Archives of Control Sciences
Volume 12(XLVIII), 2002
No. 4, *pages 397–410*

Generation of interpretable fuzzy granules by a double-clustering technique

GIOVANNA CASTELLANO, ANNA MARIA FANELLI and CORRADO MENCAR

This paper proposes an approach to derive fuzzy granules from numerical data. Granules are first formed by means of a double-clustering technique, and then properly fuzzyfied so as to obtain interpretable granules, in the sense that they can be described by linguistic labels. The double-clustering technique involves two steps. First, information granules are induced in the space of numerical data via the FCM algorithm. In the second step, the prototypes obtained in the first step are further clustered along each dimension via a hierarchical clustering, in order to obtain one-dimensional granules that are afterwards quantified as fuzzy sets. The derived fuzzy sets can be used as building blocks of a fuzzy rule-based model. The approach is illustrated with the aid of a benchmark classification example that provides insight into the interpretability of the induced granules and their effect on the results of classification.

Key words: information granulation, fuzzy clustering, hierarchical clustering, fuzzy rule-based model

1. Introduction

Information granulation [14], [15] is the process of forming meaningful pieces of information, called information granules, regarded as entities embracing collections of individual elements (e.g numerical data) that exhibit some functional or descriptive commonalities. Once formed, information granules help understand data by capturing the essence of the relationships within data.

Granular computing [6], [9], [12] which is oriented towards representing and processing information granules, is a computing paradigm that embraces a number of modeling frameworks. Fuzzy granulation [7], [13] is one such framework, whose objective is to build models at a certain level of information granularity that is conveniently quantified in terms of fuzzy sets used as basic reference points for fuzzy information processing.

Several information granulation approaches have been proposed [1], [7], [12] but only few of them addresses the issue of building meaningful linguistic labels (granules) in the space of experimental data [11].

The Authors are with Computer Science Department, University of Bari Bari, Italy, e-mail: [castellano, fanelli, mencar]@di.uniba.it

Received 15.11.2001, revised 25.04.2002.

Hence, one main issue inherently associated with information granulation is how to describe the derived granules in an interpretable form. As a consequence, one desirable feature for any information granulation approach is the ability to determine a set of granules that can be easily interpreted. In this scenario, fuzzy sets emerge as conceptual entities with a well-defined semantics, hence they offer useful features supporting the formation of granules that are descriptors of the data.

Along with these ideas, this work focuses on fuzzy granulation viewed as the composition of granulation and fuzzification, where the process of granulation is supported by a clustering mechanism. Specifically, a fuzzy double-clustering technique is proposed, which allows "interpretable" granulation from some finite data. Proceeding with clouds of numerical data, a series of clusters - information granules - are formed and afterwards quantified in terms of fuzzy sets that are semantically sound.

Finally, as a facultative step, the derived fuzzy sets can be used as building blocks of a fuzzy rule-based model which can be used to verify how much the fuzzy granules identified from data are useful in providing good mapping properties.

The paper is organized as follows. In Section 2 we introduce the proposed framework for information granulation. Section 3 deals with the double-clustering technique proposed to extract granules from data. Section 4 and 5 describe the fuzzification of the induced granules and their use in the construction of a fuzzy rule-based model. Finally, a numerical example is given in Section 6 and conclusions are covered in Section 6.

2. The information granulation approach

The development of information granules can be accomplished in two main ways: prescriptive design, where meaningful granules are expressed by an observer, and descriptive design, where numeric data are embraced to form information granules [10], [11]. The latter approach overcomes the main drawback of the prescriptive approach, that is the formation of granules that may not resemble the relationships among data.

This work focuses on the descriptive case, i.e. we concentrate on building meaningful information granules in the space of experimental data and describing them in terms of fuzzy sets. Such fuzzy sets can be used to form a rule-based model that captures relationships between such granules. Summarizing, the proposed information granulation approach for the design of interpretable fuzzy granules involves the following stages:

- **Granulation of numerical data:** To extract granules from data, a double-clustering technique is developed, that involves two consecutive steps:
  - *Data clustering:* clustering is performed in the multi-dimensional space of data to embrace similar points into granules, with a pre-fixed level of granulation (number of clusters).

– *Prototype clustering:* the prototypes obtained in the first step are further clustered along each dimension of the input space, so as to obtain a number of one-dimensional prototypes on each dimension.

- **Fuzzification of granules:** the granules induced on each input dimension are quantified in terms of interpretable fuzzy sets that describe the input variables.

A schema of the whole information granulation process is portrayed in Figure 1.

When the granulation process is completed, a fuzzy rule-based model which relies on the derived fuzzy granules can be built. This is aimed to verify how much the fuzzy granules identified from data are useful in providing good mapping properties or classification capabilities.

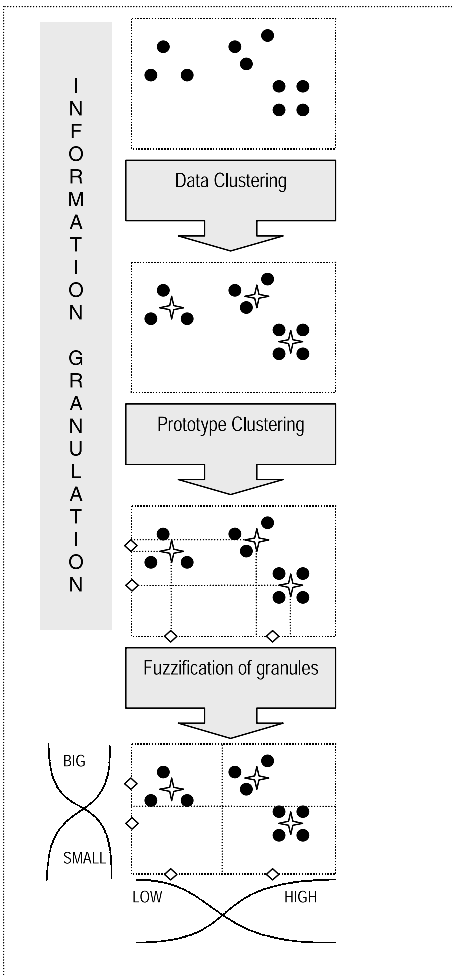


Figure 1. Diagram of the information granulation process

3. Double-clustering technique

To extract granules from data, we have developed a double-clustering technique that performs clustering at two levels. Given a set of numerical data in a n-dimensional space $(X \subseteq \boldsymbol{R}^n)$, the first level involves clustering in the n-dimensional space of data, yielding a number of n-dimensional granules embracing collections of similar data. At this level, each granule is a cluster of points represented by a prototypical point. At the second level, the cluster prototypes are projected along each dimension of the input space, so that a set of one-dimensional prototypes is established on each dimension of the input space. Then, a further clustering is performed on each dimension to assembly one-dimensional prototypes into final granules.

The proposed double-clustering technique tries to exploit the features of both multi-dimensional clustering and one-dimensional clustering. Precisely, the multidimensional clustering can capture the granularity of the data in the multidimensional space, but the fuzzification of the resulting granules typically results in nonsensical fuzzy sets. Conversely, the one-dimensional clustering can provide interpretable fuzzy sets but may loose the information about the granularity of the multidimensional data.

The complete double-clustering algorithm is summarized below.

Double-clustering algorithm

1. Perform a clustering on available dataset. The result of this operation is a set of p cluster prototypes $\bar{c}_1, \bar{c}_2, \ldots, \bar{c}_p \subseteq \boldsymbol{R}^n$, where $\bar{c}_i = (c_{i1}, c_{i2}, \ldots, c_{in})$.
2. For each dimension $j = 1, 2, \ldots, n$, perform a clustering on the set $C_j := \{c_{ij} \mid i = 1, 2, \ldots, p\}$. The result of this operation is a pre-fixed number of K_j clusters prototypes for each dimension.

In the following, we describe in more detail each step of the double-clustering technique.

3.1. Data clustering

The first step of the proposed approach is concerned with finding granules in the form of clusters, i.e. entities embracing collections of n-dimensional numerical data that exhibit some functional or descriptive commonalties. To find proper clusters in the multidimensional input space, we consider the FCM algorithm [2], that is one of the most common fuzzy clustering algorithm, widely used in many applications. However, other clustering algorithms, such as conditional FCM [8], can be used at this step without altering the nature of the proposed approach.

The result of the first clustering step is a set of p clusters in the multidimensional space, represented by p prototypes $\bar{c}_1, \bar{c}_2, \ldots, \bar{c}_p \subseteq \boldsymbol{R}^n$. Each prototype is a n-dimensional vector $\bar{c}_i = (c_{i1}, c_{i2}, \ldots, c_{in})$ that provides n one-dimensional prototypes, one for each input dimension. These one-dimensional prototypes become the points to be clustered in the second step of the double-clustering technique, as explained below.

3.2. Prototype clustering

Once information granulation has been completed in the multi-dimensional space, a second clustering is performed separately on each dimension of the input space, with the aim to group along each input dimension the prototypes obtained in the first clustering step.

Precisely, let c_{ij} be the j-th component of the i-th cluster prototype. For each dimension $j = 1, 2, \ldots, n$, we perform a clustering on the set $C_j := \{c_{ij} \mid i = 1, 2, \ldots, p\}$. Because of the low cardinality of each set C_j and the one-dimensional nature of the points to be clustered, a very simple clustering technique can be applied. In this work we apply a hierarchical clustering algorithm. Hierarchical clustering [5] is a recursive form of clustering. Methods of hierarchical clustering can be agglomerative or divisive [4]. Here, we focus on the agglomerative approach, that starts with each data point in a separate cluster and then combines the clusters sequentially, reducing the number of clusters at each step by merging the two most similar clusters. These steps can be repeated until the desired number of cluster is obtained or the distance between two closest clusters is above a certain threshold.

There are many different variations of agglomerative hierarchical algorithms [4], which primarily differ in how they evaluate the similarity between two clusters. In this work, we consider a simple method, where each cluster is represented by a centroid of the points contained in the cluster, called prototype, and the similarity between two clusters is measured by the closeness between prototypes of the clusters.

Indicating by K_j $(j = 1, 2, \ldots, n)$ the final number of clusters to be derived from each set C_j $(j = 1, 2, \ldots, n)$, the adopted hierarchical clustering algorithm works as follows:

1. Assume, without loss of generality, that the elements of C_j are sorted, that is, $i_1 < i_2 \to c_{i_1 j} \leqslant c_{i_2 j}$. Firstly, each element of C_j is considered as an elementary cluster, hence the prototypes of the clusters initially coincide with the elements of C_j. The initial set of prototypes, is defined as:

$$P^{(0)} := \left\{p_1^{(0)}, p_2^{(0)}, \ldots, p_p^{(0)}\right\} := \{c_{1j}, c_{2j}, \ldots, c_{pj}\} \quad (1)$$

   The corresponding initial set of clusters is defined as:

$$H^{(0)} := \left\{h_1^{(0)}, h_2^{(0)}, \ldots, h_p^{(0)}\right\} := \{\{c_{1j}\}, \{c_{2j}\}, \ldots, \{c_{pj}\}\} \quad (2)$$

2. For $k = 1, 2, \ldots, p - K_j$

   - Find the two nearest consecutive prototypes in $P^{(k-1)}$, denoted by $p_{i^\star}^{(k-1)}$ and $p_{i^\star+1}^{(k-1)}$.

- Define the new set of prototypes $P^{(k)} := \left\{p_1^{(k)}, p_2^{(k)}, \ldots, p_{p-k}^{(k)}\right\}$, where:

$$p_i^{(k)} := \begin{cases} p_i^{(k-1)}, & i < i^\star \\ \left(p_{i^\star}^{(k-1)} + p_{i^\star+1}^{(k-1)}\right)/2, & i = i^\star \\ p_{i+1}^{(k-1)}, & i > i^\star \end{cases} \tag{3}$$

- Define the new set of clusters $H^{(k)} := \left\{h_1^{(k)}, h_2^{(k)}, \ldots, h_{p-k}^{(k)}\right\}$, where:

$$h_i^{(k)} := \begin{cases} h_i^{(k-1)}, & i < i^\star \\ h_{i^\star}^{(k-1)} \cup h_{i^\star+1}^{(k-1)}, & i = i^\star \\ h_{i+1}^{(k-1)}, & i > i^\star \end{cases} \tag{4}$$

Once the hierarchical clustering is completed on each set C_j $(j = 1, \ldots, n)$, K_j prototypes with corresponding clusters have been derived on each input dimension. We indicate by $P := P^{(p-K_j)} =: \left\{p_1, p_2, \ldots, p_{K_j}\right\}$ and $H := H^{(p-K_j)} =: \left\{h_1, h_2, \ldots, h_{K_j}\right\}$ the set of such prototypes, with associated clusters, respectively.

The complexity of the hierarchical clustering algorithm is $O(np^2)$ because, for each of the n dimensions, the number of distance calculations is $O(p)$, repeated for $O(p)$ steps. Since usually $p \ll N$, that is the number p of prototypes is usually smaller than the number N of patterns, then the complexity of the clustering in the second step is negligible in comparison to the complexity of the clustering in the first step, which is $O(Nnp^2)$ when using FCM.

Hence, the overall complexity of the double-clustering technique is given approximately by the complexity of the first clustering step.

4. Fuzzification of information granules

The last stage of the proposed information granulation approach involves the fuzzification of the information granules defined on each dimension, i.e. the quantification of such granules in terms of interpretable fuzzy sets.

Precisely, for each dimension $j = 1, 2, \ldots, n$, the K_j extracted clusters are transformed into as many (interpretable) fuzzy sets. Different types of membership functions can be used to describe the fuzzy sets. Here we consider Gaussian membership functions. Hence, a fuzzy set A_k^j is associated with the membership function

$$\mu_{A_k^j}(x) := \exp\left(-((x - \omega_k^j)/\sigma_k^j)^2\right) \tag{5}$$

Therefore, to complete the mathematical definition of fuzzy sets, we have to derive the center ω_k^j and the width σ_k^j of each Gaussian function. To do so, we suppose the input space X to be a hyper-interval defined as $X := \times_{j=1}^{n}[m_j, M_j]$ and define the set $T_j := \left\{t_0^j, t_1^j, \ldots, t_{K_j}^j\right\}$, where

$$t_k^j := \begin{cases} 2m_j - p_1, & k = 0 \\ (p_k + p_{k+1})/2, & 0 < k < K_j \\ 2M_j - p_{K_j}, & k = K_j \end{cases} \tag{6}$$

Then, the centres ω_k^j and the widths σ_k^j of membership functions are obtained, respectively, as:

$$\omega_k^j := \left(t_{k-1}^j + t_k^j\right)/2 \tag{7}$$

$$\sigma_k^j := \left(t_k^j - t_{k-1}^j\right)/2\sqrt{-2\ln\varepsilon} \tag{8}$$

where ε is the maximum overlap between two adjacent fuzzy sets.

The resulting fuzzy sets are very interpretable, in the sense that a semantic meaning can be easily associated to each of them. Precisely, depending on the value of index k, a meaningful symbolic label can be given to A_k^j. For example, assuming $K_j := 3$, we can assign the following linguistic labels: "SMALL" for $k = 1$, "MEDIUM" for $k = 2$, and "HIGH" for $k = 3$.

5. Representation of granules via fuzzy propositions

The fuzzy sets obtained as described above, can be employed to derive fuzzy propositions that are used to represent the information granules in a linguistic fashion. Precisely, as a result of the fuzzification of granules, each n-dimensional granule (cluster) $c_i, i = 1, \ldots, p$ found in the first step has a "representative" fuzzy set on each input dimension, denoted by $A_{k_{ij}}^j$, such that $\mu_{A_{k_{ij}}^j}(c_{ij}) \geqslant \varepsilon$ (when $\mu_{A_{k_{ij}}^j}(c_{ij}) = \varepsilon$, only one of the two representative fuzzy sets is arbitrarily chosen). Collecting all the representative fuzzy sets for all the n dimensions, the n-dimensional granule c_i is described by the following proposition:

$$c_i := A_{k_{i1}}^1 \textit{ AND } A_{k_{i2}}^2 \textit{ AND } \ldots \textit{ AND } A_{k_{in}}^n \tag{9}$$

such that

$$\mu_{c_i}(\bar{x}) := \bigwedge_{j=1}^{n} \mu_{A_{k_{ij}}^j}(x_j) \tag{10}$$

where $\bar{x} = \langle x_1, \ldots, x_n\rangle$ and $\wedge$ is a τ-norm operator.

In this way it is possible to describe a n-dimensional granule (obtained by the first step of the double-clustering technique) by means of a highly comprehensible fuzzy proposition. It should be noted that, due to the second step of clustering, that provides for each dimension a number K_j of prototypes which is less than the number p of n-dimensional granules, different n-dimensional granules may be represented by the same proposition (9).

The fuzzy propositions can be employed as building blocks of a fuzzy rule-based model which can be used to verify how much the fuzzy granules identified from data are useful in providing good mapping properties in prediction or classification. For example, in case of classification tasks, a set of fuzzy rules can be easily constructed on the basis of the fuzzy propositions, as follows:

$$\begin{array}{c} Rule_r : \text{IF } \left(x_1 \text{ IS } A^1_{k_{r1}}\right) \text{ AND } \left(x_2 \text{ IS } A^2_{k_{r2}}\right) \text{ AND } \cdots \left(x_n \text{ IS } A^n_{k_{rm}}\right) \\ \text{THEN } (\bar{x} \in CLASS_1 \text{ with degree } v_{r1}) \\ \cdots \\ (\bar{x} \in CLASS_m \text{ with degree } v_{rm}) \end{array} \tag{11}$$

where m is the number of disjoint classes from which the input-output data are drawn, and $v_{rl}, l = 1, \ldots, m$ are fuzzy singletons representing the degree to which a pattern $\bar{x}$ belongs to the l-th class. For each rule, the definition of the consequent vector $\bar{v_r} = (v_{r1}, \ldots, v_{rm})$ is accomplished using the information provided by the available input-output data, denoted by $T := \left\{\langle \bar{x}_t, \bar{d}_t \rangle \mid t = 1, 2, \ldots, N\right\}$ and the membership of input data to the all the granules (clusters) that are covered by the proposition in the antecedent of the rule (remember that different granules may be represented by the same proposition). Precisely, let be $\mu_{c_i}(\bar{x})$ the membership of $\bar{x}$ in cluster c_i and let be $I_r := \{i \mid P_r \text{ represents } c_i\}$ the set of all cluster indices that are represented by the same proposition P_r. Then the consequent vector associated to the proposition P_r is obtained as:

$$\bar{v_r} := \text{ mean}_{i \in I_r} \bar{v}_i \tag{12}$$

where $\bar{v}_i$ is defines as:

$$\bar{v}_i := \frac{\sum_{t=1}^{N} \bar{d}_t \cdot \mu_{c_i}(\bar{x}_t)}{\sum_{t=1}^{N} \mu_{c_i}(\bar{x}_t)} \tag{13}$$

Based on a set of R rules as in (11), the inferred class membership degrees for a pattern $\bar{x}$ are obtained as follows:

$$y_l(\bar{x}) = \frac{\sum_{r=1}^{R} \mu_r(\bar{x}) v_{rl}}{\sum_{r=1}^{R} \mu_r(\bar{x})}, \quad l = 1, \ldots, m \tag{14}$$

where $\mu_r(\bar{x}) = \prod_{j=1}^{n} \mu_{jr}(x_j)$ is the rule activation strength.

Note that, since different n-dimensional granules may share the same representative proposition, each rule may cover different n-dimensional granules. This effect can be due to two possible reasons: granules derived in the first step are too fine for the available data (too many clusters found with the first clustering step) or the fuzzy sets are too rough (too few prototypes found in the second step).

6. Illustrative example

The effectiveness of the proposed approach has been evaluated on the well-known Iris data set [3] concerning classification of Iris flowers. Three species of iris flowers (setosa, versicolor and virginica) are known. There are 150 samples (iris flowers), 50 of each class. A sample is a four-dimensional pattern vector representing four attributes of the iris flower concerning sepal length, sepal width, petal length, and petal width.

In the first experiment, to appreciate the ability of the proposed approach in deriving interpretable fuzzy granules, the whole dataset was considered to perform information granulation. In the first clustering step, the FCM was applied to discover 4 granules (clusters) in the 4-dimensional input space. In the second clustering step, hierarchical clustering was applied to the cluster prototypes projected along each dimension, providing 3 one-dimensional clusters per dimension, that were afterwards quantified into as many fuzzy sets. A two-dimensional plot of the results of the information granulation process is illustrated in Figure 2. Following common sense, the 3 fuzzy sets on each dimension were assigned with the labels "LOW", "MEDIUM" and "HIGH". On the basis of the fuzzy sets, a fuzzy proposition was derived to represent each 4-dimensional granule in a linguistic form. Such fuzzy propositions are reported in Table 1.

Table 1. Representation of information granules via fuzzy propositions.

Granule 1	*petal_length* IS *LOW* AND *petal_width* IS *HIGH* AND *sepal_length* IS *LOW* AND *sepal_width* IS *LOW*
Granule 2	*petal_length* IS *MEDIUM* AND *petal_width* IS *LOW* AND *sepal_length* IS *MEDIUM* AND *sepal_width* IS *MEDIUM*
Granule 3	*petal_length* IS *MEDIUM* AND *petal_width* IS *MEDIUM* AND *sepal_length* IS *LOW* AND *sepal_width* IS *LOW*
Granule 4	*petal_length* IS *HIGH* AND *petal_width* IS *MEDIUM* AND *sepal_length* IS *HIGH* AND *sepal_width* IS *HIGH*

Due to the fuzzy nature of the propositions, all input patterns have non zero membership degree to each proposition, even if they do not fall in regions of pattern space covered by the granules. This effect is depicted in Figure 3, where the influence of each proposition on the input data is shown in the subspace "petal length-petal width".

To verify how much the granules induced by the proposed approach are useful in providing good mapping properties, a further experiment was carried out using a 10-fold cross validation technique. Specifically, in each of the ten trials, the training set

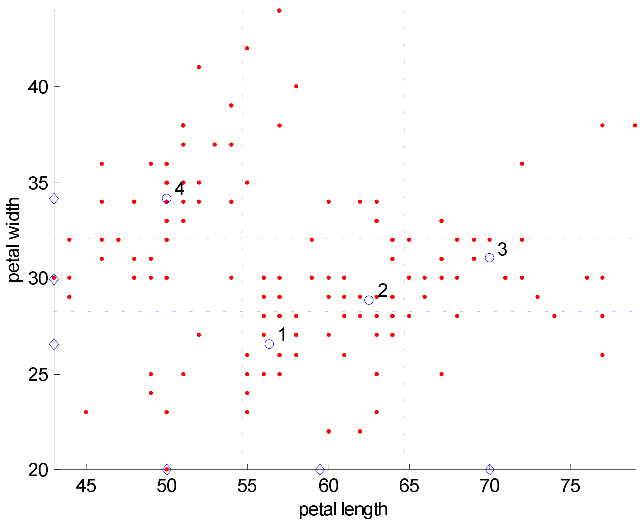
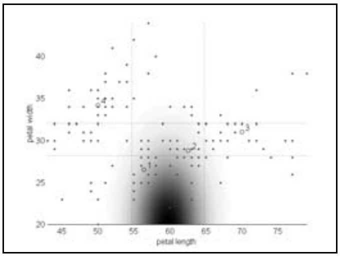


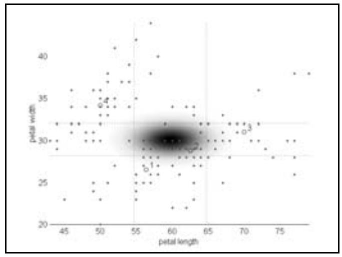
Figure 2. Result of the information granulation process plotted in two dimensions. The 4 cluster prototypes discovered by FCM (circles) are projected on each axis and further clustered to produce 3 prototypes (diamonds) on each dimension, resulting in 3 fuzzy sets per input variable. Dashed lines represent intersection points between adjacent fuzzy sets

was used to perform information granulation and to build fuzzy rule-based classifiers on the basis of the extracted granules (as described in Section 5), while the test set was used to check the classification ability of the constructed fuzzy classifiers. Such classifiers (denoted in the following by DC) were derived by considering 3 and 5 fuzzy sets per dimension (fspd) respectively. Moreover, the DC classifiers were compared with fuzzy classifiers based on standard FCM, with different number of p cluster, where $p \in \{3, 4, 5, 6, 8, 10, 15, 20\}$. In the latter case, the classification mapping was defined in the following way. Given the partition matrix U, the class membership for each cluster prototype is given by $U \cdot K$ scaled columnwise by the cluster cardinality, where $K = [k_{tl}]$ such that k_{tl} is 1 if the t-th pattern belongs to l-th class, 0 otherwise. After that, fuzzy granules were derived from prototypes by associating a Gaussian membership function with center in each prototype and circular amplitude found by trial-and-error (selection of the amplitude with minimum classification error).

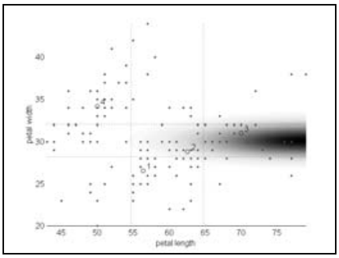
The classification results of the fuzzy classifiers are summarized in Table 2 and illustrated in Figure 4. As it can be seen, the quality of the DC fuzzy classifiers overcome the FCM-based fuzzy classifiers both in terms of classification ability and compactness of the rule base. Indeed, as the number of clusters increases, the number of rules of DC classifiers keeps low with respect to that of FCM classifiers. In addition, fuzzy rules of DC classifiers are nicely interpretable, as it can be seen from Table 3, where the rule



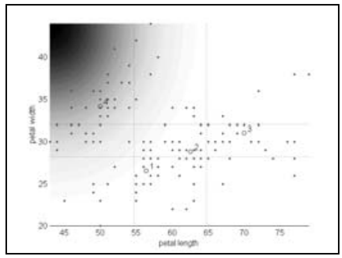
Proposition 1



Proposition 2



Proposition 3



Proposition 4

Figure 3. Influence of the 4 fuzzy propositions on the input data in two-dimensional subspace: the darker is the area, the higher is the truth value of the proposition

base of one of the DC fuzzy classifiers with 6 rules and 3 fuzzy sets per dimension is described, and from Figure 5 that plots the fuzzy sets on each dimension.

Table 2. Classification results of the fuzzy classifiers.

p	FCM			DC 3 fspd				DC 5 fspd			
	Test Error		Rules	Test Error		Rules		Test Error		Rules	
	Mean	Std		Mean	Std	Min	Max	Mean	Std	Min	Max
3	20.0%	10.8%	3	14.0%	7.6%	3	3	-	-	-	-
4	18.0%	7.3%	4	14.7%	6.5%	4	4	-	-	-	-
5	16.7%	8.6%	5	12.7%	7.0%	5	5	19.3%	11.3%	5	5
6	16.0%	8.0%	6	10.7%	8.0%	5	6	14.0%	7.6%	6	6
8	13.3%	18.1%	8	6.7%	8.0%	6	8	7.1%	8.5%	8	8
10	5.3%	7.8%	10	7.3%	4.7%	6	8	5.3%	5.0%	9	10
15	5.3%	4.0%	15	4.0%	3.3%	8	10	4.7%	3.1%	12	15
20	3.3%	4.5%	20	5.3%	5.8%	7	9	5.3%	6.5%	15	19

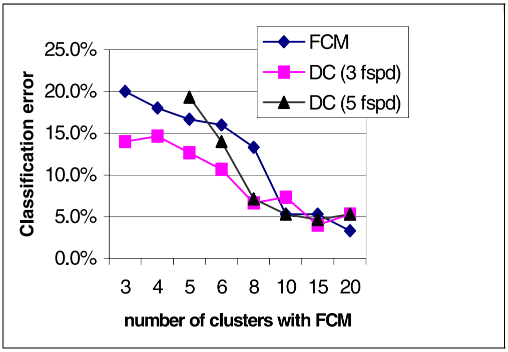


Figure 4. Comparison of the fuzzy classifiers in terms of classification error

Table 3. Example of Fuzzy Rule Base induced by the proposed approach.

R1: IF *petal_length* IS *LOW* AND *petal_width* IS *LOW* AND *sepal_length* IS *MEDIUM* AND *sepal_width* IS *MEDIUM* THEN *flower* IS *SETOSA* WITH DEGREE 0.02, *VIRGINICA* WITH DEGREE 0.86, *VERSICOLOUR* WITH DEGREE 0.11
R2: IF *petal_length* IS *LOW* AND *petal_width* IS *MEDIUM* AND *sepal_length* IS *LOW* AND *sepal_width* IS *LOW* THEN *flower* IS *SETOSA* WITH DEGREE 0.94, *VIRGINICA* WITH DEGREE 0.05, *VERSICOLOUR* WITH DEGREE 0.02
R3: IF *petal_length* IS *LOW* AND *petal_width* IS *HIGH* AND *vsepal_length* IS *LOW* AND *sepal_width* IS *LOW* THEN *flower* IS *SETOSA* WITH DEGREE 0.94, *VIRGINICA* WITH DEGREE 0.05, *VERSICOLOUR* WITH DEGREE 0.02
R4: IF *petal_length* IS *MEDIUM* AND *petal_width* IS *HIGH* AND *sepal_length* IS *LOW* AND *sepal_width* IS *MEDIUM* THEN *flower* IS *SETOSA* WITH DEGREE 0.01, *VIRGINICA* WITH DEGREE 0.63, *VERSICOLOUR* WITH DEGREE 0.36
R5: IF *petal_length* IS *MEDIUM* AND *petal_width* IS *MEDIUM* AND *sepal_length* IS *HIGH* AND *sepal_width* IS *HIGH* THEN *flower* IS *SETOSA* WITH DEGREE 0.01, *VIRGINICA* WITH DEGREE 0.17, *VERSICOLOUR* WITH DEGREE 0.82
R6: IF *petal_length* IS *HIGH* AND *petal_width* IS *MEDIUM* AND *sepal_length* IS *HIGH* AND *sepal_width* IS *HIGH* THEN *flower* IS *SETOSA* WITH DEGREE 0.01, *VIRGINICA* WITH DEGREE 0.11, *VERSICOLOUR* WITH DEGREE 0.87

7. Conclusions

In this paper, an information granulation approach has been proposed which lies on a double-clustering technique to derive granules that can be easily quantified in terms of interpretable fuzzy sets.

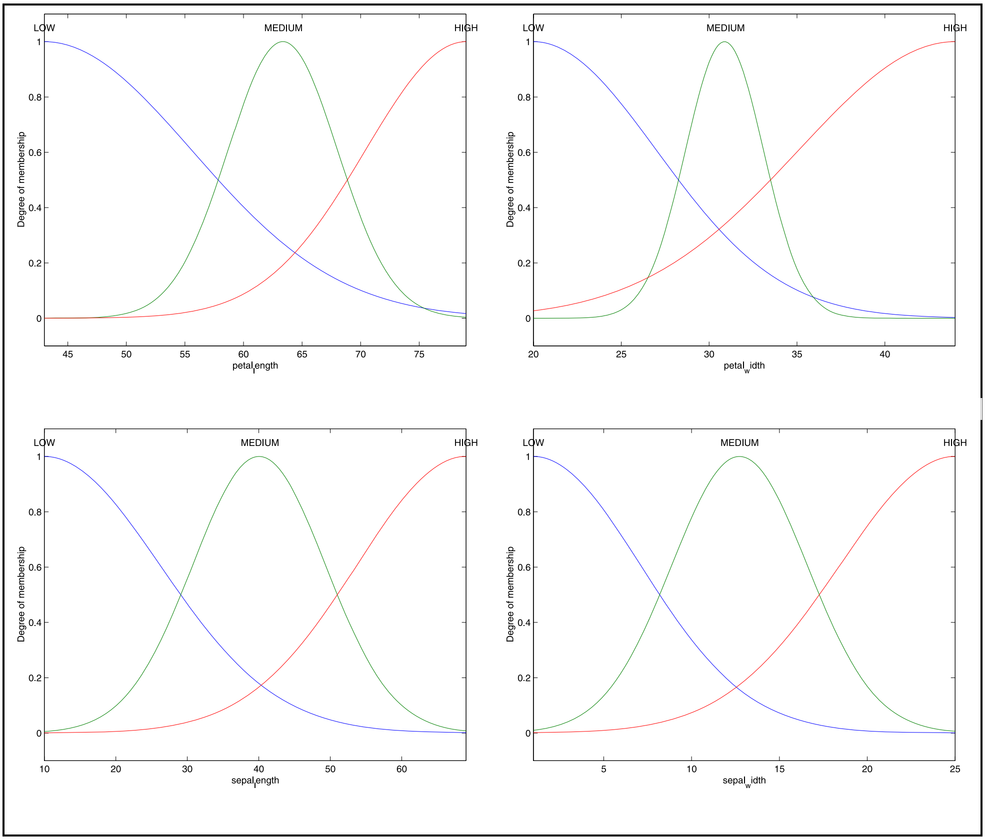


Figure 5. Fuzzy sets used in the rule base of Table 3

The approach has been illustrated with the aid of a benchmark classification data set that showed how the induced information granules can reveal the underlying structure in the data and how they impact on the results of classification.

In this work, the first step of the double-clustering technique is based on the standard FCM method. However, a variety of other clustering algorithms can be used as well. In particular, a different version of the proposed double-clustering technique could rely on the Conditional FCM proposed by Pedrycz, to exploit the contextual information in discovering granules from data.

A further extension may concern the optimization of the extracted fuzzy granules. To do so, the granules can be embedded into a neurofuzzy network that can be trained to optimally adjust the shape and the distribution of fuzzy granules so as to better capture the relationships among data.

References

[1] W. GAWRONSKI: Balanced Control of Flexible Structures. *Lecture Notes in Control and Information Sciences*, **211**, Springer-Verlag, (1996).

[2] A. ILCHMANN: Non-Identifier-Based High-Gain Adaptive Control. *Lecture Notes in Control and Information Sciences*, **189**, Springer-Verlag, (1993).

[3] S.M. JOSHI: Control of Large Flexible Space Structures. *Lecture Notes in Control and Information Sciences*, **131**, Springer-Verlag, (1989).

[4] T. KACZOREK: Linear Control Systems Vol.1. Research Studies Press Ltd., 1992.

[5] T. KATO: Pertubation Theory for Linear Operators. Springer-Verlag, 1976.

[6] J. LA SALLE and S. LEFSCHETZ: Stability by Liapunov's Direct Method with Applications. Academic Press, 1961.

[7] Z. LIN: Low Gain Feedback. *Lecture Notes in Control and Information Sciences*, **240**, Springer-Verlag, (1999).

[8] H. LOGEMANN and S. TOWNLEY: Discrete-time low-gain control of uncertain infinite-dimensional systems. *IEEE Trans. on Automatic Control*, **42** (1997), 22-37.

[9] H. LOGEMANN and S. TOWNLEY: Adaptive control of infinite-dimensional systems without parameter estimation: an overview. *IMA J. Mathematical Control and Information*, **14** (1997), 175-206.

[10] H. LOGEMANN and S. TOWNLEY: Low-gain control of uncertain regular linear systems. *SIAM J. Cont. Opt.*, **35** (1997), 78-116.

[11] D.E. MILLER and E.J. DAVISON: The self-tuning robust servomechanism problem. *IEEE Trans. on Automatic Control*, **34** (1989), 511-523.

[12] P.K. SINHA: Multivariable Control. Marcel Dekker Inc., 1984.

[13] T. WILLIAMS: Transmission-zero bounds for large space structures, with applications. *J. Guidance, Control, and Dynamics*, **12** (1989), 33-38.