-DFPC is a promising clustering paradigm. Also, the minor differences between the accuracy of algorithms with and without using BIC suggest that combining BIC with a clustering algorithm is a good approach to automatically determine the number of clusters.

As immediate future work, we plan to analyze how to combine the advantages of the interval type-2 FL in data clustering and the SOM as knowledge extracting tool as was used in the approach based on type-1 FL reported in (Meschino et al., 2015), called SOM-based Fuzzy Predicate Clustering (SFPC). It is expected that using this approach, cluster prototypes can be improved due to the SOM abilities to generalize the data-space, obtaining more suitable descriptions of the clusters and higher clustering accuracies than those obtained with the T2-DFPC method. This approach is currently being developed.

We are currently also performing extra experiments on noisy datasets to evaluate the robustness of the method proposed.

Acknowledgements

The authors would like to acknowledge and thank the support for researching from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) from Argentina, and the European Commission under the Erasmus Mundus Action 2 Programme – Eureka SD Project.

References

Bache, K., & Lichman, M. (2013). *UCI machine learning repository* Retrieved from *http://archive.ics.uci.edu/ml*.

Bishop, C. M. (2006). *Pattern recognition and machine Learning. Information science and statistics*. Springer.

Bouchet, A., Pastore, J. I., Andrade, R. E., Brun, M., & Ballarin, V. (2011). Arithmetic mean based compensatory fuzzy logic. *International Journal of Computational Intelligence and Application, 10*(2), 231–243.

Celikyilmaz, A., & Turksen, I. B. (2008). Enhanced fuzzy system models with improved fuzzy clustering algorithm. *IEEE Transactions on Fuzzy Systems, 16*(3), 779–794.

Cerquitelli, T., Servetti, A., & Masala, E. (2016). Discovering users with similar internet access performance through cluster analysis. *Expert Systems with Applications, 64*, 536–548. doi:10.1016/j.eswa.2016.08.025.

Charytanowicz, M., Niewczas, J., Kulczycki, P., Kowalski, P. A., Lukasik, S., & Zak, S. (2010). A complete gradient clustering algorithm for features analysis of x-ray images. In E. Piętka, & J. Kawa (Eds.). In *Information technologies in biomedicine: Vol. 2* (pp. 15–24). Berlin Heidelberg: Springer.

Comas, D. S., Meschino, G. J., Brun, M., & Ballarin, V. L. (2014a). Label-based type-2 fuzzy predicate classification applied to the design of morphological W-operators for image processing. In *First latin American congress on computational intelligence* (pp. 55–60).

Comas, D. S., Meschino, G. J., Pastore, J. I., & Ballarin, V. L. (2011). A survey of medical images and signal processing problems solved successfully by the application of Type-2 Fuzzy Logic. *Journal of Physics Conference Series, 332*(1). doi:10.1088/1742-6596/332/1/012030.

Comas, D. S., Pastore, J. I., Bouchet, A., Ballarin, V. L., & Meschino, G. J. (2014b). Type-2 fuzzy logic in decision support systems. *Studies in Computational Intelligence, 537*. doi:10.1007/978-3-642-53737-0_18.

Comas, D. S., Pastore, J. I., Bouchet, A., Ballarin, V. L., Passoni, L. I., & Meschino, G. J. (2012). Diseño automático de Sistemas de Lógica Difusa Tipo 2 mediante Mapas Auto-organizados. In *Quinto torneo regional de inteligencia computacional* (pp. 1–8).

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems, 47*(4), 547–553.

Deng, Z., Jiang, Y., Chung, F.-L., Ishibuchi, H., & Wang, S. (2013). Knowledge-leverage-based fuzzy system and its modeling. *IEEE Transactions on Fuzzy Systems, 21*(4), 597–609.

Drobics, M., Bodenhofer, U., & Winiwarter, W. (2002). Mining clusters and corresponding interpretable descriptions - a three-stage approach. *Expert System, 19*, 224–234.

Dubey, R. B., Hanmandlu, M., Gupta, S. K., & Gupta, S. K. (2010). The brain MR Image segmentation techniques and use of diagnostic packages. *Academic Radiology, 17*(5), 658–671.

Dubois, H., & Prade, D. (1980). *Fuzzy sets and Systems: theory and applications*: Vol. 1. New York: Academic Press Inc.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Human Genetics, 7*, 179–188.

Fraley, C., & Raftery, A. E. (1998). How many clusters? which clustering method? Answers via model-based cluster analysis. *Computer Journal, 41*(8), 578–588.

Ghazavi, S. N., & Liao, T. W. (2008). Medical data mining by fuzzy modeling with selected features. *Artificial Intelligence in Medicine, 43*(3), 195–206. doi:10.1016/j.artmed.2008.04.004.

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems, 17*(2), 107–145.

Hsu, C.-C., Lin, S.-H., & Tai, W.-S. (2011). Apply extended self-organizing map to cluster and classify mixed-type data. *Neurocomputing, 74*(18), 3832–3842. doi:10.1016/j.neucom.2011.07.014.

Ishibuchi, H., & Tanaka, H. (1990). Multiobjective programming in optimization of the interval objective function. *European Journal of Operational Research, 48*(2), 219–225.

Jain, A. K., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys, 31*(3), 264–323.

John, R. I., & Innocent, P. R. (1998). Type 2 fuzzy sets and neuro-fuzzy clustering of radiographic tibia images. In *Proceedings of the Sixth IEEE Int. Conf. on Comput. Intell.* ((IEEE, Ed.)).

Juang, C.-F., Chiu, S.-H., & Chang, S.-W. (2007). A self-organizing ts-type fuzzy network with support vector learning and its application to classification problems. *IEEE Transactions on Fuzzy Systems, 15*(5), 998–1008.

Kundu, S. (1997). Min-transitivity of fuzzy leftness relationship and its application to decision making. *Fuzzy Sets System, 86*(3), 357–367.

Kwan, R. K. R. K.-S., Evans, A. C., & Pike, G. B. (1999). MRI simulation-based evaluation of image-processing and classification methods. *IEEE Transactions on Medical Imaging, 18*(11), 1085–1097.

Lee, C. C. (1990). Fuzzy logic in control systems fuzzy logic controler Part I. *IEEE Transaction on Systems Man and Cibernetics, 20*(2), 404–418 http://ieeexplore.ieee.org/document/52551/.

Mansoori, E. G. (2011). FRBC: A fuzzy rule-based clustering algorithm. *IEEE Transactions on Fuzzy Systems, 19*(5), 960–971.

Melin, P., & Castillo, O. (2013). A review on the applications of type-2 fuzzy logic in classification and pattern recognition. *Expert Systems with Applications, 40*(13), 5413–5423.

Mendel, J. M. (2007). Type-2 fuzzy sets and systems: An overview. *IEEE Computational Intelligence Magazine, 2*(1), 20–29.

Meschino, G. J., Comas, D. S., Ballarin, V. L., Scandurra, A. G., & Passoni, L. I. (2013). Using SOM as a tool for automated design of clustering systems based on fuzzy predicates. *Advances in Intelligent Systems and Computing, 198* AISC. doi:10.1007/978-3-642-35230-0_9.

Meschino, G. J., Comas, D. S., Ballarin, V. L., Scandurra, A. G., & Passoni, L. I. (2015). Automatic design of interpretable fuzzy predicate systems for clustering using self-organizing maps. *Neurocomputing, 147*(1). doi:10.1016/j.neucom.2014.02.059.

Moore, R., & Lodwick, W. (2003). Interval analysis and fuzzy set theory. *Fuzzy Sets System, 135*(1), 5–9.

Ordoñez, A., Ordoñez, H., Corrales, J. C., Cobos, C., Wives, L. K., & Thom, L. H. (2017). Grouping of business processes models based on an incremental clustering algorithm using fuzzy similarity and multimodal search. *Expert Systems with Applications, 67*, 163–177. doi:10.1016/j.eswa.2016.08.061.

Piltaver, R., Luštrek, M., Gams, M., & Martinčić-Ipšić, S. (2016). What makes classification trees comprehensible? *Expert Systems with Applications, 62*, 333–346. doi:10.1016/j.eswa.2016.06.009.

Qi, R., Wu, D., Sheng, L., Henson, D., Schwartz, A., Xu, E., et al. (2013). On an ensemble algorithm for clustering cancer patient data. *BMC Systems Biology, 7*(Suppl 4) (Suppl 4), S9. doi:10.1186/1752-0509-7-S4-S9.

Ruspini, E. H. (1969). A new approach to clustering. *Information and Control, 15*, 22–32.

Sengupta, A., & Pal, T. K. (2000). On comparing interval numbers. *European Journal of Operational Research, 127*(1), 28–43.

Singh, B. K., Verma, K., & Thoke, A. S. (2016). Fuzzy cluster based neural network classifier for classifying breast tumors in ultrasound images. *Expert Systems with Applications, 66*, 114–123. doi:10.1016/j.eswa.2016.09.006.

Swenson, E. R., Bastian, N. D., & Nembhard, H. B. (2016). Data analytics in health promotion: Health market segmentation and classification of total joint replacement surgery patients. *Expert Systems with Applications, 60*, 118–129. doi:10.1016/j.eswa.2016.05.006.

Tizhoosh, H. R. (2005). Image thresholding using type II fuzzy sets. *Pattern Recognition, 38*(12), 2363–2372.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*, 338–353.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences, 8*, 199–249.

Zadeh, L. A. (1989). Knowledge representation in fuzzy logic. *IEEE Transactions on Knowledge and Data Engineering, 1*(1), 89–100.

Zarandi, M. H., Zarinbal, M., & Izadi, M. (2011). Systematic image processing for diagnosing brain tumors: A Type-II fuzzy expert system approach. *Applied Soft Computing, 11*(1), 285–294.

Zeng, J., & Liu, Z.-Q. (2006). Type-2 fuzzy hidden Markov models and their application to speech recognition. *IEEE Transactions on Fuzzy Systems, 14*(3), 454–467.

Zhao, Q., Hautamaki, V., & Fränti, P. (2008). Knee point detection in BIC for detecting the number of clusters. In J. Blanc-Talon, S. Bourennane, W. Philips, D. Popescu, & P. Scheunders (Eds.), *Advanced concepts for intelligent vision systems*. Berlin Heidelberg: Springer.