Int. J. Fuzzy Syst. (2020) 22(6):1731–1747
https://doi.org/10.1007/s40815-020-00903-z

Data Description Through Information Granules: A Multiview Perspective

Abdullah Balamash[3,4] · Witold Pedrycz[1,2,3] · Rami Al-Hmouz[3] · Ali Morfeq[3]

Received: 17 October 2019 / Revised: 25 May 2020 / Accepted: 9 June 2020 / Published online: 27 July 2020

Abstract In light of the remarkable diversity of data, arises an interesting and challenging problem of their description and concise interpretation. In a nutshell, in the proposed description pursued in this study, we consider a framework of information granules. The study develops a general scheme composed of two functional phases: (i) clustering data and features forming segments of original data and delivering a meaningful partition of data, and (ii) development of information granules. In both phases, we discuss a suite of performance indexes quantifying the quality of segments of data and the resulting information granules. Along this line, discussed are collections of information granules and their mutual relationships. A series of publicly available data sets is used in the experiments—their granular signature is quantified, and the quality of these findings is analyzed.

Keywords Information granules · Multiview perspective · Clustering · Reconstruction · Classification · Prediction · Granular signature of data

✉ Abdullah Balamash
asbalamesh@kau.edu.sa

1 Department of Electrical & Computer Engineering, University of Alberta, Edmonton, AB T6R 2V4, Canada

2 Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

3 Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia

4 Center of Excellence in Intelligent Engineering Systems (CEIES), King Abdulaziz University, Jeddah 21589, Saudi Arabia

1 Introduction

Data analysis and data analytics, in general, are inherently aimed at revealing and description of interpretable and stable relationships among variables as well as quantifying their changes over time and space. Along with large volumes of data and their diversity, comes a genuine need to develop a flexible, user-centric and computationally efficient environment producing meaningful results.

The key research hypothesis is that in the realization of the above stated agenda of data analytics, the concepts of a *multiview* perspective [1–4] of data with the use of information granules play a pivotal role both at the methodological as well as algorithmic level of ensuing constructs. The formation and engagement of the multiview organization of processing of data contributes in a tangible way to the efficient way of solving of a spectrum of tasks of data analysis, especially facilitating a thorough user-centric interpretation of results and producing readable yet fully legitimate outcomes supported by the existing experimental evidence. The varying (adjustable) perspective delivered by information granules helps establish a sound tradeoff between the representation capabilities of various views at the data and the efficiency of fundamental categories of tasks of data science such as association analysis, classification, prediction, link analysis and others.

Another important research hypothesis is that by engaging the multiview perspective of data analytics of the same data, we establish a coherent and holistic view of the data and ensuing models under consideration. Data are represented through a collection of information granules. The diversity of the constructed granules manifests itself by the fact that information granules are built based on subsets of data and subsets of features while the quality of granules is being assessed.

The term multiview data analysis has been used in the literature in the past, however this term comes with a different meaning. The study reported in [5] offers an interesting view focused on feature selection. In our case, the multiview character of information granules is concerned with the perspective established with regard to mutual organization involving some sections of the data and subsets of features. Furthermore, the multiview is formed in the conceptually appealing and computationally sound setting of information granules.

A number of well-focused research aims of this study are presented (subsequently leading to the formulation of a coherent and comprehensive methodological framework of the investigation):

i. A multiview formation of data subspaces leading to dimensionality reduction, enhanced readability (interpretability) of the data and increased efficiency of ensuing analysis (such as e.g., prediction, association analysis, or classification). The varying (adjustable) levels of detail captured by the individual views (perspectives) are helpful in reducing computing overhead of individual optimization tasks.
ii. The multiview facets built for the data are also concerned with granulation of the feature space, therefore leading to so-called *meta-features* (viz. collections of features, which exhibit some semantics and offer a view at the data at the higher level of abstraction). Both of these categories of views outlined in (i)–(ii) give rise to information granules of meta-features and information granules established in the joint data-feature space. Each view (perspective) gives rise to its focused perception of the same data and ensuing results produced in this setting.
iii. Formation of optimization criteria quantifying the quality and practical relevance of the multiview perspective at the data. The essential criteria fall under the umbrella of representation capabilities of the data (which are commonly linked to the inevitable compression error) and the relevance of the established cognitive perspective in solving main categories of data analysis problems. An important and intriguing problem comes with a way on how to balance these two requirements and cope with their conflicting nature (higher representation capabilities do not directly translate into more efficient and computationally sound performance of data analysis).
iv. While numeric prototypes are sound initial descriptors of segments of data and features, they are elevated to granular counterparts, which in turn offer better abstract and holistic descriptors of data.

The ultimate objective is to derive structural information [6] in the data and feature (attribute) space and construct information granules on combinations of subsets of data and features. Their quality is evaluated in view of various criteria depending on further use of information granules in system modeling (classification and prediction) and data representation. The constructed information granules are ranked with respect to the pertinent performance criteria (either reconstruction-based, prediction-oriented, or classification-based). An overall scheme of processing underlying a way of moving from data to information granules is displayed in Fig. 1. Here, the main phases are highlighted along with numeric and granular descriptors. The overall scheme outlined here entails also a significant level of originality as the comprehensive concept and its algorithmic environment have not been investigated.

The data-feature segmentation can be concisely captured in the following way:

$$(D1, F1)\ldots(D_i, F_j), \quad i = 1, 2, \ldots, c; \quad j = 1, 2, \ldots, r, \tag{1}$$

where the data and feature sets are exhaustive and mutually exclusive, namely

$$\begin{matrix} \bigcup_{i=1}^{c} D_i = \boldsymbol{D} D_i \cap D_j = \emptyset, i \neq j \\ \bigcup_{i=1}^{r} F_i = \boldsymbol{F} F_i \cap F_j = \emptyset, i \neq j \end{matrix}, \tag{2}$$

where c and r are the number of segments present in data and feature space.

The study offers some original insights into the problem of data description that have not been studied in the past: (i) the development of information granules in the data and feature space delivers a new focused view at the essence of the overall data set; (ii) the ensuing information granules built on a basis of numeric prototypes establish a so-called granular blueprint of data and help focus on the essence of

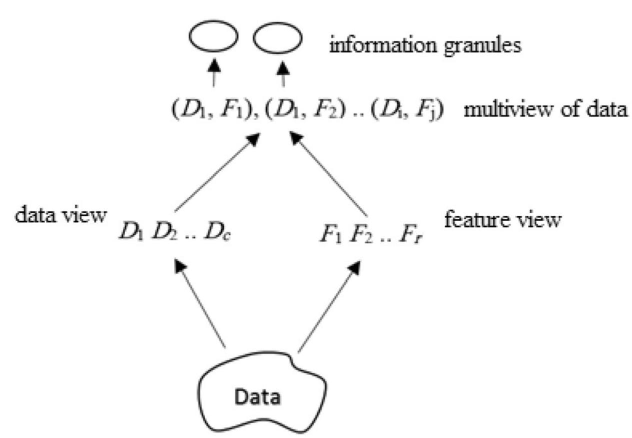

Fig. 1 Overall processing scheme: from data to information granules; building data and feature views

the relationships present there, and (iii) the construction of classifiers and predictors at the granular level by engaging information granules as a backbone of such constructs.

To systematically organize the presentation on the concepts and their construction, the paper is structured as follows. Section 2 elaborates on the development of subsets of data and features (data views) with the use of fuzzy clustering, Fuzzy C-Means (FCM) [7], being more specific. Subsequently, in Sect. 3, the characterization of these views is offered through several performance indexes, say a reconstruction error, classification content and prediction content of numeric representatives of data views. Section 4 is devoted to the construction of information granules through the principle of justifiable granularity. Information granules form a blueprint of classifiers and predictors; these topics are covered in Sect. 5. Experiments using publicly available data are presented in Sect. 6. Conclusions and directions of future research are included in Sect. 7.

2 Development of Subsets of Data (Clusters) Through Clustering Completed in Data Space and Feature Space

Information granules are commonly constructed with the help of clustering techniques [8] regarded as a prerequisite design vehicle. Clustering is regarded as a sound departure point of further constructs. Here, we consider the Fuzzy C-Means (FCM) algorithm as a representative of vehicle of clustering. While the FCM is commonly used to cluster data, it can be also considered to cluster features, viz. build a collection of features. In what follows, we consider patterns (data) $\boldsymbol{x}_1, \boldsymbol{x}_2,\ldots, \boldsymbol{x}_N$ expressed in the n-dimensional space of real numbers $\boldsymbol{R}^n$. Recall that clustering realized by the FCM algorithm returns a collection of prototypes and a partition matrix. The number of clusters in the data space is set to c, and the number of clusters in the feature space is set to r. In what follows, we recall the essence of building data segments and feature segments.

2.1 Clustering in the Data Space

The FCM is guided by the following well-known objective function:

$$Q = \sum_{i=1}^{c} \sum\nolimits_{k=1}^{N} u_{ik}^{m} ||\boldsymbol{x}_k - \boldsymbol{v}_i||^2, \tag{3}$$

where c stands for the number of clusters, m is a fuzzification coefficient ($m > 1$) and ||.|| is a weighted Euclidean distance [9, 10], namely $||\boldsymbol{a} - \boldsymbol{b}||^2 = \sum_{j=1}^{n} \frac{(a_j - b_j)^2}{\sigma_j^2}$, $\dim(\boldsymbol{a}) = n$ with the weights being the standard deviations of the corresponding variables. The optimization (viz. partitioning the data) carried out in the data space is realized iteratively: one starts with a randomly initialized partition matrix U and then updates the parameters to be optimized, viz. the partition matrix and the collection of the prototypes $\boldsymbol{v}_1, \boldsymbol{v}_2,\ldots, \boldsymbol{v}_c$.

Note that the partition matrix is fuzzy, viz. the entries assume values in between 0 and 1. In other words, fuzzy sets formed by the FCM embrace almost all data but with some degrees of membership [11]. To form the constructed subsets of data, we make them Boolean (two-valued) by admitting those data which belong to the ith cluster to the highest extent, viz. $\boldsymbol{D}_i = \{\boldsymbol{x}_k | \, u_{ik} = \max_{j=1,2,\ldots,c} u_{jk}\}$.

In conclusion, one can regard the prototypes $\boldsymbol{v}_i$ as the concise numeric descriptors of $\boldsymbol{D}_i$. The prototypes are just a manifestation of the data composing the clusters and as such are the most meaningful outcomes of fuzzy clustering to be used in further investigations.

2.2 Clustering in the Feature Space

When it comes to revealing structure in feature space, we reformulate the problem and look at the objects that are subject to clustering. Let us organize the original data into vectors positioned in $\boldsymbol{R}^N$, namely $z_1, z_2,\ldots, z_n$, where $z_j = [x_{j1}\ x_{j2}\ \ldots\ x_{jN}]$, $j = 1, 2,\ldots, n$.

The objective function guiding the process of clustering of features (thus building subsets of features) is expressed as follows:

$$Q = \sum_{i=1}^{c} \sum_{j=1}^{n} g_{ij}^{m} ||z_j - \boldsymbol{t}_i||^2. \tag{4}$$

Here, the distance is expressed as follows:

$$||z_j - \boldsymbol{t}_i||^2 = \sum_{k=1}^{N} (z_{jk} - t_{ik})^2. \tag{5}$$

The partition matrix G conveys crucial information about the subsets of features forming so-called met features. The features belonging to the jth cluster to the highest extent are denoted by $\boldsymbol{F}_j$. In virtue of identifying elements of the partition matrix F, the subsets $\boldsymbol{F}_j$, $\boldsymbol{F}_l$, etc., are mutually disjoint.

In summary, the results of clustering completed in the data and feature space come as data sets ad feature sets. We form all possible combinations of subsets produced by the clustering completed in the data space and feature space. For instance, the subset $(\boldsymbol{D}_i, \boldsymbol{F}_j)$ describes the data belonging to $\boldsymbol{D}_i$ and having features belonging to $\boldsymbol{F}_j$. Having c and r clusters in the data and feature space, respectively, we have cr subsets (segments) in the Cartesian product of these two spaces. In what follows, we evaluate the quality of such subsets; by computing the pertinent measures one can order the subsets and evaluate their distribution.

3 Characterization of Data Views ($\boldsymbol{D}_i$, $\boldsymbol{F}_j$)

The performance of each cluster (data view) can be evaluated in various ways. Depending upon applications, there are several main indexes to be considered: (i) reconstruction error, (ii) classification content, and (iii) prediction capabilities.

First, we elaborate on the reconstruction criterion. In total there are cr information granules (clusters) each associated with the reconstruction error. The results obtained for the corresponding clusters are arranged in a matrix form organizing results for all combinations of $\boldsymbol{D}_i$ and $\boldsymbol{F}_j$.

3.1 Reconstruction Error

Denote the reconstruction error produced for ($\boldsymbol{D}_i$, $\boldsymbol{F}_j$) by V_{ij}, i = 1, 2.., c; j = 1, 2,…, r.

This error expresses the representation capabilities of the prototype $\boldsymbol{v}_{\mathrm{ij}}$ associated with ($\boldsymbol{D}_i$, $\boldsymbol{F}_j$) by computing the following expression V_{ij}:

$$V_{ij} = \frac{1}{\mathrm{card}(\boldsymbol{D}_i)} \frac{1}{\mathrm{card}(\boldsymbol{F}_j)} \sum_{\substack{k=1 \\ \boldsymbol{x}_k \in D_i}}^{N} ||x_k - \boldsymbol{v}_{ij}||^2_{F_j}. \tag{6}$$

The same as in clustering algorithms, the distance ||.|| is the weighted Euclidean involving the standard deviation of the variables; let us emphasize that the calculations are completed for features forming $\boldsymbol{F}_j$. The prototype $\boldsymbol{v}_{\mathrm{ij}}$ standing in (6) is computed as follows:

$$v_{ij,l} = \frac{1}{N_i} \sum_{x_k \in D_i} x_k, \tag{7}$$

where l runs through indexes of features forming $\boldsymbol{F}_j$. Obviously, the coordinates of $\boldsymbol{v}_{\mathrm{ij}}$ stand for those variables which form $\boldsymbol{F}_j$. We organize the values of the reconstruction error into a c by r matrix form containing the values of the V_{ij}. Furthermore, the values of V_{ij} can be arranged in an increasing order by ranking the relevance of numeric descriptors by starting from the most relevant ones (viz. with the smallest values of this criterion).

Furthermore, depending on the nature of the data under consideration, the quality of information granules can be assessed by viewing their discriminatory and predictive content (abilities).

3.2 Classification Content of Information Granule

When dealing with classification problem, one determines a class content of information granule. Consider that in the classification problem we encounter t classes $\omega_1, \omega_2,\ldots, \omega_t$. The quality of the information granule (cluster) formed by ($\boldsymbol{D}_i$, $\boldsymbol{F}_j$) is assessed by looking at the distribution of data in ($\boldsymbol{D}_i$, $\boldsymbol{F}_j$) across different classes. With regard to the matrix of segments of data and features, the result is the same across the columns in the given row.

Then we calculate the probability of classes present in this information granule $\boldsymbol{p}_i$= [p_{i1} p_{i2} …, p_{it}], i = 1,2,…,c. The less homogenous the information granule is, the higher its vagueness becomes. The quantification is realized by means of the entropy [12] measure defined as follows:

$$h(u) = \begin{cases} 2u, u \in [0, 1/2] \\ 2(1-u), u \in [1/2, 1] \end{cases}. \tag{8}$$

The vagueness of the ith granule is expressed as follows:

$$C_i = \sum_{l=1}^{t} h(p_{il}). \tag{9}$$

3.3 Predictive Content of Information Granules

Comparing the performance indexes is completed by looking at the diversity of output data falling within the bounds of the information granules. The diversity is quantified by means of the variance of the output variable of data falling within the bounds of ($\boldsymbol{D}_i$, $\boldsymbol{F}_j$). In more detail, recall that the data come in the form ($\boldsymbol{x}_k$, y_k), where y_k is the output variable. We calculate the variance

$$\sigma^2_{iy} = \frac{1}{N_i - 1} \sum_{(x_k, y_k) \in D_i} (y_k - \bar{y})^2, \tag{10}$$

where

$$\bar{y}_i = \frac{1}{N_i} \sum_{(x_k, y_k) \in D_i} y_k \tag{11}$$

and

$$R_i = \sigma^2_{iy} \tag{12}$$

i = 1, 2,…,c.

4 Construction of Information Granules

The data subsets (segments) ($\boldsymbol{D}_i$, $\boldsymbol{F}_j$) embracing some data and formed over a certain collection of features give rise to information granules. The granules are built with the use of the principle of justifiable granularity [13–16].

In a nutshell, this principle produces an information granule in such a way it meets the requirements of coverage and specificity whose product is maximized; see Fig. 2.

The design of information granule is realized in such a way that the granule is (i) experimentally justifiable and (ii) is semantically sound. The experimental justification means that there is enough data embraced (contained) in the constructed granule making its existence legitimate in terms of the experimental data (hence the aspect of experimental justification). The semantic soundness states that the granule has to exhibit some interpretation capabilities and its precision needs to be sufficient enough. The coverage is expressed in the following way:

$$\operatorname{cov}(G_{ij}) = \frac{1}{N_{ij}} \operatorname{card}\left\{ \boldsymbol{x}_k \in (D_i, F_j) \,|\, ||\boldsymbol{x}_k - \boldsymbol{v}_{ij}||_{\boldsymbol{F}_j} \le n_j \rho_{ij}^2 \right\}. \tag{13}$$

The interpretation of the coverage criterion requires some attention. This criterion quantifies the amount of experimental evidence behind the constructed information granule. In more detail, we count the number of data whose distance computed over all features present in $\boldsymbol{F}_j$, say $||x_k - v_{ij}||_{F_j} = \sum_{l=1}^{n_j} \frac{(x_{kl} - v_{ij,l})^2}{\sigma_{il}^2}$, with σ_{il} being equal to the standard deviation of data residing within the corresponding segment of the data) equal or smaller than $n_j \rho_{ij}^2$ the threshold implied by the radius of the constructed information granule ρ_{ij}^2.

The specificity regarded as a measure of precision is given as follows:

$$\operatorname{sp}(G_{ij}) = 1 - \rho_{ij}. \tag{14}$$

Note again that the highest specificity is achieved for the radius set to zero. However, at this case, the coverage is practically equal to zero. On the other hand, the highest coverage implies a zero value of specificity. The increase of coverage implies the decrease of specificity and vice versa. If these conflicting criteria have to be optimized, one has to proceed with a bi-criteria optimization or formulate the problem as a scalar optimization by taking an aggregate of the criteria. The product of coverage and specificity could serve as a viable alternative here.

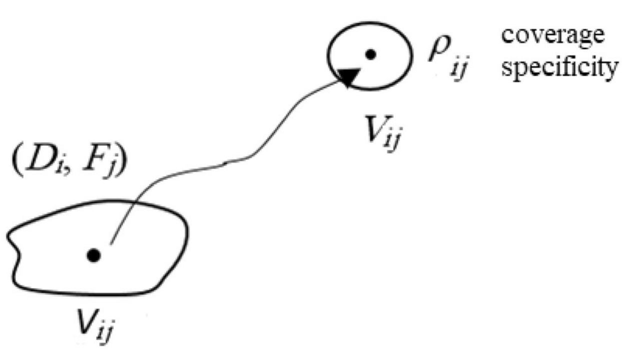

Fig. 2 From data segment to information granules

An information granule V_{ij} associated with $(\boldsymbol{D}_i, \boldsymbol{F}_j)$ is the pair $(\boldsymbol{v}_{ij}, \rho_{ij})$, where the radius ρ_{ij} is optimized by considering the optimization problem

$$\rho_{ij,\text{opt}} = \arg \operatorname{Max}_{\rho_{ij} \in [0,1]} \left[\operatorname{cov}(G_{ij}) \operatorname{sp}(G_{ij}) \right]. \tag{15}$$

The higher the value of the optimized product of coverage and specificity, the more suitable (relevant) the constructed information granule becomes. Proceeding with the constructed information granules (after maximization of (13)), we can conveniently display them in the coverage-specificity plane; see Fig. 3. The location of information granules helps identify the best of them in terms of the specificity and coverage criteria.

5 Granular Predictors and Classifiers

The collection of information granules $G_{ij} = (\boldsymbol{v}_{ij}, \rho_{ij})$, $i = 1, 2, \ldots, c$; $j = 1, 2 \ldots, r$, forming the concise description of data are regarded as building modules (blueprint) so that they can give rise to granular predictors and classifiers. We briefly outline the essence of the underlying architecture; noticeable is a role of the granules as a skeleton of the construct.

5.1 Predictors

Let us consider that for each information granule, there is a numeric representative of the output variable. Any input $\boldsymbol{x}$ is matched vis-a-vis the individual information granules giving rise to the corresponding activation (matching) levels $u_1, u_2, \ldots, u_{cr}$:

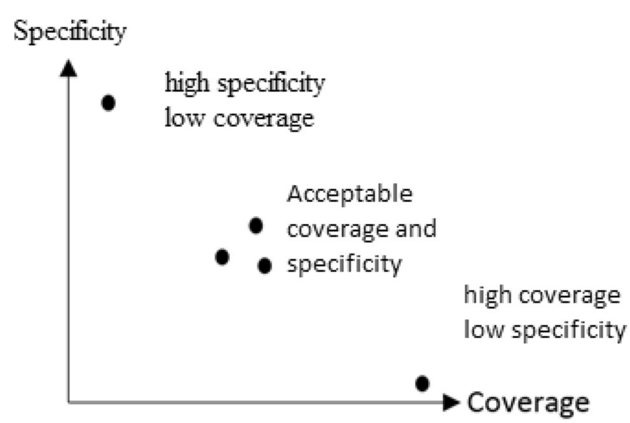

Fig. 3 Characterization of information granule in the coverage-specificity plane

$$u_{ij} = \frac{1}{\sum_{\substack{i_1=1\\ j_1=1}}^{c,r} \left(\frac{\|\mathbf{x}-\mathbf{v}_{ij}\|_{F_j}}{\|\mathbf{x}-\mathbf{v}_{i_1 j_1}\|_{F_j}} \right)^{2/(m-1)}}. \tag{16}$$

The prediction result is computed by taking a linear combination of the numeric representatives of the individual information granules and their radii, namely

$$\hat{y} = \sum_{i=1}^{c} \bar{u}_i \bar{y}_i \qquad \hat{\rho} = \sum_{i=1}^{c} \bar{\rho}_i \bar{y}_i, \tag{17}$$

where

$$\bar{u}_i = \sum_{j=1}^{r} u_{ij}, \qquad \bar{\rho}_i = \sum_{j=1}^{r} \rho_{ij}. \tag{18}$$

Thus, the prediction result arises as information granule $\hat{Y} = (\hat{y}, \hat{\rho})$.

5.2 Classifiers

As presented so far, each information granule comes with a vector of probability of classes $\boldsymbol{p}_i$. $\boldsymbol{p}_i = [p_{i1}\ p_{i2}\ \ldots p_{it}]$, $i = 1,2,\ldots,c$. They are coming as a result of counting the number of patterns belonging to the individual classes [17]. More specifically, denoting by Ni the number of data contained in $\boldsymbol{D}_i$, $\mathbf{n}_{i1}$, $n_{i2},\ldots, n_{it}$ are the counts of number of data belonging to the corresponding classes. The vector $\boldsymbol{p}_i$ is composed of the ratios

$$p_i = \left[\frac{n_{i1}}{N_i} \frac{n_{i2}}{N_i} \cdots \frac{n_{it}}{N_i} \right]. \tag{19}$$

The process of class assignment proceeds in a similar way as discussed in case of predictors. The final class membership $\boldsymbol{p}$ is computed in the following way:

$$p = \sum_{i=1}^{c} \bar{u}_i p_i. \tag{20}$$

Using the maximum rule, one selects this class i_0 for which the coordinate of $\boldsymbol{p}$ attains the highest value, viz. $i_0 = \arg\mathrm{Max}_{i=1,2,\ldots,p} p_i$.

In other words, i_0 is the index of the largest coordinate of the vector $\boldsymbol{p}$.

5.3 Illustrative Example

In this example, we assume a synthetic data of six data points of four features. The first three data points are from a certain normal distribution (classification class 1), and the next three from another normal distribution (classification class 2).

$$X = \begin{bmatrix} -0.3748 & -0.6411 & 2.8948 & 0.8533 \\ 0.9164 & 1.0290 & 2.1972 & 4.0396 \\ 1.0432 & 1.3009 & -1.4089 & 2.4230 \\ 4.8278 & 2.9201 & 2.6086 & 6.5614 \\ 5.2599 & 2.2045 & 5.6848 & 7.7025 \\ 5.3587 & 1.3298 & 2.2939 & 9.2319 \end{bmatrix}$$

X is clustered into $c = 2$ data clusters, and $r = 2$ feature clusters as follows (Fig. 4).

Accordingly, we have four information granules: (D_1, F_1), (D_1, F_2), (D_2, F_1), and, (D_2, F_2). The prototype of each information granule is computed by averaging all its data points. For example, the computation of v_{11} is illustrated in Fig. 5.

Now using Eq. (16), we compute the membership matrix u_{ij}. Figure 6 illustrates the semantics of u_{ij}.

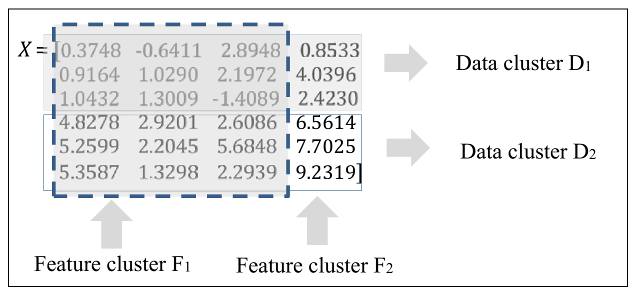

Fig. 4 Clustering the synthetic data into ($c = 2$) data clusters and ($r = 2$) feature clusters

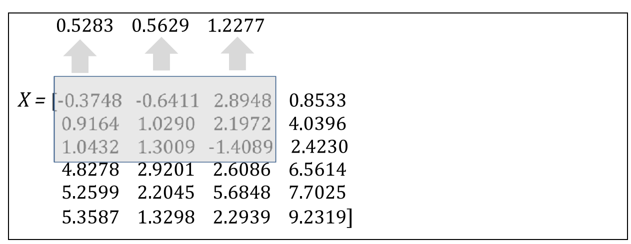

Fig. 5 Computation of prototype v_{ij} for information granule (D_i, F_j)

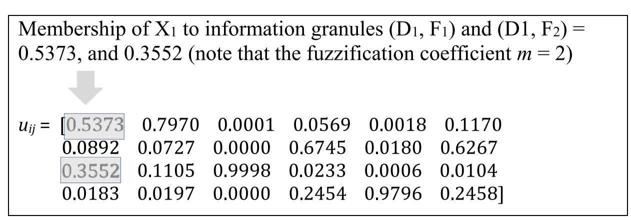

Fig. 6 The membership values u_{ij}

Membership of X_1 to data granule D_1 = 0.5373 + 0.3552 = 0.8925

$$u_{ij} = \begin{bmatrix} \boxed{0.5373} & 0.7970 & 0.0001 & 0.0569 & 0.0018 & 0.1170 \\ 0.0892 & 0.0727 & 0.0000 & 0.6745 & 0.0180 & 0.6267 \\ \boxed{0.3552} & 0.1105 & 0.9998 & 0.0233 & 0.0006 & 0.0104 \\ 0.0183 & 0.0197 & 0.0000 & 0.2454 & 0.9796 & 0.2458 \end{bmatrix}$$

Membership of X_4 to data granule D_2

$$\bar{u}_i = \begin{bmatrix} 0.8925 & 0.9076 & 1.0000 & 0.0802 & 0.0024 & 0.1274 \\ 0.1075 & 0.0924 & 0.0000 & \boxed{0.9198} & 0.9976 & 0.8726 \end{bmatrix}$$

Fig. 7 The membership values $\bar{u}_i$

Membership of X_4 to data clusters D_1 and D_2 is 0.0802 and 0.9198

$$\bar{u}_i = \begin{bmatrix} 0.8925 & 0.9076 & 1.0000 & \boxed{0.0802} & 0.0024 & 0.1274 \\ 0.1075 & 0.0924 & 0.0000 & \boxed{0.9198} & 0.9976 & 0.8726 \end{bmatrix}$$

$$p_i = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The maximum value, which mean that X_4 belongs to class 2

$$p = [0.0802\ 0.9198] * \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = [0.0802\quad 0.9198]$$

Fig. 8 Classification class prediction

Using Eq. (18), we compute the membership through data clusters ($\bar{u}_i$). For example, $\bar{u}_i$ is computed as shown in Fig. 7.

Using Eq. (19), we compute (p_i), the ratio of each classification class in each data cluster D_i. Then, using Eq. (20), the assigned class to a certain data point X_i is computed as shown in Fig. 8.

6 Experimental Studies

In this section, we elaborate on the development of information granules and their quality. Both classification and regression type of data are considered; see Table 1.

We proceed with the clustering algorithm in the data space or feature space as described in Sect. 2. The number of clusters in the data space is c while for clustering features we consider r clusters. The clustering results are transformed to the binary version. The values of these numbers are selected based on the behavior of the objective function versus the varying values of these parameters. For each data set, we report the results in a certain format. The numbers of segments in the data and feature space are

Table 1 Summary of data

Data set	Purpose	Number of used features	Number of instances	Number of classes	Data type
Gender Voice [18]	Determining male or female based on voice characteristics	21	3168	2	Classification
Wine [19]	Predicting wine quality	11	4898	7	
Concrete compressive strength [19]	Predicting concrete compressive strength	8	1030	No classes	Regression
Abalone [19]	Predicting the age of abalone	7	4177		

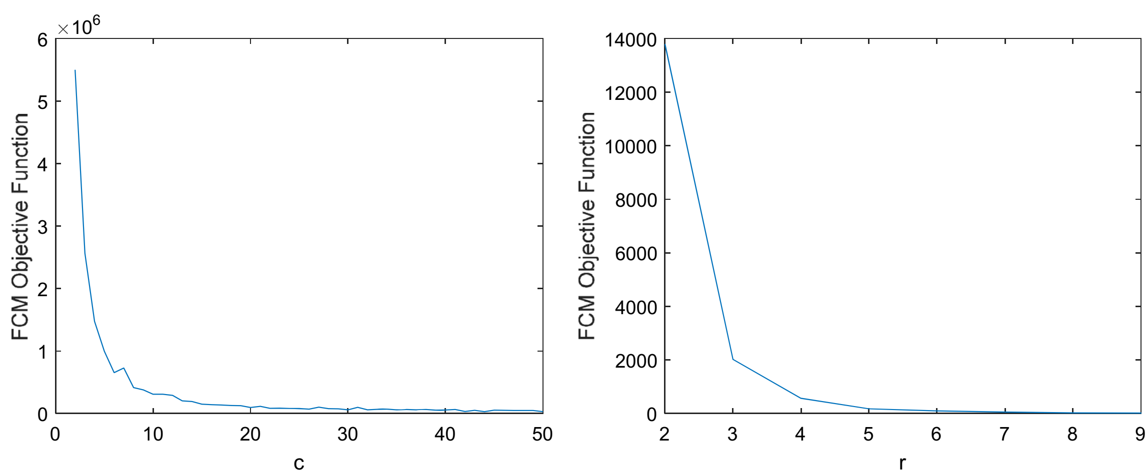

Fig. 9 Performance indexes (objective functions) for successive values of c and r for the *Gender Voice* data set

Table 2 V_{ij} (*Gender voice* data set) for $i = 1, 2,\ldots, c$; $j = 1, 2,\ldots,r$

r/c	2	4	7
2	0.9034 0.6605	0.5702 0.3176	0.6500 0.2096
	0.6469 0.8373	0.7180 0.1565	0.7109 0.1170
		0.3774 0.2925	0.3406 0.2795
		0.8934 0.6456	0.7655 0.5814
			1.0411 0.4863
			0.8444 0.1584
			0.3761 0.1446
3	0.9034 0.0432 0.9692	0.5702 0.5529 0.2000	0.5106 0.2496 0.1061
	0.6469 2.5030 0.0044	0.7180 0.4607 0.0044	0.6568 0.1177 0.0053
		0.3774 0.8703 0.0037	0.3533 0.4763 0.0038
		0.8934 0.0091 0.9638	0.8608 0.0011 0.9420
			0.9346 0.0180 0.4124
			0.7712 0.0745 0.2253
			0.4830 0.2374 0.0029
4	0.9523 0.1212 0.0432 0.9692	0.5685 0.5966 0.5529 0.2000	0.6806 0.1611 0.1146 0.2571
	0.6184 1.1015 2.5030 0.0044	0.7418 0.3369 0.4607 0.0044	0.7385 0.2695 0.3419 0.0045
		0.3842 0.2686 0.8703 0.0037	0.3454 0.2638 0.8310 0.0038
		0.9452 0.0653 0.0091 0.9638	0.8125 0.0144 0.0003 0.8720
			1.1050 0.0185 0.0023 0.7283
			0.8909 0.0998 0.0237 0.2258
			0.3799 0.3156 0.1976 0.1181

Table 3 V_{ij} (*Concrete* data set) for $i = 1, 2,\ldots, c$; $j = 1, 2,\ldots,r$

r/c	2	4	7
2	53.9519 83.9738	35.9535 77.5321	43.4919 107.0560
	35.6485 58.3637	40.2511 75.7059	17.5937 80.7913
		30.5896 66.3636	27.6481 54.6669
		16.1722 37.9906	34.6507 34.6476
			21.8200 29.0736
			25.9452 33.2534
			12.8776 15.0407
3	47.3335 57.2611 83.9738	39.0413 34.4096 77.5321	32.2958 49.0899 107.0560
	20.1269 43.4093 58.3637	34.5679 43.0927 75.7059	21.4409 15.6702 80.7913
		20.0424 35.8632 66.3636	11.6319 35.6563 54.6669
		14.9211 16.7977 37.9906	26.0518 38.9501 34.6476
			13.2607 26.0997 29.0736
			25.1975 26.3190 33.2534
			13.5048 12.5640 15.0407

selected on a basis of the changes of the performance indexes (objective functions) regarded as functions of c and r; see Fig. 9. They are treated as functions of the number of segments and tend to stabilize when moving towards higher values of c and r. Now we will demonstrate our scheme using one classification data set (Gender Voice data set), and one regression data set (*Concrete* Data set), then follow the same procedure for more data sets.

As we have cr information granules, the quality of obtained information granules is reported by means of the reconstruction index (6). The values of V_{ij} computed with the use of (6) for individual granules are presented in Tables 2 and 3. It is clear from these tables, and based on Fig. 9 that when the number of data clusters reaches 7 and above, and when the number feature clusters reaches 4 and above, we get a low value of the reconstruction error. This

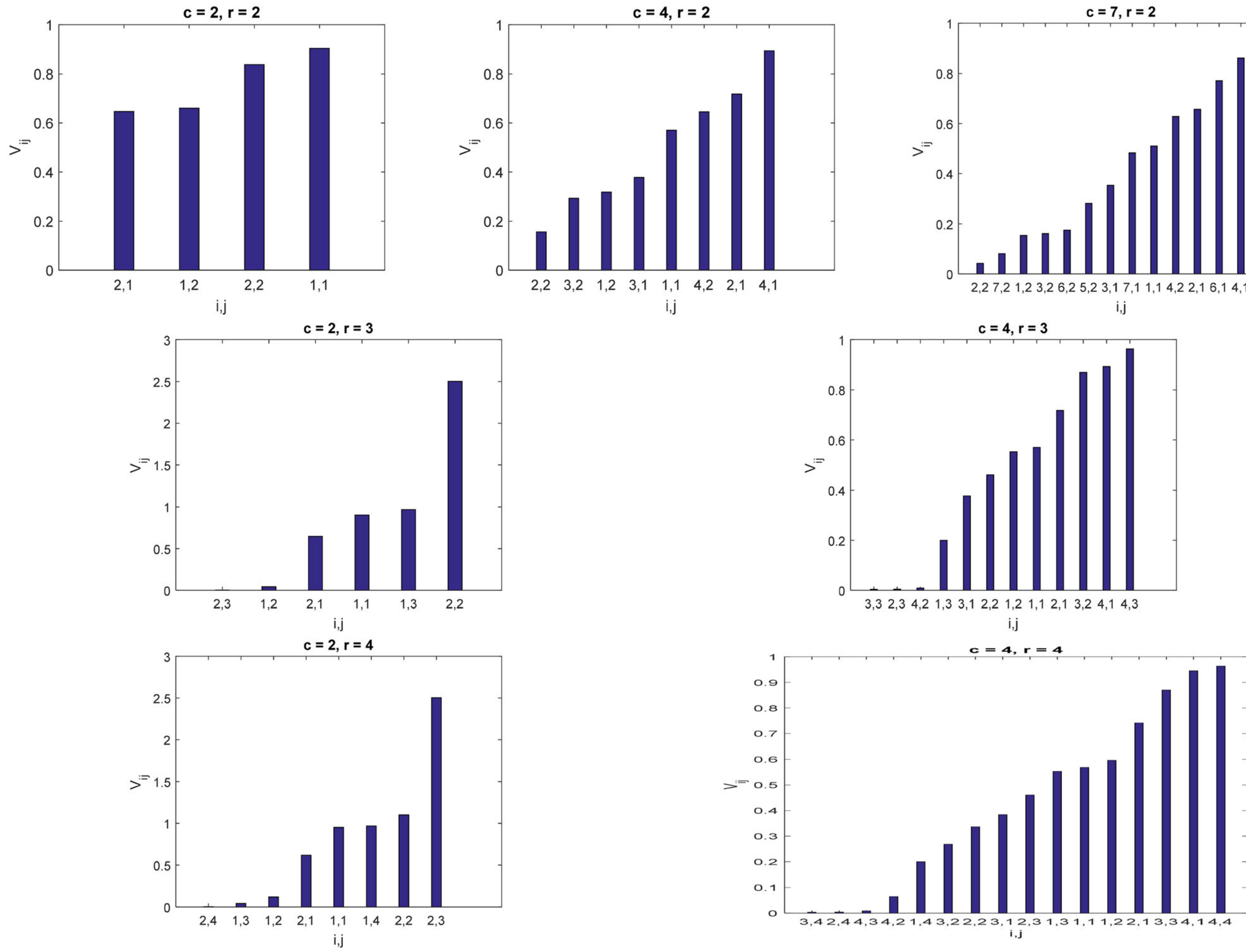

Fig. 10 V_{ij} starting from the best information granules for the *Gender Voice* data set

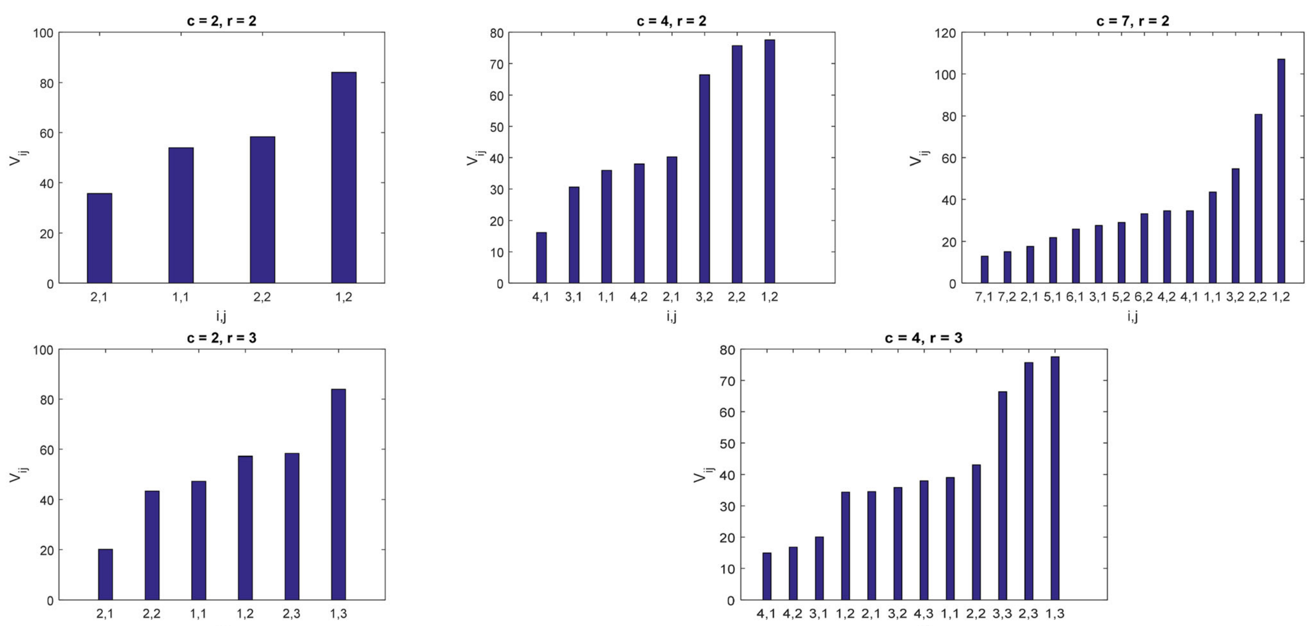

Fig. 11 V_{ij} starting from the best information granules for the *Concert* data set

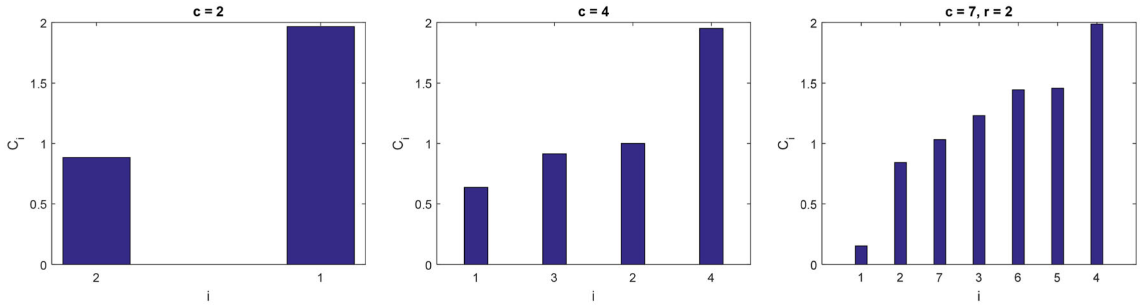

Fig. 12 Entropy of information granules for the *Gender Voice* data set

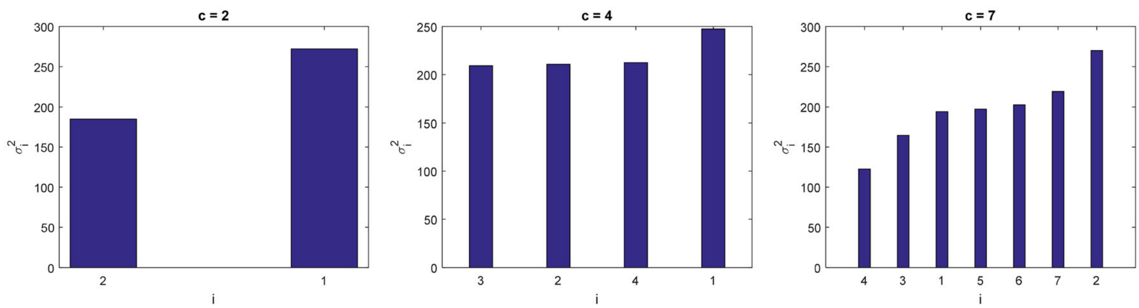

Fig. 13 Variance of information granules for the *Concert* data set

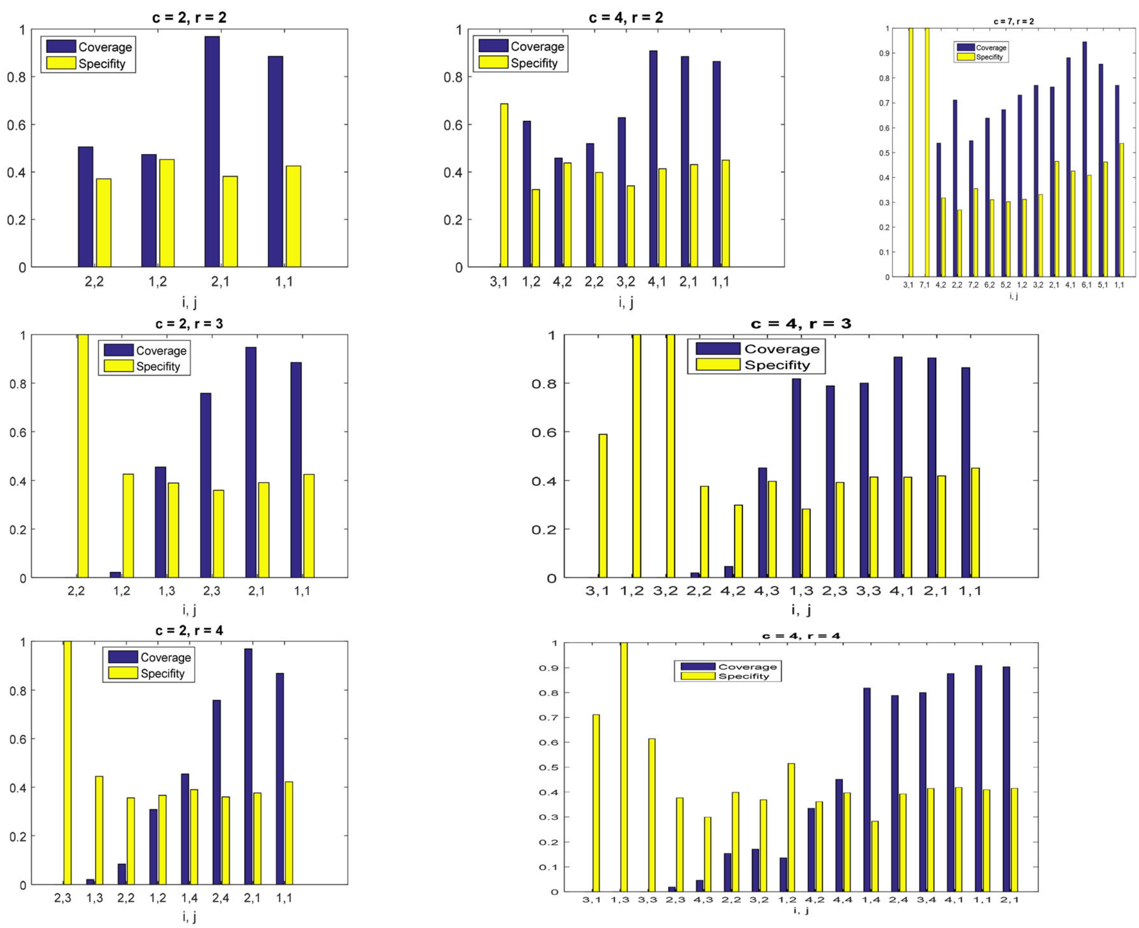

Fig. 14 Characterization of information granules in the coverage and specificity space for the *Gender Voice* data set

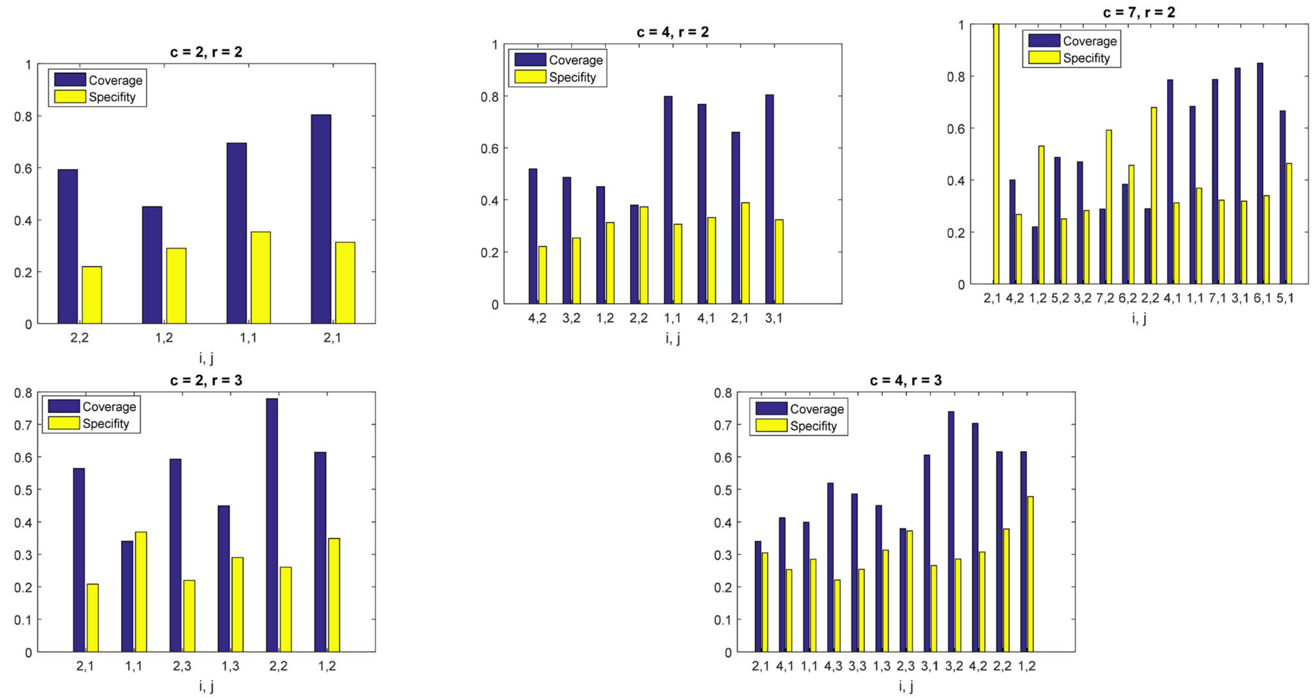

Fig. 15 Characterization of information granules in the coverage and specificity space for the *Concrete* data set

Table 4 V_{ij} (*Abalone* data set) for $i = 1, 2,\ldots, c$; $j = 1, 2,\ldots,r$

r/c	2	4	7
2	0.0810 0.0353	0.0203 0.0185	0.0090 0.0056
	0.0875 0.0659	0.0389 0.0115	0.0231 0.0062
		0.0198 0.0211	0.0083 0.0157
		0.0550 0.0612	0.0089 0.0119
			0.0097 0.0159
			0.0153 0.0272
			0.0384 0.0676
3	0.0668 0.0353 0.1095	0.0119 0.0185 0.0370	0.0079 0.0056 0.0114
	0.0197 0.0659 0.2230	0.0431 0.0115 0.0307	0.0293 0.0062 0.0107
		0.0091 0.0211 0.0413	0.0064 0.0157 0.0122
		0.0108 0.0612 0.1434	0.0067 0.0119 0.0131
			0.0062 0.0159 0.0167
			0.0074 0.0272 0.0312
			0.0081 0.0676 0.0990
4	0.0668 0.0312 0.1095 0.0476	0.0119 0.0178 0.0370 0.0207	0.0088 0.0069 0.0056 0.0114
	0.0197 0.0476 0.2230 0.1207	0.0431 0.0102 0.0307 0.0151	0.0335 0.0252 0.0062 0.0107
		0.0091 0.0172 0.0413 0.0326	0.0069 0.0059 0.0157 0.0122
		0.0108 0.0432 0.1434 0.1152	0.0071 0.0064 0.0119 0.0131
			0.0064 0.0060 0.0159 0.0167
			0.0073 0.0074 0.0272 0.0312
			0.0084 0.0078 0.0676 0.0990

Table 5 V_{ij} (*WINE* data set) for $i = 1, 2,\ldots, c$; $j = 1, 2,\ldots,r$

r/c	7	10	12
2	0.9546 0.0369	0.8440 0.0270	0.7760 0.0279
	0.8692 0.0245	0.8650 0.0314	0.8821 0.0311
	0.8049 0.0297	0.7561 0.0257	0.8054 0.0176
	0.8992 0.0358	0.8386 0.0289	0.8022 0.0182
	0.9098 0.0341	0.8765 0.1214	0.9459 0.1191
	0.8284 0.1261	0.8980 0.0349	0.7223 0.0201
	0.8763 0.2865	1.0262 0.0573	0.9422 0.0482
		0.8064 0.0223	0.8052 0.0183
		1.0210 0.3428	1.0506 0.3546
		0.7163 0.0329	0.9457 0.0260
			0.9371 0.0349
			0.7178 0.0275
3	0.9822 0.7062 0.0369	0.8794 0.5253 0.0270	0.8240 0.4654 0.0293
	0.9190 0.4207 0.0245	0.9417 0.1750 0.0314	0.9483 0.1421 0.0315
	0.8611 0.2993 0.0297	0.8157 0.2191 0.0257	0.8718 0.2238 0.0179
	0.9058 0.8395 0.0358	0.9028 0.2611 0.0289	0.9100 0.2190 0.0180
	0.9571 0.4848 0.0341	0.9561 0.1602 0.1214	1.0340 0.1214 0.1197
	0.9002 0.1828 0.1261	0.9620 0.3221 0.0349	0.7804 0.2047 0.0201
	0.7861 1.6882 0.2865	1.1128 0.2461 0.0573	1.0179 0.2505 0.0486
		0.8729 0.2075 0.0223	0.8711 0.1716 0.0189
		0.8767 2.3190 0.3428	0.9137 2.2834 0.3546
		0.7224 0.6612 0.0329	0.9950 0.2298 0.0254
			1.0014 0.3057 0.0350
			0.7145 0.7469 0.0275
4	0.9171 1.0147 0.7062 0.0369	0.6911 0.9579 0.5180 0.0308	0.6461 0.8940 0.4577 0.0279
	0.9449 0.9061 0.4207 0.0245	0.9490 0.9266 0.2321 0.0267	0.9948 0.9489 0.1429 0.0311
	0.8289 0.8772 0.2993 0.0297	0.7574 0.7891 0.2527 0.0228	0.8495 0.8797 0.2275 0.0176
	0.7000 1.0087 0.8395 0.0358	0.9051 0.9990 0.1542 0.1222	0.9281 0.8364 0.2188 0.0182
	1.0137 0.9287 0.4848 0.0341	0.9396 0.8931 0.1763 0.0254	1.0286 1.0419 0.1215 0.1191
	0.8503 0.9251 0.1828 0.1261	0.8565 1.0066 0.2959 0.0363	0.6962 0.8210 0.2088 0.0201
	0.5762 0.8911 1.6882 0.2865	1.1349 1.1173 0.2515 0.0543	0.9750 1.0410 0.2510 0.0482
		0.9366 0.9056 0.1794 0.0321	0.8900 0.8681 0.1728 0.0183
		0.6464 0.9949 2.3199 0.3419	0.6365 1.0522 2.2834 0.3546
		0.5407 0.8090 0.6305 0.0327	1.0706 1.0036 0.2241 0.0260
			0.8423 1.0872 0.3211 0.0349
			0.5316 0.8059 0.7477 0.0275

can be verified by computing the average reconstruction error.

Figures 10 and 11 (bar plot) display the values of V_{ij} starting from the best information granules (viz. with the lowest value of V_{ij}). In general, the error is low for all information granules when the values of both c and r are relatively high (4 or higher for r, and 7 or higher for c).

In case of classification data, the quality of information granules is evaluated with the aid of entropy (9). The obtained values of entropy are shown in an increasing order by proceeding with the lowest value; see Fig. 12. While for the regression data, the quality of information granules is evaluated with the aid of the variance (10) shown in an increasing order by proceeding with the lowest value (see Fig. 13). When the number of information granules increases, the average variance and average vagueness decrease.

Proceeding with the characterization of information granules built on a basis of the numeric prototypes, we display the optimal values of coverage and specificity (viz. the values obtained when the product of coverage and specificity achieved the highest value). Some selected results are displayed in Figs. 14 and 15. It can be noted that

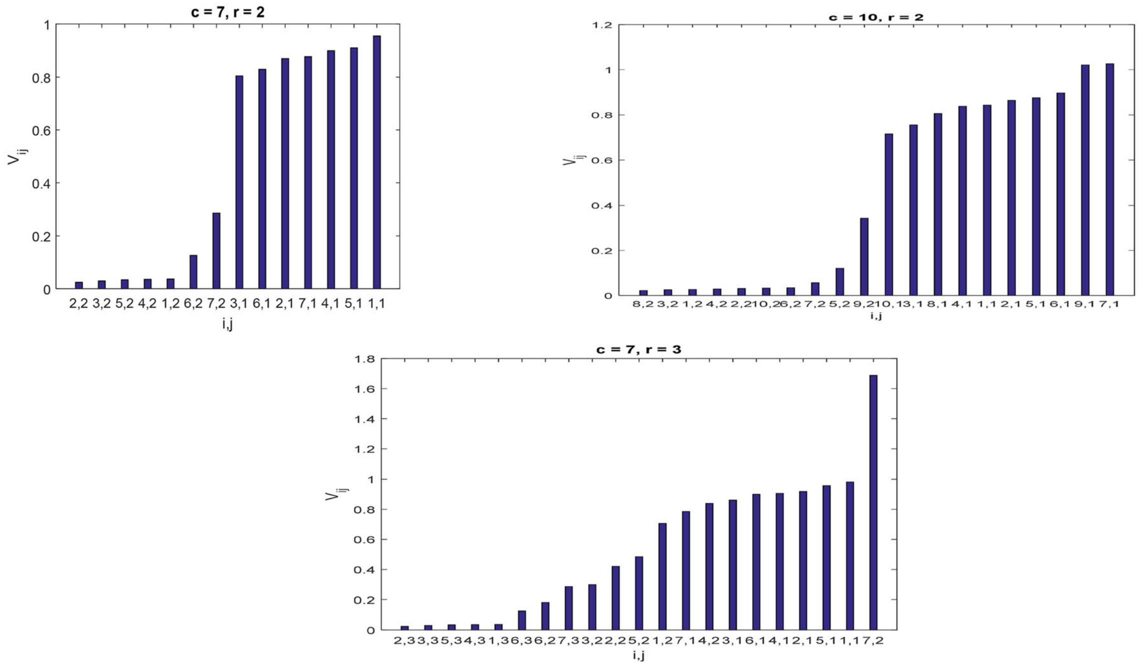

Fig. 16 V_{ij} starting from the best information granules for the *Wine* data set

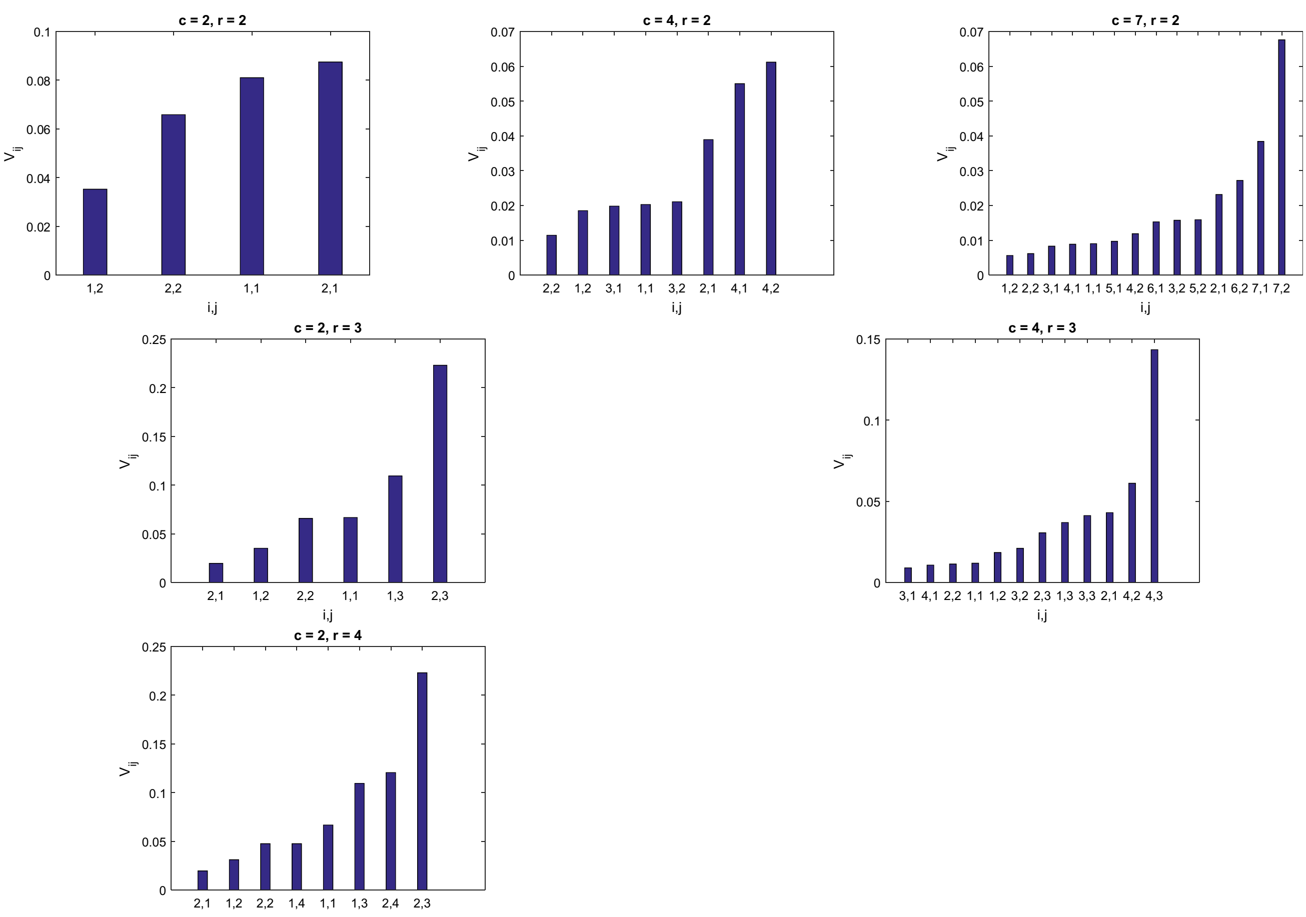

Fig. 17 V_{ij} starting from the best information granules for the *Abalone* data set

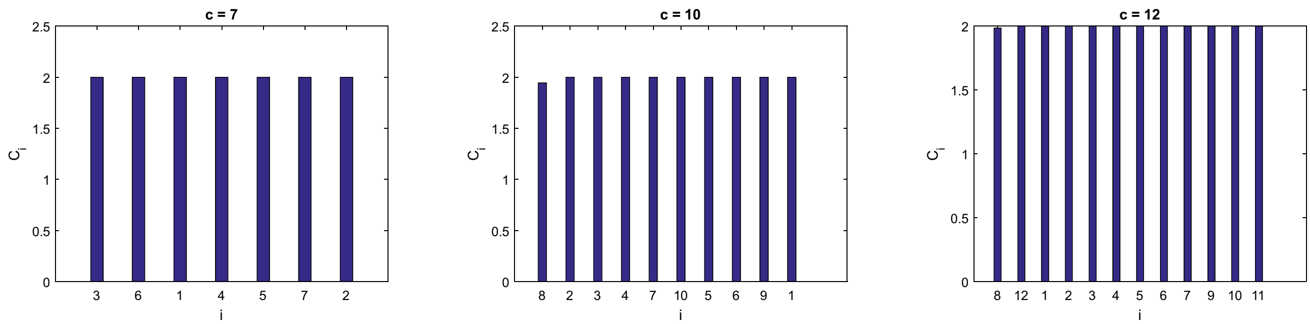

Fig. 18 Entropy of information granules for the *Wine* data set

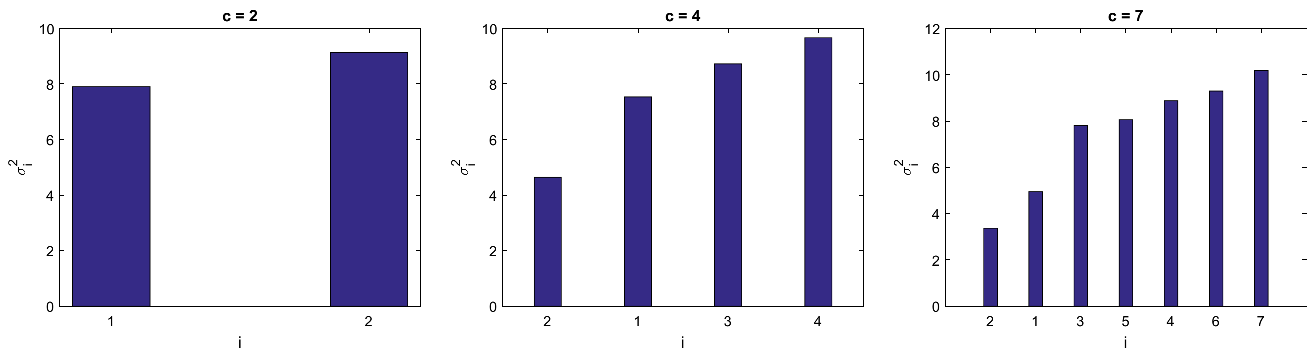

Fig. 19 Entropy of information granules for the *Abalone* data set

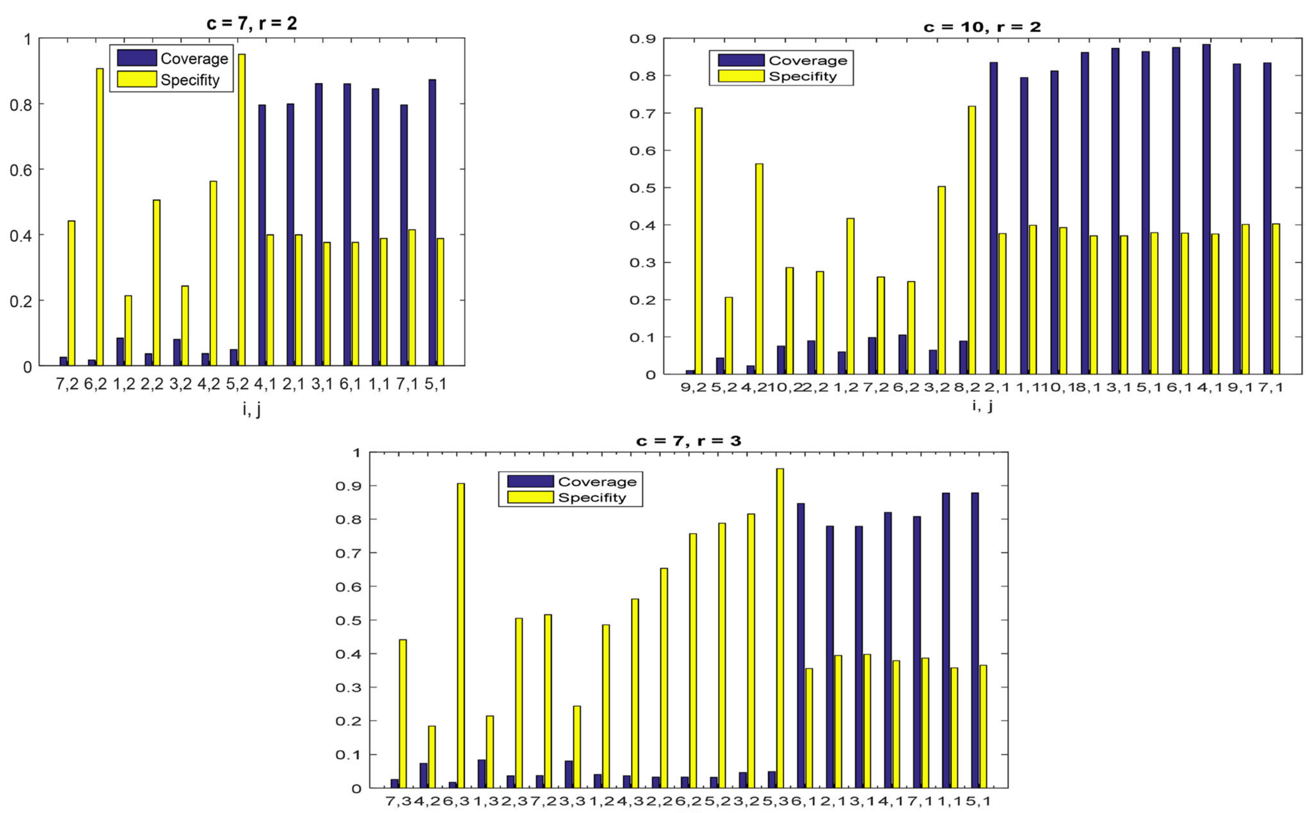

Fig. 20 Characterization of information granules in the coverage and specificity space for the *Wine* data set

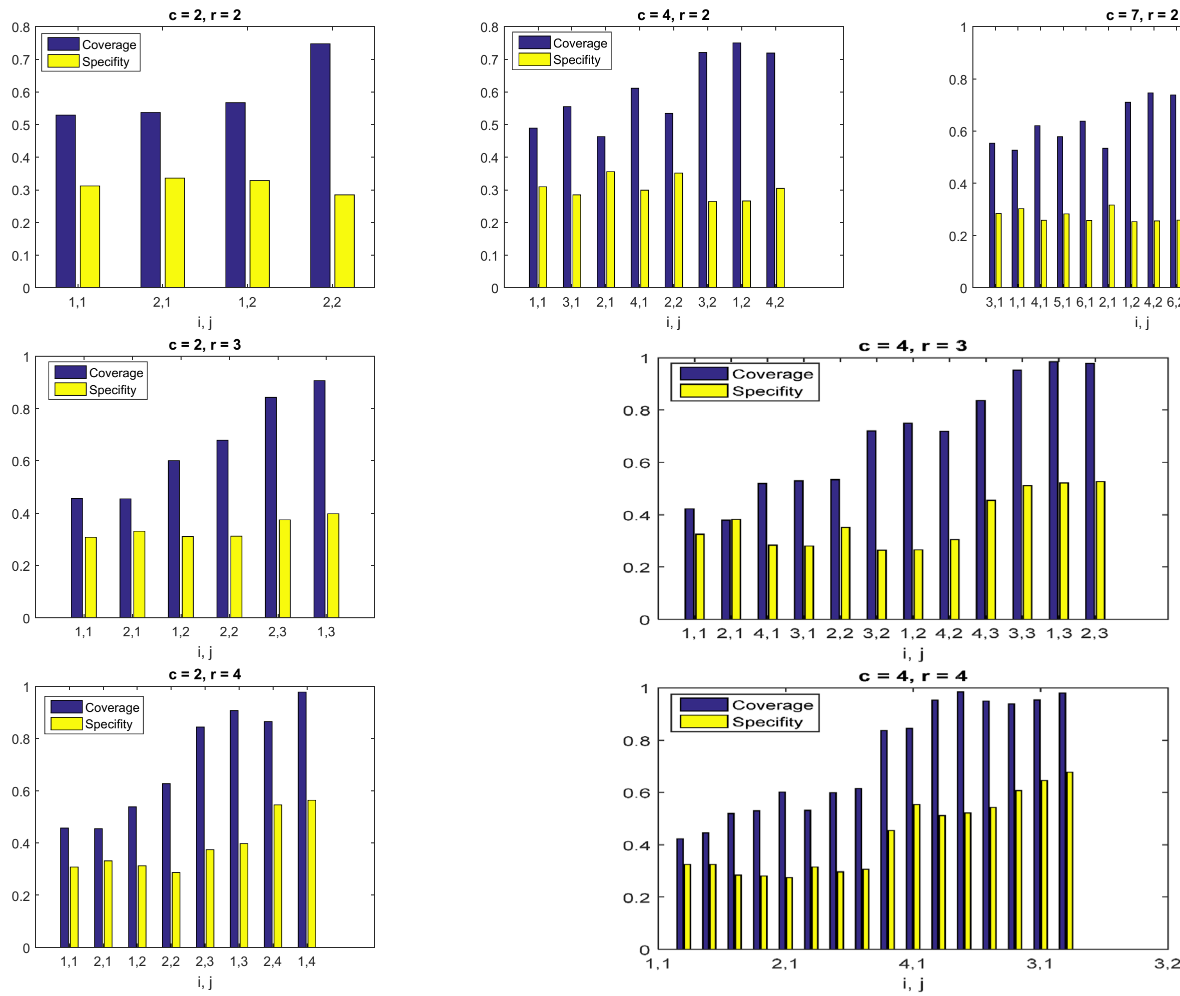

Fig. 21 Characterization of information granules in the coverage and specificity space for the *Abalone* data set

when the total number of information granules is high, the average product of coverage and specificity becomes low.

Proceeding with the remaining data sets, we report the results in a similar manner in Tables 4, 5 and Figs. 16, 17, 18, 19, 20 and 21. These tables and figures support all conclusions we have reached by inspecting the above two data sets.

7 Conclusions

The study was devoted to the concise development of data by constructing their numeric representatives followed by the augmentation of the prototypes expressed in terms of information granules. The developed optimization environment helps quantify the quality of information granules (in terms of entropy and diversity) and numeric prototypes (evaluated by means of the reconstruction error). Information granules form a blueprint of data and constitute an initial setting for a variety of constructs and classifiers, predictors and association networks. It is worth stressing that information granules are functional building modules that are used as generic components in the development of a plethora of models including predictors and classifiers. Equally important is the fact that the study delivered a way to quantify the quality of information granules with regard to their classification or prediction capabilities. Likewise, the multiview perspective at experimental data is essential to cope with massive data as one constructs essential information granules pertinent that are central to facilitate an efficient way of building classifiers and predictors.

While Sect. 5 elaborates on the fundamentals of the modeling constructs (which owing to the use of information granules can be referred to as granular predictors, granular classifiers, etc.), more detailed studies could follow that focus on the detailed architectures and some following learning schemes.

Acknowledgements This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, Saudi

Arabia, under Grant No. (KEP-5-135-39). The authors, therefore, acknowledge with thanks DSR technical and financial support.

Appendix A

Used symbols

Symbol	Description
D_i	Data cluster i
F_j	Feature cluster j
r	Number of feature clusters
c	Number of data clusters
Q	FCM objective variable
x_k	Data point k
z_j	Feature j
N	Total number of data points
n	Total number of features
v_i	Data cluster i prototype
m	Fuzzification coefficient
u_{ik}	The membership value of a data point x_k to the data cluster i
g_{ij}	The membership value of a feature z_j to the feature cluster i
V_{ij}	Reconstruction error produced for (D_i, F_j)
$\lVert\cdot\rVert_{F_j}$	Distance completed for features forming F_j
ρ_{ij}	The probability class j exists in information granule i
v_{ij}	Data cluster i prototype computed by averaging cluster data points just for features forming F_j
h	Entropy
C_i	Vagueness of the i_{th} information granule
σ_{iy}	The variance of the output values of data cluster i
$\bar{y}_i$	The average of the output values of data cluster i
R_i	
N_{ij}	Number of data points in information granule $G_{ij} \equiv (D_i, F_j)$
cov	Coverage
sp	Specify
$\hat{y}$	Predicted y value
ρ	Predicted class

References

1. Sun, S., Shawe-Taylor, J., Mao, L.: PAC-Bayes analysis of multi-view learning. Inf. Fusion **35**, 117–131 (2017)
2. Jiang, B., Qiu, F., Wang, L.: Multi-view clustering via simultaneous weighting on views and features. Appl. Soft Comput. **47**, 304–315 (2016)
3. Pontes, B., Giráldez, R., Aguilar-Ruiz, J.S.: Biclustering on expression data: a review. J. Biomed. Inform. **57**, 163–180 (2015)
4. Zong, L., Zhang, X., Yu, H., Zhao, Q., Ding, F.: Local linear neighbor reconstruction for multi-view data. Pattern Recognit. Lett. **84**, 56–62 (2016)
5. Yin, J., Sun, S.: Multiview uncorrelated locality preserving projection. IEEE Trans. Neural Netw. Learn. Syst. (2019). https://doi.org/10.1109/TNNLS.2019.2944664
6. Henriques, R., Antunes, C., Madeira, S.C.: A structured view on pattern mining-based biclustering. Pattern Recognit. **48**, 3941–3958 (2015)
7. Bezdek, J.C., Ehrlich, R., Full, W.: FCM: the fuzzy c-means clustering algorithm. Comput. Geosci. **10**, 191–203 (1984)
8. Lunwen, W.: Study of granular analysis in clustering. Comput. Eng. Appl. **5**, 29–31 (2006)
9. Liberti, L., Lavor, C., Maculan, N., Mucherino, A.: Euclidean distance geometry and applications. SIAM Rev. **56**, 3–69 (2014)
10. Dattorro, J.: Convex optimization & euclidean distance geometry. Lulu.com (2010)
11. Zadeh, L.A.: Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst. **90**, 111–127 (1997)
12. de Sa, J., Silva, L., Santos, J., Alexandre, L.: Minimum error entropy classification. In Studies in Computational Intelligence. Springer (2013)
13. Wang, X., Pedrycz, W., Gacek, A., Liu, X.: From numeric data to information granules: a design through clustering and the principle of justifiable granularity. Knowl. Based Syst. **101**, 100–113 (2016)
14. Pedrycz, W., Homenda, W.: Building the fundamentals of granular computing: a principle of justifiable granularity. Appl. Soft Comput. **13**, 4209–4218 (2013)
15. Pedrycz, W., Al-Hmouz, R., Morfeq, A., Balamash, A.: The design of free structure granular mappings: the use of the principle of justifiable granularity. IEEE Trans. Cybern. **43**, 2105–2113 (2013)
16. Pedrycz, W.: The principle of justifiable granularity and an optimization of information granularity allocation as fundamentals of granular computing. J. Inf. Process. Syst. **7**, 397–412 (2011)
17. Balamash, A., Pedrycz, W., Al-Hmouz, R., Morfeq, A.: Granular classifiers and their design through refinement of information granules. Soft. Comput. **21**, 2745–2759 (2017)
18. Machine Learning Data (MLData). https://www.mldata.io/dataset-details/gender_voice/. Accessed 1 Sept 2019
19. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/index.php. Accessed 1 Sept 2019

Abdullah Balamash received the Ph.D. degree in electrical and computer engineering from the University of Arizona, Tucson, AZ, USA, in 2004. In August 1996, he received a scholarship from the Saudi government to pursue his M.S. and Ph.D. degrees. He is an Associate Professor at King Abdulaziz University (KAU), Jeddah, Saudi Arabia. In 2018, he became the head of the Electrical and Computer engineering Department at King Abdulaziz University. In 2015, he became the deputy director of the center of excellence in intelligent engineering systems. In January 1995, he joined the Department of Electrical and Computer Engineering of KAU, Jeddah, as a Teaching Assistant. From 1991 to 1994, he worked for the Saudi Consolidated Electricity Company as a System Engineer. His research interests are in machine learning and soft computing.

Witold Pedrycz (IEEE Fellow 98) is Professor and Canada Research Chair (CRC) in Computational Intelligence in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada. He received MSc, PhD and DSci, all from the Silesian University of Technology, Gliwice, Poland, in 1977, 1980, and 1984, respectively. He is also with the Systems Research Institute of the Polish Academy of Sciences, Warsaw, Poland. In 2009 Dr. Pedrycz was elected a foreign member of the Polish Academy of Sciences. In 2012 he was elected a Fellow of the Royal Society of Canada. Witold Pedrycz has been a member of numerous program committees of IEEE conferences in the area of fuzzy sets and neurocomputing. In 2007 he received a prestigious Norbert Wiener award from the IEEE Systems, Man, and Cybernetics Society. He is a recipient of the IEEE Canada Computer Engineering Medal, a Cajastur Prize for Soft Computing from the European Centre for Soft Computing, a Killam Prize, and a Fuzzy Pioneer Award from the IEEE Computational Intelligence Society. His main research directions involve Computational Intelligence, fuzzy modeling and Granular Computing, knowledge discovery and data mining, fuzzy control, pattern recognition, knowledge-based neural networks, relational computing, and Software Engineering. He has published numerous papers in this area. He is also an author of 15 research monographs covering various aspects of Computational Intelligence, data mining, and Software Engineering. Dr. Pedrycz is intensively involved in editorial activities. He is an Editor-in-Chief of **Information Sciences**, Editor-in-Chief of **WIREs Data Mining and Knowledge Discovery** (Wiley), and **Int. J. of Granular Computing** (Springer). He currently serves on the Advisory Board of **IEEE Transactions on Fuzzy Systems** and is a member of a number of editorial boards of other international journals.

Rami Al-Hmouz received the BSc degree in electrical engineering from Mutah University in 1998, the MSc degree in computer engineering from the University of Western Sydney in 2004, and the PhD degree in computer engineering from the University of Technology, Sydney, in 2008. Currently, he is a professor at King Abdulaziz University (KAU), Jeddah, Saudi Arabia. His research interests are in machine learning, computer vision, and granular computing. Dr. Al-Hmouz is a senior member of the Institute of Electrical and Electronics Engineers.

Ali Morfeq received the Ph.D. degree from the University of Colorado at Boulder, Boulder, CO, USA, in 1990. In 1985, he received the M.S. degree from Oregon State University, Corvallis, OR, USA. He is currently the Vice Dean of the Faculty of Engineering, King Abdulaziz University (KAU), Jeddah, Saudi Arabia. He worked as a Faculty Member with the Electrical and Computer Engineering (ECE) Department, KAU. He is an Associate Professor at King Abdulaziz University (KAU), Jeddah, Saudi Arabia. He taught several courses. He also served in several administration positions in the university such as the Chairman of ECE Department, and as the Vice Dean of admission and registration of the university, and as the Head of many academic and administrative committees. Additionally, he worked as a Consultant to many firms including big industrial companies and several hospitals. His research interests are databases, data science, and soft computing.