2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

From Data to Granular Data and Granular Classifiers

Rami Al-Hmouz[1], Witold Pedrycz[1] [2] [3], Abdullah Balamash[1] , and Ali Morfeq[1]
(1) Electrical and Computer Engineering Department
King Abdulaziz University, Jeddah, Saudi Arabia
ralhmouz@kau.edu.sa, asbalamesh@kau.edu.sa, morfeq@kau.edu.sa
(2) Department of Electrical & Computer Engineering
University of Alberta, Edmonton, Alberta, T6R 2V4 Canada
(3) Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland
wpedrycz@ualberta.ca

***Abstract*— Information granules emerging as a result of an abstract and more condensed and global view at numeric data play an essential role in various pattern recognition pursuits. In this study, we investigate an idea of granular prototypes (representatives) and discuss their role in the realization of classification schemes. A two-stage procedure of a formation of information granules is discussed. We show how the commonly used clustering methods are viewed as a prerequisite for the construction of granular prototypes. In this regard, a certain version of the principle of justifiable granularity is investigated. In the sequel, a characterization of information granules expressed in terms of their information (classification) content is provided and its usage in the realization of a classifier is studied. Experimental studies involving both synthetic and publicly available data are reported.**

Keywords*— *pattern classification, information granules, Granular Computing, clustering, principle of justifiable granularity, Fuzzy C-Means

I. INTRODUCTION

Clustering and classification algorithms are predominantly realized at the numeric level. Classification schemes along with their linear and nonlinear discrimination functions are numeric constructs and in their design we usually use numeric data. Granular Computing [1- 4] with its underlying concept of information granules [5], establishes a different view at classification processes and classification results. Interestingly enough, information granules have been studied from different perspectives resulting in different terminology and conceptual focus; one can refer here to so-called symbolic data [6]. Clustering has been a synonym of information granulation. Depending upon the nature of the detailed clustering algorithm, we refer to fuzzy clustering, rough clustering, etc. Irrespectively of the mechanisms being involved there, the produced entities are numeric. In this study, we propose to move further along this line by building upon such numeric results of clustering (say, *numeric* prototypes) and form granular representatives (*granular* prototypes) of more general character and exploit them in the ensuing classification pursuits. Our key objective of this study is to establish a general conceptual and design framework of granular representation of numeric data. A formation of granular prototypes (representatives) of data consists of two main phases. First, we move from numeric data to a collection of numeric prototypes. Second, we augment (abstract) the prototypes by making them granular. The intent is also to arrive at the characterization of such granular prototypes in terms of its information content (classification-oriented) and investigate classification mechanisms. Through this characterization, granular prototypes can also deliver a sound and comprehensive description of the data and in this way may serve as a sound data analysis vehicle. In comparison with the existing literature, the study exhibits several evident aspects of originality: (i) we identify a genuine and fully legitimate need behind an emergence of information granules and their function in the representation of numeric data, (ii) we deliver a systematic development environment in which initially formed numeric representatives (formed as a result of clustering or fuzzy clustering) are made granular by invoking the principle of justifiable granularity, and (iii) the diversity of data are quantified in terms of information granularity.

The study is structured as follows. We look at the construction of information granules by discussing a two-phase development process (Section II) and then elaborate on at the characterization of granular prototypes (Section III). The ensuing classification mechanism is discussed in Section IV. Numeric experiments are reported in Section V. Conclusions are covered in Section VI.

II. CONSTRUCTION OF INFORMATION GRANULES

The development of information granules is realized in two main phases. First, numeric data $\boldsymbol{x}_k$ positioned in $\boldsymbol{R}^n$ are clustered in the sequel giving rise to some condensed numeric description of the structure. A common characterization of the structure of data is provided in terms of a finite number of prototypes (centroids, medoids or other meaningful representatives of the data) that can be deemed as a family of concise descriptors of the data. Prototypes are distributed in the data space in a way they capture the essence of the data (commonly they are localized in the regions where we encounter the highest density of data). In case of fuzzy clustering, the results of clustering are conveyed in the form of the prototypes and a fuzzy partition matrix capturing information about membership grades. There is a plethora of clustering and fuzzy clustering algorithms including such commonly used techniques as K-Means and Fuzzy C–Means (FCM) [7][8] and their numerous variants. Irrespectively from the algorithmic details of the methods being considered, the commonly

reported results come as *numeric* prototypes and a *numeric* partition matrix [7-10]. Given the data set comprising "N" data (patterns) $\boldsymbol{x}_1, x_2, \ldots, x_N$ located in the n-dimensional space of real vectors, $\boldsymbol{R}^n$, clustering these data into c clusters produces prototypes $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_c$ and the fuzzy partition matrix $\boldsymbol{U} = [u_{ik}]$, i=1, 2, ,…, c; k=1, 2, …, N where $u_{ik} \in [0,1]$ with the two "standard" requirements as to the allocation of membership values across the clusters. More specifically, for the k-th data point one has $\sum_{i=1}^{c} u_{ik}=1$ and the boundary conditions with respect to the content of the i-th cluster stating that $0 < \sum_{k=1}^{N} u_{ik} < N$

As remarked in [11], it is somewhat surprising to see that the representatives (such as prototypes) of the numeric data are only numeric entities. One might have anticipated that in order to fully reflect the diversity of the represented data, the representatives themselves could have been sought as constructs emerging at the higher level of abstraction, viz. coming in the form of information granules. Information granularity becomes inherent as a conceptual vehicle to quantify the diversity of numeric data embraced by the information granule. To adhere to this line of thought, starting with the numeric prototypes, we build information granules around the numeric ones.

The principle of justifiable granularity discussed in [2] delivers a general design methodology by emphasizing and forming a sound tradeoff between justifiability of the granule (being viewed in terms of the numeric evidence, namely the data being embraced by the constructed information granule) and the specificity of the granule itself. If the experimental justifiability is of primary concern, we look at the σ-count of the information granule. Let us consider an interval-valued prototype and discuss its realization along the individual coordinates. . By moving to the right from the j-th coordinate of the i-th prototype, we determine the endpoint of the interval, say b_{ij} such that the value of the following ratio

$$\frac{\sum_{k:x_{kj} \in [v_{ij}, b_{ij}]} u_{ik}}{\sum_{k:x_{kj} \geq v_{ij}} u_{ik}} \geq p \qquad (1)$$

does not fall below a certain predetermined threshold p, $p \in [0,1]$. In the analogous way, we form the lower bound of the interval a_{ij}. Note that the expression standing in the nominator of (1) is a total sum of all elements (data) falling within the interval.

An overall scheme discussed above is illustrated in Figure 1. While the increasing values of p give rise to larger (more abstract, general) information granules, these intervals should not overlap so this restricts the admissible values of p to the range $[0, p_{\max}]$ where $p_{\max}$ is the maximal value of the ratio shown above for which none of the two prototypes overlap each other.

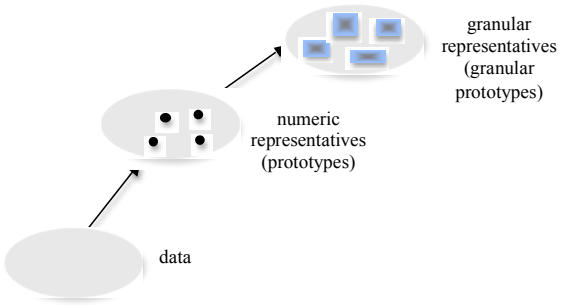


Figure 1. The essence of the two-phase development of information granules: from numeric data – numeric prototypes and their granular counterparts

III. A CHARACTERIZATION OF GRANULAR PROTOTYPES

The information granules – granular prototypes formed in the way described above, can be described by looking at its (i) information content, and (ii) size (volume). Information content depends on the nature of the data themselves. In classification problems where data come with class labels, the mixture of data (patterns) embraced by the granular prototype determine information content of the granular prototype. More specifically, for the i^{th} granular prototype, we count the number of patterns falling within the bounds of the granular prototype and belonging to different classes. This result in the vector of class membership $\boldsymbol{f}_i$. Ideally, given "r" classes encountered in the classification problem, the granular prototype could include only patterns belonging to the same class. The signature of this prototype is then a Boolean vector $\boldsymbol{f}_i$ =[0 ….0 1 0…0]. In the worst case where a substantial mixture of pattern occurs, the entries of the above vector are getting close to $1/r$. Of course, when the values of p get higher (and the size of the information granules increases), the likelihood of having a more visible mixture of classes present within the granular prototype increases.

The size (volume) of the granular prototype is another important descriptor. We express it by taking a product of the ratios of the length of the interval and the ranges of the corresponding variables. For each coordinate (variable), we form the granular (interval-valued) prototype $\boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_c$ and

$$\text{Vol}(\boldsymbol{V}_i) = \prod_{j=1}^{n} \frac{|b_{ij} - a_{ij}|}{range_j} \qquad (2)$$

where $range_j$ is the range of the values assumed by the j-th variable.

Having a collection of information granules of the prototypes, we invoke a classification mechanism.

IV. CLASSIFICATION WITH THE USE OF GRANULAR PROTOTYPES

The classification of any new pattern $\mathbf{x}$ is carried out in two steps. First, a membership degree of $\boldsymbol{x}$ to each $\boldsymbol{V}_i$ is determined. As usual, any ensuing computing involves a determination of distance of $\boldsymbol{x}$ from the corresponding information granule. We discuss here a detailed optimization problem that to a significant extent resembles what has been done in the optimization completed by the FCM itself.

Furthermore because of the granular nature of the entity used in the computation of distances, it is apparent that a degree of membership of $\boldsymbol{x}$ to $\boldsymbol{V}_i$ (no matter how it is determined) has to be reflective of the fact that such distance could be granular (not numeric) and in the simplest case should exhibit an interval nature of possible values, say $u_i = [u_i^-, u_i^+]$. Second, the granular (interval-valued) membership are aggregated with the information content of the corresponding prototypes.

Computation of interval-valued membership grades of matching

Before proceeding with the granular prototypes, it is instructive to briefly recall a way in which the degrees of membership are determined when dealing with numeric prototype, Considering the prototypes V_1, V_2, …, V_c and a certain pattern $\boldsymbol{x}$, we formulate the following optimization problem whose solution is a set of optimal degrees of membership (activation levels). Note that the critical point of the method is related with the determination of the distance between $\boldsymbol{x}$ and $\boldsymbol{V}_i$ (which itself is interval-valued). It becomes apparent that there is no single distance one can deem to be fully reflective of the distance between a numeric data and the granular prototype. Sound boundaries of the distance (optimistic and pessimistic) can be formed by looking at the shortest and longest distance as determined for the individual variables of the prototype. Denoting the interval-valued coordinates of $\boldsymbol{V}_i$, say $\boldsymbol{V}_{ij} = [v_{ij}^-, v_{ij}^+]$, j=1, 2, …, n, and assuming the Euclidean type of distance, we use the following self-explanatory expression, see also Figure 2. Here we have

$\|\boldsymbol{x}-\boldsymbol{V}_i\|_{\min} = (x_1 - b_{i1})^2 + (x_2 - a_{i2})^2$

$\|\boldsymbol{x}-\boldsymbol{V}_i\|_{\max} = (x_1 - a_{i1})^2 + (x_2 - b_{i2})^2$

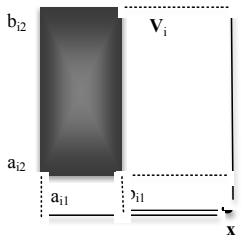


Figure 2. Computing bounds of the distance from the interval-valued prototype; shown is a 2-dimensional case (n=2)

Obviously, if x_j is included in the interval $[v_{ij}^-, v_{ij}^+]$ and this inclusion occurs for all the variables, the corresponding distance becomes equal to zero. The bounds of the distance result in the matching levels (membership grades) computed as

$$\mathbf{w}_\mathbf{i}^{*} = \frac{1}{\sum_{\mathrm{j}=1}^{\mathrm{c}} \left(\frac{\| \mathbf{x} - \mathbf{V}_\mathrm{i} \|_{\min}}{| \mathbf{x} - \mathbf{V}_\mathrm{j} \|_{\max}} \right)^{2/(m-1)}}$$

$$\mathbf{w}_\mathbf{i}^{**} = \frac{1}{\sum_{\mathrm{j}=1}^{\mathrm{c}} \left(\frac{\| \mathbf{x} - \mathbf{V}_\mathrm{i} \|_{\max}}{| \mathbf{x} - \mathbf{V}_\mathrm{j} \|_{\min}} \right)^{2/(m-1)}} \qquad (3)$$

where m is a fuzzification coefficient, m >1. Subsequently we form the bounds of the membership grades $\boldsymbol{u}_i = [u_i^-, u_i^+]$ in the following form,

$$u_i^- = \min(\boldsymbol{w}_i^{*}, \boldsymbol{w}_i^{**}) \quad u_i^+ = \max(\boldsymbol{w}_i^{*}, \boldsymbol{w}_i^{**})$$

Aggregation of information content of granular prototypes

In light of the construction discussed above, the result of classification is expressed as an aggregation of interval-valued class assignment. More specifically, we develop a vector of interval-valued class membership $\boldsymbol{\Omega}$ in the following form

$$\Omega = \sum_{i=1}^{c}{}_{\oplus} (\mathbf{u}_\mathrm{i} \otimes \mathbf{f}_\mathrm{i}) \qquad (4)$$

(the symbols shown in circles underline that the computing involves information granules, say fuzzy numbers, rather than plain numeric entities) [12]. In other words, as a result of this computing we obtain the interval-valued membership intervals in the form

$$\Omega = \begin{bmatrix} [\omega_1^-, \omega_1^+] \\ [\omega_2^-, \omega_2^+] \\ \\ [\omega_r^-, \omega_r^+] \end{bmatrix}$$

In the classification rule we distinguish between the following two situations:

(i) $\boldsymbol{x}$ is located inside a certain prototype $\mathbf{V}_{i0}$; then a vector of class membership $\boldsymbol{\Omega}$ is given as $\boldsymbol{\Omega} = \boldsymbol{f}_{i0}$ (so the class membership vector is fully inherited from the class content of the prototype),

(ii) $\boldsymbol{x}$ is positioned outside all $\boldsymbol{V}_i$s. Then the formula shown above is invoked. The interval–valued of the L-th entry of the class membership is computed as follows

$$[\boldsymbol{\omega}_\mathbf{L}^-, \boldsymbol{\omega}_\mathbf{L}^+] = [\sum_{i=1}^{c} u_i^- \mathbf{f}_{\mathbf{iL}}, \sum_{i=1}^{c} u_i^+ \mathbf{f}_{\mathbf{iL}}]$$

Interestingly, by looking at these rules, it becomes apparent that in this way we have generalized the well-known K-NN classification rule exploited in pattern recognition. As the class membership grades are interval-valued, to arrive at a single numeric value and thus identify a single class (say, L_0), a simple alternative is to find the maximal upper coordinate of $\boldsymbol{\Omega}$, namely $L_0 = \arg\max_L \boldsymbol{\omega}_L^+$

V. Experimental Studies

In this section, we evaluate the proposed granular intervals on two collections of data: synthetic data and machine learning repository datasets.

(1) Synthetic dataset: two-class two dimensional 1,800 instances of patterns have been generated randomly using normal distributions with different means and standard deviations, and then divided into the training (1,200) and testing(600) sets. The data are shown in Figure 3.

We process the data with the use of the FCM algorithm with c=5. Then the obtained prototypes are:

	x_1	x_2
v_1	0.433	0.933
v_2	0.943	0.807
v_3	-0.094	-0.183
v_4	-0.598	1.247
v_5	0.5	0.45

The interval prototypes, Figure 4, are generated for selected values of p (0.1, 0.12, 0.14,…..); we keep varying p until no intersection among intervals of the prototypes ($p_{\max}$=0.58) occurs.

In an ideal case, each interval-valued prototype would embrace data coming from the same class. The classification error ε is defined in three ways by looking at the patterns positioned (i) inside interval-valued prototypes (ε_{in}), (iii) outside intervals (ε_{out}), (iii) located anywhere (ε_g).

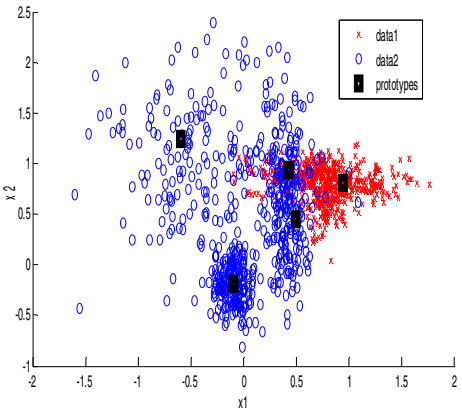


Figure 3: Synthetic training dataset

Table 1 shows the classification errors for the individual interval-valued prototypes along with the total volumes of the prototypes reported for some selected values of ps.
The classification errors inside and outside intervals for training and testing dataset are shown in Figure 5.

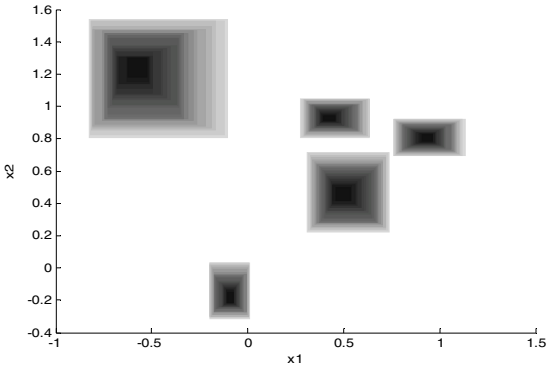
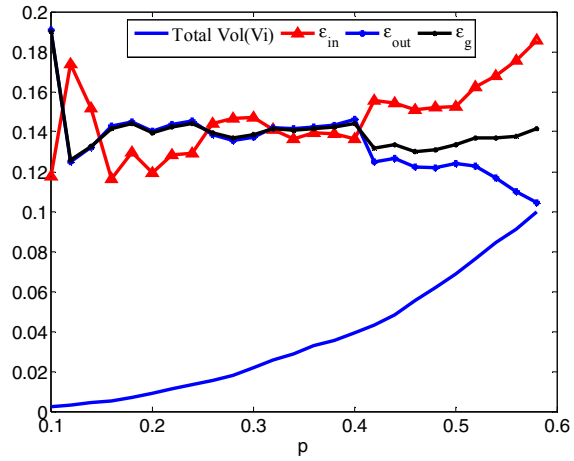


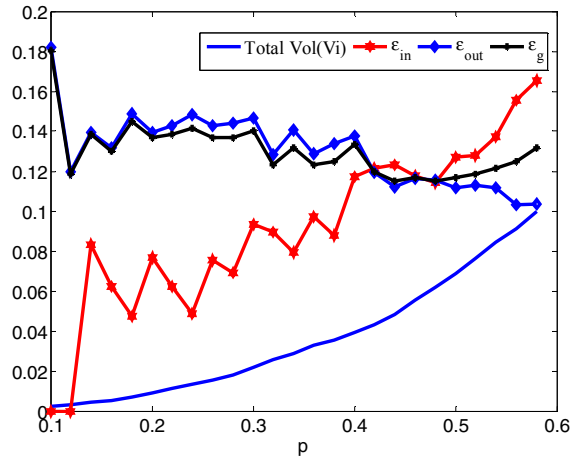
Figure 4: Interval prototypes of synthetic dataset (c=5, p=0.1, 0.12, 0.14 , …0.58).

The classification error ε_g for the training data at (p_{max} = 0.58) is 0.1425 and the minimum ε_g is 0.1292 occurs at (p_{opt} = 0.42) for the same set. The total volume of interval-valued prototypes increases with p, but this does not guarantee lower classification errors in a sense increasing the interval volumes might aggregate data from both classes. Table 2 shows ε_g for training and testing data set when considering different values of c.
In general, as the number of the prototypes increases, the classification error ε_g decreases for both p_{max} and p_{opt}, however, increasing the number of prototypes does not guarantee the decrease of the classification error as it will be shown for the machine learning datasets.

(2) Machine learning repository datasets: Four data sets have been chosen from UCI repository for machine learning datasets [13].



(a)



(b)

Figure 5: Classification error as a function of p: (a) training dataset (b) testing data set

Each dataset has been divided (60-40% split) into the training and testing sets. The description of the selected dataset is shown in Table 3.
Figure 6 displays the values of the classification errors (ε_{in}, ε_{out} and ε_g) for the Iris dataset with respect to p for c=3 and c= 4 (left vertical axis). The plots also show the total volume

of the interval prototype with respect to p (right y axis). As the Iris dataset contains three classes, we start with three interval prototypes hoping that three clusters could be enough to differentiate between the classes in the sense that each prototypes is reflective of an individual class (ε_{in}=0- at p=0.58 for c=3 and p=0.44 for c=4).

The intervals starts intersect (overlap) at p=0.7 and 0.64 where the values of ε_g are equal to 0.108 and 0.084 for c=3 and c=4, respectively. However, the minimum value of ε occurred at p_{opt}= 0.34 (c=3) and 0.62 (c=4).

Table 4 shows the classification error ε_g for machine learning datasets at p_{max} and p_{opt} along with the values of c. In most of the cases, the minimum value of ε_g does not occur at p_{max}, Also, considering more prototypes in the classification process might improve the process to a certain extent, however it depends on the nature of the data.

p		V_1	V_2	V_3	V_4	V_5	ε_{in}	ε_g	Total Vol (Vi)
0.1	$\boldsymbol{f_i}$	0.25	0	0	0	0	0.07	0.22	0.002
	Vol($\boldsymbol{V_i}$)	0	0	0	0.001	0.001			
0.3	$\boldsymbol{f_i}$	0.424	0	0	0	0.24	0.16	0.14	0.022
	Vol($\boldsymbol{V_i}$)	0.002	0.002	0.002	0.011	0.005			
0.4	$\boldsymbol{f_i}$	0.333	0	0	0	0.239	0.13	0.14	0.04
	Vol($\boldsymbol{V_i}$)	0.003	0.004	0.003	0.02	0.01			
0.58	$\boldsymbol{f_i}$	0.326	0.007	0	0.041	0.437	0.19	0.13	0.1
	Vol($\boldsymbol{V_i}$)	0.009	0.009	0.008	0.053	0.021			

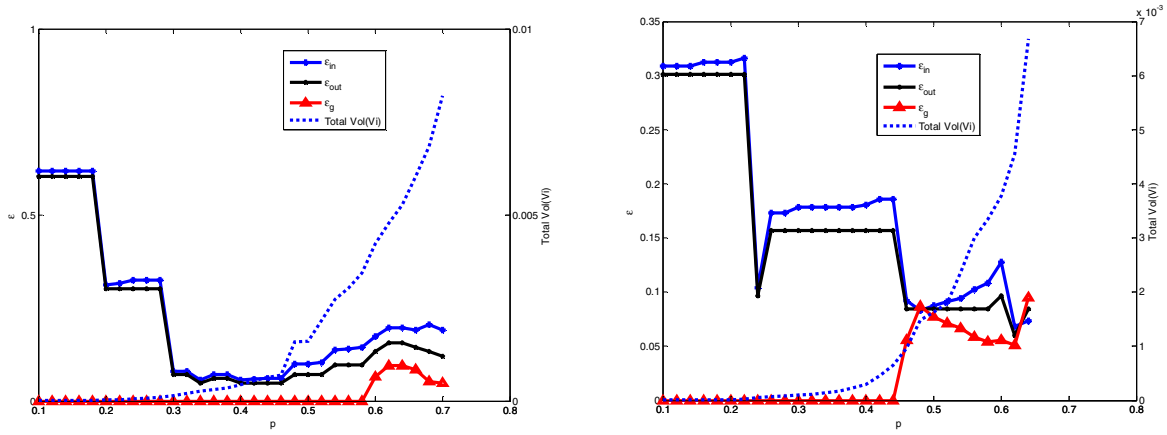
Table 1: Interval-valued prototype for the training data: $\boldsymbol{f_i}$, Vol($\boldsymbol{V_i}$), ε and Total Vol($\boldsymbol{V_i}$)

c	p_{max}	Training (p_{max})	Testing (p_{max})	p_{opt}	Training (p_{opt})	Testing (p_{opt})
2	0.60	0.167	0.167	0.52	0.159	0.168
4	0.64	0.129	0.145	0.16	0.112	0.112
7	0.52	0.133	0.143	0.14	0.108	0.108
9	0.46	0.128	0.128	0.14	0.105	0.105
11	0.48	0.127	0.128	0.20	0.123	0.130
15	0.54	0.098	0.098	0.54	0.098	0.098

Table 2: classification error ε_g for different value of c for synthetic dataset.

Dataset	Number of instances	Number of variables	Number of classes	Training set	Testing set
Iris	150	4	3	75	75
Page Blocks Classification	5473	10	5	2736	2737
Breast Cancer Wisconsin (Diagnostic)	569	30	2	284	285
Seeds	210	7	3	105	105

Table 3: Description of the selected machine learning datasets.



(a) (b)

Figure 6: Classification error for the Iris data --training data: (a) c=3 (b) c=4

Dataset	c	p_{max}	Training set	Testing set	p_{opt}	Training set	Testing set
Iris	3	0.7	0.108	0.119	0.34	0.048	0.090
	4	0.64	0.084	0.149	0.62	0.060	0.134
	7	0.7	0.108	0.119	0.34	0.048	0.090
Page Blocks Classification	5	0.76	0.093	0.095	0.74	0.093	0.095
	7	0.64	0.093	0.095	0.54	0.093	0.095
	10	0.52	0.093	0.096	0.52	0.093	0.096
Breast Cancer Wisconsin (Diagnostic)	2	0.82	0.064	0.070	0.82	0.064	0.070
	4	0.68	0.121	0.137	0.68	0.121	0.137
	8	0.60	0.067	0.078	0.60	0.067	0.078
Seeds	3	0.66	0.106	0.106	0.66	0.106	0.128
	5	0.58	0.069	0.074	0.58	0.069	0.074
	9	0.46	0.106	0.106	0.46	0.106	0.106

Table 4: Classification error ε_g obtained for selected machine learning datasets versus different number of prototypes.

Dataset	Prototypes intervals (c, p)	Other methods reported in the literature
Iris	0.090 (7,0.48)	0.0447 [14]
		0.033 +- 2.3 [15]
		0.0133 [16]
Page Blocks Classification	0.095 (7, 0.54)	0.0305 [17]
		0.0272 [18]
		0.1446 [19]
Breast Cancer Wisconsin (Diagnostic)	0.078 (8, 0.60)	0.0228 [20]
		0.0190+- 0.005 [21]
		0.038 +-0.014 [22]
Seeds	0.074 (5, 0.58)	0.04 [23]
		0.0810 [24]

Table 5: Comparison of classification error rates ε_g for selected machine learning data

.

The granular classifier competes with other methods reported in the literature as shown in Table 5. It is worth mentioning that the value of p plays an important role in the overall process. In the current experiments, the value of this parameter is fixed for all prototypes and for all input variables. In other words, there could be different volumes of interval prototypes for the same value of prototypes depending on the distributions of data around the prototypes.

VI. Conclusions

The idea of building granular prototypes viewed as sound general and non-numeric representatives of data and being inherently associated with classification problems helps re-formulate the essence of the classifiers and make sense of the nature of classification results. Granular prototypes can be treated as generalized landmarks in the feature space while the way in which the determination of class assignment is completed generalizes an idea of condensed K-NN (K- nearest neighborhood) classification rule with the objects exhibiting well-defined information content (class characterization).

As the study brings forward some initial concepts along with algorithmic considerations, there are several clearly delineated directions worth pursuing. First, the construction of granular prototypes involves setting a value of the crucial parameter (p), which is common across all granules and all variables. In this sense, a growth of information granules is the same and is limited by the non-overlap requirement. Second, our focus was on interval constructs (which was predominantly motivated by the readability of the resulting concept). We are, however, well poised to consider more advanced formalisms of information granules (fuzzy sets, rough sets, etc.). Also, optimizing the value of p along each variable (feature) would lead to higher classification rates reported for the patterns located within the intervals of the prototypes.

Acknowledgments

This study was funded by King Abdulaziz University (KAU), under grant No. (4-135-1434/HiCi). The authors, therefore, acknowledge technical and financial support of KAU.

References

[1] A. Bargiela, and W. Pedrycz, *Granular Computing: An Introduction*. Kluwer Academic Publishers, Dordrecht, (2003).

[2] W. Pedrycz, *Granular Computing: Analysis and Design of Intelligent Systems*, CRC Press/Francis Taylor, Boca Raton, (2013).
[3] L.A. Zadeh, "Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic". *Fuzzy Sets and Systems*, 90, 1997, pp. 111-117.
[4] L.A Zadeh, "From computing with numbers to computing with words-from manipulation of measurements to manipulation of perceptions". *IEEE Trans. on Circuits and Systems*, 45, 1999, pp. 105-119
[5] L.A. Zadeh, "A note on Z-numbers", *Information Sciences*, 181, 2011, pp. 2923-2932.
[6] L. Billard, and E. Diday, *Symbolic Data Analysis*. J. Wiley, Chichester, 2006.
[7] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York, Plenum, 1981.
[8] W. Pedrycz, *Knowledge-Based Clustering: From Data to Information Granules*. J. Wiley, Hoboken, 2005.
[9] W. Pedrycz, and J. Valente de Oliveira, "A development of fuzzy encoding and decoding through fuzzy clustering", *IEEE Trans. Instrum. Meas*, vol. 57, no. 4, 2008, pp. 829–837.
[10] R. Xu, and D. Wunsch, "Survey of clustering algorithms", *IEEE Transactions on Neural Networks 16*, 3, 2005, pp. 645–678.
[11] W. Pedrycz, A. Bargiela, "An optimization of allocation of information granularity in the interpretation of data structures: toward granular fuzzy clustering", *IEEE Trans on Systems, Man, and Cybernetics, Part B*, 42, 2012, pp. 582-590.
[12] W. Pedrycz, and F. Gomide, *Fuzzy Systems Engineering: Toward Human-Centric Computing.* John Wiley, Hoboken, NJ, 2007.
[13] Machine Learning Repository : http://archive.ics.uci.edu/ml/
[14] E. Bauer, and R. Kohavi. "An empirical comparison of voting classification algorithms: Bagging, boosting, and variants", *Machine Learning*, 36, 1999, pp. 105–139.
[15] H. Cevikalpa, and B. Triggs, "Hyperdisk based large margin classifier", Pattern Recognition, Volume 46, Issue 6, 2013, pp. 1523–1531.
[16] F. J. de Souza, M. M. B. R. Vellasco, and M. A. C. Pacheco, "Hierarchical neuro-fuzzy quad tree models", *Fuzzy Sets Syst.*, vol. 130/2, 2002, pp. 189–205.
[17] J.R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA, 1992.
[18] S. Eschrich, N. V. Chawla and L. O. Hall, "Generalization Methods in Bioinformatics". In *2002* Proc Data Mining in Bioinformatics. pp.26-32
[19] M. Adil, Bagirov, "Max–min separability", *Optimization Methods and Software*, Vol. 20, Iss. 2-3, 2005, pp.277-296.
[20] G. Salama, M. Abdelhalim, and M. Zeid, "Breast Cancer Diagnosis on Three Different Datasets Using Multi-Classifiers", *International Journal of Computer and Information Technology*, Vol. 01, Issue 01, 2012, pp. 36-43.
[21] H. A. Abbas, "An evolutionary artificial neural networks approach for breast cancer diagnosis", *Journal of Artificial Intelligence in Medicine*, Volume 25 Issue 3, 2002, Pages 265-281.
[22] B. Moghaddam, and G. Shakhnarovich, *Boosted Dyadic Kernel Discriminants* (Advances in Neural Information Processing Systems 15, (2002), MIT Press, pp. 745-752.
[23] M. Charytanowicz, J. Niewczas, P. Kulczycki, P.A. Kowalski, S. Lukasik, and S. Zak, "A Complete Gradient Clustering Algorithm for Features Analysis of X-ray Images", *Information Technologies in Biomedicine*, E. Pietka, J. Kawa (eds.), Springer-Verlag, Berlin-Heidelberg, 2010, pp. 15-24
[24] O. Hryniewicz, "Statistical and Possibilistic Methodology for the Evaluation of Classification Algorithms", *Communications in Computer and Information Science*, Volume 303, 2013, pp 255-269.