2003 International Symposium on Advanced Intelligent Systems
September 25-28, 2003, Jeju, Korea

DCClass: a Tool to Extract Human Understandable Fuzzy Information Granules for Classification

Giovanna Castellano, Anna M. Fanelli, Corrado Mencar

Department of Informatics – University of Bari

v. E. Orabona, 4 – 70126 – Bari – ITALY

e-mail: {castellano, fanelli, mencar}@di.uniba.it

***Abstract*—In this paper we describe DCClass, a tool for fuzzy information granulation with transparency constraints. The tool is particularly suited to solve fuzzy classification problems, since it is able to automatically extract information granules with class labels. For transparency pursuits, the resulting information granules are represented in form of fuzzy Cartesian product of one-dimensional fuzzy sets. As a key feature, the proposed tool is capable to self-determining the optimal granularity level of each one-dimensional fuzzy set by exploiting class information. The resulting fuzzy information granules can be directly translated in human-comprehensible fuzzy rules to be used for class inference. The paper reports preliminary experimental results on a medical diagnosis problem that shows the utility of the proposed tool.**

***Index Terms*—DCClass, Fuzzy Information Granules, Transparency Constraints, Fuzzy Rules, Classification.**

I. Introduction

Granular Computing (GrC) is an emerging paradigm that deals with representing and processing information in the form of some aggregates, called information granules [1]. The representation frame for information granules is fundamental and is driven by the application domain. Nevertheless, among different forms of information aggregation, fuzzy sets are one of the most appealing, due to their closeness to the human way to abstract concepts from the observed environment [11].

A key task in GrC is the so called information granulation process, which is responsible in the formation of information aggregates from a set of available data. A methodological and algorithmic issue is the formation of transparent fuzzy information granules, meaning that they should provide a clear and understandable description of patterns held in data. Such fundamental property can be formalized by a set of constraints that must be satisfied during the information granulation process [5], [7].

In this work, we propose a tool for fuzzy information granulation with transparency constraints, which is particularly appropriate to solve classification problems. The tool is called DCClass (Double Clustering for Classification) and it is based on an enhanced version of our Crisp Double Clustering (CDC) algorithm proposed in [3],[4]. DCClass provides a set of information granules represented in form of Cartesian product of one-dimensional fuzzy sets. As a key feature of the proposed tool, the granularity of the derived one-dimensional fuzzy sets is optimally calculated by exploiting available class information, thus recovering the user from an arbitrary choice of the granularity level of each fuzzy set.

Each of the information granules returned by DCClass is associated to class labels, so fuzzy classification rules can be directly defined. Such rules constitute the knowledge base for a fuzzy inference system, which can be conveniently used to solve fuzzy classification problems, as well as to validate the derived information granules in terms of their adherence to the available data.

To show the effectiveness of the proposed tool, a medical case study is considered in this work. Preliminary experimental results show that DCClass is able to extract human understandable fuzzy information granules, which define fuzzy classification rules that can be used for diagnosis prediction, as well as to provide a knowledge base for further understanding and knowledge refinement.

II. Transparent Fuzzy Classification

In this Section, we provide a description of the problem of fuzzy classification properly solved by a set of fuzzy rules. In addition, we formulate a set of constraints that we adopt as definition of transparency. For sake of simplicity, we consider only two-class classification problems, though the extension to multi-class problems is straightforward.

The classification problem is implicitly defined by a dataset of N pairs:

$$\mathbf{D} = \left\{(\mathbf{x}_i, c_i),\ i = 1,2,\ldots,N\right\} \tag{1}$$

where each $\mathbf{x}_i = (x_1, x_2, \ldots, x_n)$ is a vector in $\mathbb{R}^n$ and c_i is the class discriminant that can assume one of two possible values $C = \{C_1, C_2\}$. Without loss of generality, we assume that the input vector belongs to a closed hyper-interval, defined as:

$$\mathbf{X} = [m_1, M_1] \times \cdots \times [m_n, M_n] \tag{2}$$

being $[m_i, M_i] \subset \mathbb{R}$, $i = 1,2,\ldots,n$.

Given such dataset, the problem is to derive a set of R fuzzy rules that accurately classify the elements of the available dataset as well as newly observed examples. The rules have the following form:

$$\text{IF } \mathbf{x} \text{ is } \mathbf{A}^{(r)} \text{ THEN class is } c^{(r)} \tag{3}$$

where $r = 1,2,\ldots,R$ is the rule index, $\mathbf{A}^{(r)}$ is an information granule represented as a multi-dimensional fuzzy set defined on the entire domain $\mathbf{X}$, and $c^{(r)} \in C$ is the class discriminant of the rule. When an input vector is given, the inferred class membership values are computed as follows:

$$\mu_{C_j}(\mathbf{x}) = \frac{\sum_{r:c^{(r)}=C_j} \mu_{\mathbf{A}^{(r)}}(\mathbf{x})}{\sum_{r=1}^{R} \mu_{\mathbf{A}^{(r)}}(\mathbf{x})}, \quad j = 1,2 \tag{4}$$

where $\mu_{\mathbf{A}^{(r)}}(\mathbf{x})$ is the membership value of pattern $\mathbf{x}$ in the information granule $\mathbf{A}^{(r)}$. When crisp classification in required, the class with maximum membership value is selected.

Many different rule extraction techniques exist in literature that can provide accurate rulesets in an efficient way (see, e.g. [6],[10]) However, most of these techniques do not take into account the transparency requirement of the discovered knowledge base: as a consequence, the resulting rules are accurate but lack in human understandability.

Informally speaking, the transparency of the fuzzy rules concerns the possibility for human users to read and understand the knowledge base acquired from data. Such requirement can be translated in a formal fashion by a set of constraints to be satisfied in the knowledge extraction process. In this work, we adopt the following constraints, which are commonly embraced in specialized literature (see, e.g. [5]) and stand out for their simplicity and general applicability:

1. Each information granule must be defined as a Cartesian product of one-dimensional fuzzy sets:

$$\mathbf{A}^{(r)} = A_1^{(r)} \times A_2^{(r)} \times \cdots \times A_n^{(r)} \tag{5}$$

being each one-dimensional fuzzy set $A_i^{(r)}$ defined over the interval $[m_i, M_i]$. The membership function of the multidimensional granule is induced by the membership functions of the one-dimensional fuzzy sets and a T-norm operator $\otimes$ as follows:

$$\mu_{\mathbf{A}^{(r)}}(\mathbf{x}) = \mu_{A_1^{(r)}}(x_1) \otimes \cdots \otimes \mu_{A_n^{(r)}}(x_n) \tag{6}$$

2. Each one-dimensional fuzzy set must be normal, unimodal and convex. In other words, for each fuzzy set there exist only one element (called *prototype*) with maximum membership 1.0, while all other elements have a membership value that decreases as the distance from the prototype increases;
3. For each input dimension i, the extreme values m_i and M_i should be prototypes for some fuzzy sets that are respectively called leftmost and rightmost fuzzy sets;
4. For each element in the interval $[m_i, M_i]$ there must exist at least one fuzzy set that yields a membership value greater that a specified coverage threshold ε (usually $\varepsilon = 0.5$);
5. Two fuzzy sets defined on the same input dimension must not overlap too much, i.e. their possibility measure should not exceed a specified threshold. In this work, we set the possibility threshold to be equal to the coverage threshold ε.

It should be noted that constraint 1 concerns the multidimensional granules, while all the other constraints refer to their one-dimensional projections. In particular, constraint 2 can be guaranteed if a proper choice of membership functions is made. In this work, we adopt Gaussian membership functions, characterized by two parameters, namely the center ω and the width σ:

$$\mu_A(x) = \exp\left(-\frac{(x-\omega)^2}{2\sigma^2}\right) \tag{7}$$

It is easy to show that Gaussian membership functions satisfy constraint 2, whatever is the choice of the parameters ω and σ. In contrast, the remaining constraints can be satisfied only for proper choices of the algorithm to generate fuzzy information granules. When transparency constraints are satisfied, meaningful labels can be assigned to each one-dimensional fuzzy sets, like LOW, MEDIUM, HIGH, etc., depending on the relative position of their prototypes. Moreover, multidimensional granules can be represented as conjunction of labels, like 'x_1 is LOW AND x_2 is HIGH', thus conveying immediately interpretable knowledge that can be used in fuzzy inference.

In addition to the abovementioned constraints, it is desirable to follow some guidelines that suggest a low number of rules, as well as a low number of one-dimensional fuzzy sets per dimension. As a consequence, the rule extraction algorithm should take into account also these guidelines in order to provide a more legible knowledge base.

III. DCCLASS

In this Section, we describe the proposed DCClass tool. The tool is based on our CDC algorithm, with a substantial modification to accommodate class information within the information granulation process.

DCClass is defined as a composition of three sequential steps that are depicted with an illustrative example in Figure 1 and are described as follows. First, available data is compressed by means of a vector quantization algorithm, with the aim of derive a set of multidimensional prototypes that capture multidimensional relationships in data (fig. 1a). Among different vector quantization strategies, those that exploit class information are preferred because of their superior accuracy. In particular, in this work we adopt the LVQ1 (Learning Vector Quantization, version 1) algorithm as described in [8]. It should be noted that each multidimensional prototype is associated to a class label, which will be used in the subsequent steps of granules formation.

Since a direct fuzzification of the derived prototypes may result in nonsensical fuzzy information granules, a second step is performed. Specifically, the multidimensional prototypes are projected onto each input axis, being each projection associated to a class label according to the original

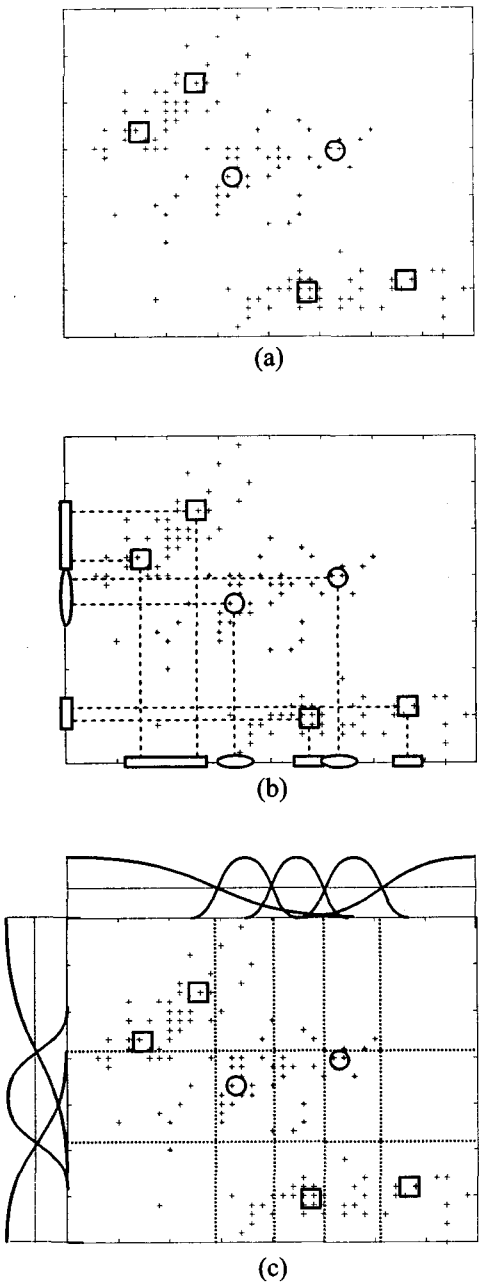


Figure 1: Illustrative example of the three stages of DCClass. (a) Data points belonging to two classes (crosses and stars) are compressed by LVQ (squares and circles). (b) Prototypes of the codebook are projected onto each dimension and then clustered. (c) Midpoints between clusters are used to define fuzzy sets with transparency constraints

prototype (fig. 1b). For each dimension, the projections are clustered together according to the following criterion: the adjacent projections of the same class are grouped together, while the projections of different classes belong to different clusters. The rationale behind such step consists in merging similar projections in a single cluster, as long as they belong to the same class. In this way, fuzzy sets will be shared by different information granules, thus improving the legibility of the resulting knowledge base.

As a final step, midpoints between edges of adjacent one-dimensional clusters (i.e. the mean value between the maximum element of a cluster and the minimum element of its next adjacent cluster) are calculated. Such midpoints are used to define the intersection points between the membership functions of adjacent fuzzy sets (fig. 1c). The fuzzy sets so derived satisfy all the transparency constraints defined previously. Finally, multidimensional granules are formed by a combination of one-dimensional fuzzy sets, one for each dimension. Combinatorial explosion of the number of granules is avoided by selecting only the fuzzy granules that well represent the multidimensional prototypes discovered in the first step.

The combination of the three steps provides a definition of fuzzy information granules that capture multidimensional relationships on data and are represented in form of human understandable fuzzy sets. Moreover, as a key feature, DCClass automatically provides the granularity level of fuzzy sets for each dimension, by exploiting information on class distribution. Only the number of multidimensional prototypes to be discovered in the first step has to be specified.

Finally, the resulting fuzzy information granules are directly translated into classification rules as follows. If an information granule is defined as a Cartesian product of fuzzy sets $A_{i_1}, A_{i_2}, \ldots, A_{i_n}$ and is associated to class C_j, then a classification rule is formed as follows:

$$\begin{aligned} &\text{IF } x_1 \text{ is } A_{i_1} \text{ and } \cdots \text{ and } x_n \text{ is } A_{i_n} \\ &\text{THEN class is } C_j \end{aligned} \tag{8}$$

Each fuzzy granule defines a rule. As a consequence, the number of resulting rules is upper bounded by the number of multidimensional prototypes discovered in the first step of DCClass computation.

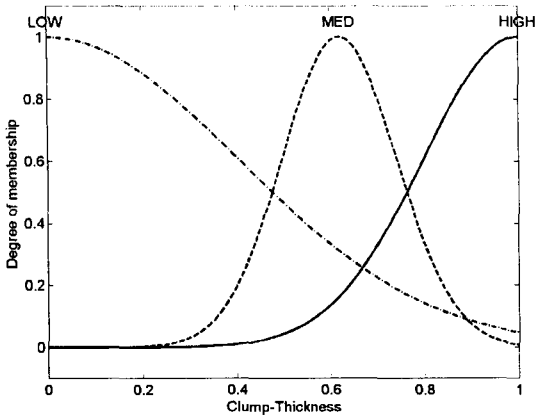
IV. Experimental Results

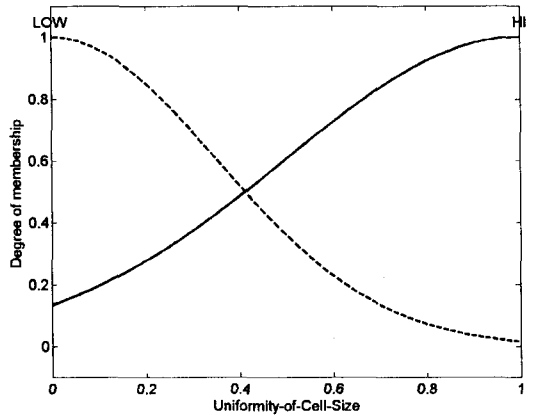
To illustrate the effectiveness of the proposed tool, the medical case study of Wisconsin Breast Cancer (WBC) diagnosis is considered. In particular, the WBC dataset has been retrieved from the UCI Machine Learning Repository [2], where data and detailed information can be found. The dataset consists of 683 examples[1] described by nine continuous attributes labeled with one out of two possible diagnoses: malignant and benignant cancer.

The dataset is split according to the ten-fold stratified cross validation, so each simulation run is repeated ten times and average results are drawn. In addition, we fix the maximum number of rules (i.e. the number of multidimensional prototypes to be discovered in the first step of DCClass) to six, and we adopt the product as the T-norm for the Cartesian product in (6).

The application of DCClass provides ten rule sets with average classification error on the test sets of 3.975%, while the mean number of rules is 3.6. When compared with similar approaches, like NEFCLASS [9], the achieved results confirm that the proposed tool is a valid technique to extract accurate knowledge from data.

[1] The original dataset consists of 699 examples from which we discard 16 cases with missing values.





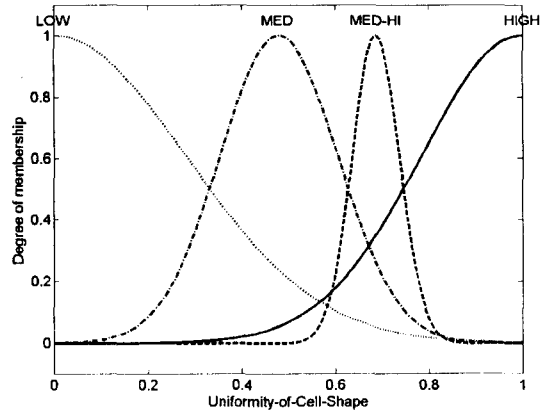


Figure 2: Fuzzy sets defined by DCClass for three input variables

To assess the transparency of the resulting information granules, we choose a rule set of five rules that provides a classification error of 1.471% on the test set. The one-dimensional fuzzy sets derived from the application of DCClass satisfy the transparency constraints defined in Section II, as illustrated in Figure 2 for three input dimensions. For such fuzzy sets, the association of meaningful linguistic label is straightforward. In Figure 3, we report two of the five rules to illustrate how transparent rules are formed on the basis of information granules discovered by DCClass.

If Clump-Thickness is LOW AND Uniformity-of-Cell-Size is LOW AND Uniformity-of-Cell-Shape is LOW AND Marginal-Adhesion is LOW AND Single-Epithelial-Cell-Size is LOW AND Bare-Nuclei is LOW AND Bland-Chromatin is LOW AND Normal-Nucleoli is VERY-LOW AND Mitoses is MED-LOW **Then** *MALIGNANT*.

If Clump-Thickness is MED AND Uniformity-of-Cell-Size is HIGH AND Uniformity-of-Cell-Shape is HIGH AND Marginal-Adhesion is HIGH AND Single-Epithelial-Cell-Size is HIGH AND Bare-Nuclei is HIGH AND Bland-Chromatin is HIGH AND Normal-Nucleoli is HIGH AND Mitoses is HIGH **Then** *BENIGNANT*.

Figure 3: Two of the five rules discovered by DCClass

V. CONCLUSIONS

We have presented DCClass, a tool to extract human interpretable fuzzy information granules from data, which are particularly suited to solve classification problems. As supported by experimental results, the tool is able to discover both accurate and transparent granules that can be conveniently represented in form of fuzzy rules. The proposed tool can be further improved by optimally reducing the number of one-dimensional fuzzy sets per input and the number of input variables for each rule, so as to provide simpler classification rules without significant loss of accuracy. Algorithmic solutions with this aim are under investigation.

REFERENCES

[1] Bargiela, A.; Pedrycz, W., *Granular Computing: An Introduction*, Kluwer Academic Press, 2002

[2] Blake, C.L.; Merz, C.J., *UCI Repository of machine learning databases*. Irvine, CA: University of California, [http://www.ics.uci.edu/~mlearn/MLRepository.html], 1998

[3] Castellano G.; Fanelli, A.M.; Mencar, C., "Fuzzy Granulation of multi-dimensional data by a Crisp Double-Clustering algorithm", in *Proc. of 7th Multi-Conference on Systemics, Cybernetics and Informatics (SCI2003)*, Orlando, Fl, US, 2003

[4] Castellano, G.; Fanelli, A.M.; Mencar, C., "Generation of interpretable fuzzy granules by a double clustering technique", *Archives of Control Sciences: Special issue on Granular Computing*, vol. 12 no. 4, pp. 397-410, 2002

[5] De Oliveira, J.V., "Towards Neuro Linguistic Modeling: Constraints for Optimization of Membership Functions," *Fuzzy Set and Systems*, 106: 357-380, 1999

[6] Jang, J-S.R.; Sun, C-T., "Neuro-Fuzzy Modeling and Control", *Proceedings of IEEE*, vol. 83, pp.378-406, IEEE, 1995

[7] Jin Y.; Von Seelen W.; Sendhoff, B., "On generating FC3 Fuzzy rule systems from data using evolution strategies" *IEEE Transactions on Systems, Man and Cybernetics - Part B*, vol. 29, no. 6, pp. 829-845, 1999

[8] Kohonen, T., *Self Organizing Maps*, Springer Verlag, Heidelberg, 1995

[9] Nauck, D.; Kruse, R., "Obtaining interpretable fuzzy classification rules from medical data," *Artificial Intelligence in Medicine*, vol. 16, pp. 149-169, 1999

[10] Sugeno, M.; Kang, G.T., "Structure identification of fuzzy model," *Fuzzy Sets and Systems*, vol. 28, pp. 15-33, 1988

[11] Zadeh, L.A., "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic", *Fuzzy Sets and Systems*, vol. 90, pp. 117-117, Elsevier, 1997