Cluster characterization through a representativity measure

Marie-Jeanne Lesot and Bernadette Bouchon-Meunier

Laboratoire d'Informatique de Paris 6
Université Pierre et Marie Curie
8 rue du capitaine Scott, 75 015 Paris, France
{Marie-Jeanne.Lesot, Bernadette.Bouchon-Meunier}@lip6.fr

Abstract. Clustering is an unsupervised learning task which provides a decomposition of a dataset into subgroups that summarize the initial base and give information about its structure. We propose to enrich this result by a numerical coefficient that describes the cluster representativity and indicates the extent to which they are characteristic of the whole dataset. It is defined for a specific clustering algorithm, called Outlier Preserving Clustering Algorithm, OPCA, which detects clusters associated with major trends but also with marginal behaviors, in order to offer a complete description of the inital dataset. The proposed representativity measure exploits the iterative process of OPCA to compute the typicality of each identified cluster.

1 Introduction

Given a set of numerical descriptions of datapoints, information extraction often consists in clustering [7, 6], i.e. decomposing the dataset into homogeneous and distinct subgroups that summarize the initial base. This learning task gives information about the data structure and highlights the major trends present in the dataset. Yet it usually overlooks [3, 9] the marginal behaviors represented by atypical points or outliers, although the latter are also necessary to characterize the dataset.

More precisely, a complete characterization should contain information both on major trends and atypical behaviors, together with knowledge about the group representativity, as illustrated by the following example: consider a device having three different modes, semantically described as "high", "low" and "abnormally low". This linguistic characterization includes the atypical behavior together with the major ones, which is indeed necessary to describe accurately the process. Moreover it indicates that the three components are not equivalent and it specifies the most important cases: the adverb "abnormally" underlines the peculiarity of the third mode and conveys information about its representativity. It is necessary to the description and, for instance, it makes it possible to distinguish this device from another one having three normal modes, described as "high", "low" and "very low".

Therefore, we propose to enrich clustering results by defining numerical coefficients that measure the cluster representativity, or equivalently their exceptionality, to indicate the extent to which they are characteristic of the whole dataset. Such coefficients complete the information about the data subgroups and can for instance help analysts to provide linguistic labels for the clusters, e.g. to choose the appropriate modifying adverbs.

We thus define an *exceptionality coefficient*, denoted *cExc*, which takes the reference value 1 for characteristic subgroups and a high value for clusters with low representativity that correspond to minor behaviors. It is based on a specific clustering algorithm, the Outlier Preserving Clustering Algorithm (OPCA) [9], which simultaneously detects the major and minor trends present in a dataset; more precisely, *cExc* is deduced from the iterative process of OPCA and does not require expensive additional computations to provide additional information.

The article is organized as follows. Section 2 recalls the principle of the Outlier Preserving Clustering Algorithm on which the exceptionality coefficients are based. Section 3 defines the exceptionality coefficients and compares them with other numerical coefficients which also characterize a dataset, namely the *local outlying factors* proposed by Breunig *et al.* [2] and the typicality degrees proposed by Rifqi [10], we underline their differences and their respective properties. Lastly section 4 illustrates and interprets the results obtained on artificial and real datasets.

2 Outlier preserving clustering algorithm

In this section, we briefly recall the principle of the Outlier Preserving Clustering Algorithm, OPCA [9], which simultaneously identifies subgroups corresponding to major and marginal trends: it is able to build classic clusters, as provided by classic clustering algorithms, but also one-point clusters, corresponding to outliers, as provided by outlier detection methods, and lastly intermediate clusters, corresponding to small sets of similar outliers, which may be overlooked by both clustering and outlier detection techniques.

2.1 Justification

Usually, outliers are considered as noisy points which are to be excluded from the learning dataset, so that they do not disturb the learning process. For instance, robust clustering algorithms [4, 5] decrease the weight associated to such datapoints, so as to reduce their influence on the final result. Yet, in our characterization aim, they are to be considered differently: outliers should not be rejected, but considered as minor but significant groups that should be present as any other cluster.

One approach could be to decompose clustering into two cases, distinguishing the outlier handling from the other points: one could perform a first step of outlier detection method [1, 8, 2] as a preliminary to clustering; one can also model the outliers in a specific cluster [3] or define a specific distribution to handle them [11].

Using such an approach, it could be possible to perform a second clustering step on this specific cluster to identify relevant subgroups among outliers. OPCA takes a different approach which does not distinguish between the two types of points and handles them the same way: as different clustering algorithms are sensitive to different types of clusters, it appears that combining them makes it possible to exploit their different properties, and thus to optimize simultaneously, compactness and separability (see [9] for more details)..

2.2 Principle

OPCA is based on the combination, in an iterative process, of the single linkage agglomerative hierarchical algorithm and the fuzzy c-means, denoted AHC_{min} and *fcm* respectively. The coupling makes it possible to exploit the different properties of each method, namely the ability of AHC_{min} to take into account the group separability and the ability of *fcm* to provide compact subgroups; the iterative process makes it possible to adapt the outlier definition to local contexts, in particular to take into account variable significant distance scales, depending on the local data density.

More precisely, the algorithm considers a group G and first determines the most appropriate dividing method: if G contains well-separated subgroups, it is decomposed by AHC_{min}; if G has a low separability but has still a low compactness, it is split by *fcm*. This process is then iterated on each obtained subgroup. The criteria selecting the applied algorithm respectively evaluate the presence of gaps in G data distribution and estimate G compactness as a function of its diameter and its variance, compared to the average variance of its possible subgroups (see [9] for more details).

2.3 Results and limitations

Figure 1 illustrates the results obtained with artificial one-dimension datasets where datapoints are represented as stars, possibly repeated and vertically shifted in case of multi-occurrences, and clusters are represented by their membership functions defined as Gaussians of order 4 [9]. Dataset 1 (fig. 1a) corresponds to an outlier detection task in a simple case, the algorithm manages to identify the three intuitive subgroups. Dataset 2 (fig. 1b) shows a more complex case with variable densities and a definition of outliers that must depend on the locally significant distances; the obtained fuzzy subsets correspond to a natural description of the data which represents both major and marginal trends. Lastly, Dataset 3 (fig. 1c) shows a dataset which leads to a fuzzy profile identical to the first one; yet the importance of the leftmost cluster is higher than on graph (a) and it should be interpreted differently: a linguistic characterization could be “high”, “low” and “abnormally low” in the first case and “high”, “intermediate” and “low” in the second one, the absence of adverb showing that the three groups

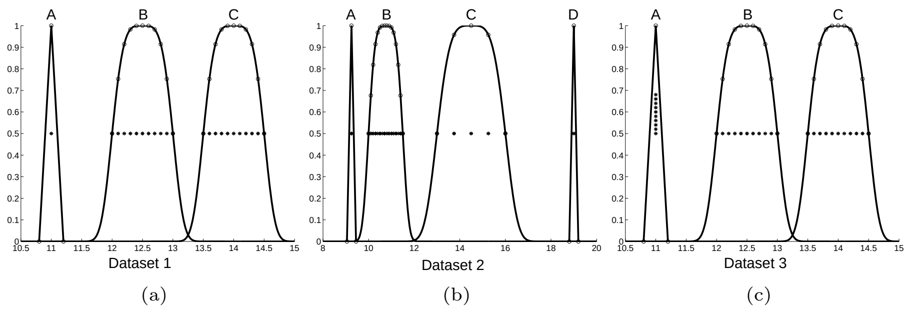

Fig. 1. Profiles of artificial one-dimension databases obtained with the outlier preserving clustering method. The fuzzy membership functions are defined as Gaussian functions of order 4 (see [9]).

are equivalent. Such instances show that fuzzy profiles are not sufficient to characterize a dataset; we propose to enrich it through exceptionality coefficients to measure the extent to which a cluster is representative of the whole dataset.

3 Representativity measure

The previous clustering algorithm highlights the structure of a dataset through the identification of all relevant subgroups. In fact, it provides a much richer information about the extracted subgroups owing to its iterative process: beside the data assignment, it offers a chronological account of the group constitution; figure 2 shows a tree representation of the group identification history for the three previous datasets. This can be exploited to measure the cluster representativity and it is the basis of the definition of exceptionality coefficients.

In what follows, we will use the notion of *isolation date* or *identification date* for a cluster, associated with the use of OPCA: it corresponds to the iteration step from which a group is not further divided but considered as a group that must appear as such in the final data decomposition; for instance it equals 2 for cluster C of Dataset 2.

3.1 Exceptionality coefficient definition

To justify our definition of exceptionality coefficients, we first consider the extreme case of single-point clusters, i.e. outliers. Such points are considered as non-characteristic of the whole dataset with an intensity that depends on their distance to the nearest point: a local outlier, as point B of Dataset 2, better represents the whole dataset than a global outlier, as point D, and thus must have a lower $cEcx$. Now this information is contained in the chronological account of the group constitution through the isolation date: the significant distance scale decreases when the number of iterations of OPCA increases, which means that

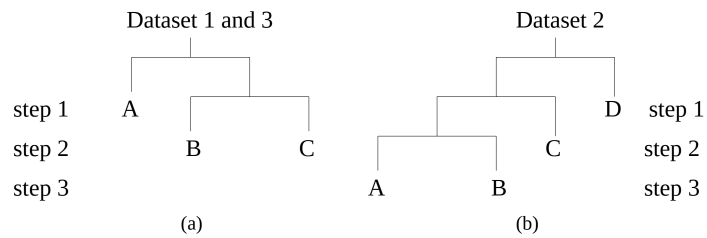

Fig. 2. Chronological account of the groups constitution. (a) for Datasets 1 and 3; (b) for Dataset 2. Letters correspond to clusters as indicated on figure 1.

a datapoint isolated at the end of the process is only a local outlier. Therefore representativity increases when the isolation date increases.

In case of larger groups, this principle still holds, but it must be balanced by the group size: isolation is associated with separability, insofar as an isolated cluster is well-separated from its neighbors; thus it is not directly equivalent to exceptionality. For instance a large and well-separated cluster may represent the whole dataset better than smaller groups identified later which may be justified only at a more refined distance scale.

We propose to measure cluster representativity by the following exceptionality coefficient, where G denotes the group under consideration, $|G|$ its size, and $h(G)$ its isolation date, i.e. the iteration number at which it is not further divided, lastly $\bar{s}_t$ is the average size of groups handled at step t. We then define the *exceptionality coefficient* characterizing cluster G as

$$cExc(G) = \frac{1}{Z} \frac{\max_{\Gamma} h(\Gamma)}{h(G)} \frac{\bar{s}_{h(G)}}{|G|} \tag{1}$$

where Z is a normalization factor defined to set a reference value: a cluster is fully representative of the whole dataset if it gets a $cExc$ value of 1, its exceptionality increases when $cExc$ increases. The isolation date $h(G)$ is compared to the total number of iterations (this constant factor could be included in Z but helps the lisibility). The group size $|G|$ is considered with respect to the size of equivalent groups $\bar{s}_{h(G)}$, i.e. the size of clusters built at the same step; it makes it possible to take into account the local context and measures a local lack of balance in terms of size.

According to this definition, exceptionality is maximal for early identified clusters of small size, which are indeed the least representative groups, thus it corresponds to the desired properties.

3.2 Comparison with other approaches

In this section, we compare the proposed exceptionality coefficients with existing numerical coefficients which also characterize the data, namely *local outlying factors* [2] and typicality degrees [10]. We underline their respective properties which provide different insights on the dataset.

Local outlying factors Breunig *et al.* [2] propose to extend the classic binary definition of outliers to a graded notion that quantifies the "outlierness" of each point, and define the *local outlying factor*, denoted *lof*, which measures a local isolation degree: for any point x, $lof(x)$ is defined as a quotient between the density around x and the density around x neighbors, where the density around a point y is a function of the average distance between y and its nearest neighbors.

The first difference between *cExc* and *lof* concerns the characterized elements: *lof* is computed for each individual datapoint whereas *cExc* applies to data subgroups. This difference is linked to the objectives of the two quantities: *lof* belongs to an outlier detection framework which considers points individually and does neither take into account nor provide any information about natural grouping, whereas *cExc* aims at enriching clustering results.

lof and *cExc* can be compared directly in the special case of single-point clusters, which are simultaneously datapoints and subgroups. Then both can be interpreted as local isolation degrees, they only differ in the way they define it: through the density quotient, *lof* compares average distances to nearest neighbors, whereas *cExc* compares isolation dates. The latter is equivalent to comparing isolation distances as the iterative property of the clustering algorithm implies that the significative distance scale decreases at each step.

For larger groups, more differences appear: *lof* values vary and are minimal for points "deep inside the cluster" [2], which are less isolated and thus less outlying than data near the cluster limits. On the contrary, by definition, *cExc* has the same value for all points in the same cluster. One could characterize the whole group with *lof* for instance computing the average *lof* value for points belonging to a same cluster, this could give results similar to *cExc* (provided the hyperparameters for *lof* definition are well chosen). Yet in this case, it seems preferable to exploit thoroughly the information offered by OPCA and compute *cExc* instead of carrying out additional independant computations to evaluate *lof*.

Typicality degrees Exceptionality coefficients can also be seen as evaluating the cluster typicality and are to be compared to the typicality degrees defined by Rifqi [10] in the context of fuzzy prototype construction. These degrees are computed in a supervised learning framework on fuzzy data where each point is associated with a class label and described as a membership function on a numerical domain; they indicate for a datapoint x the extent to which it is typical of the class it belongs to, i.e. the extent to which it is similar to other points of its class and/or distinct from members of the other class. For each point, the typicality degree is computed as the aggregation of its average within-class similarity and inter-class dissimilarity.

Typicality degrees are first different from exceptionality coefficients because, as *lof*, they characterize each datapoint individually and not data subgroups. But the major distinction is to be made about the considered reference: typicality degrees indicate the extent to which a point is characteristic of the class it belongs to, whereas *cExc* and *lof* take the whole dataset as reference.

This reference difference is particularly sensitive on single-point clusters: in the typicality case, outliers appear as having an obviously high within-class similarity as they are the only member of their class, and their isolation leads to a high inter-class dissimilarity. Thus they are viewed as especially typical points whereas for *lof* and *cExc* they correspond to exceptional behaviors. This is due to the fact that typicality degrees give a more local information related to a specific class and not to the whole dataset.

Thus the representativity measure defined by *cExc* is quite different from *lof* and typicality degrees, which themselves are distinct: as a summary, typicality degrees characterize each point in a supervised framework indicating the extent to which it is typical of the class it belongs to, *lof* characterize each datapoint in an unsupervised framework, indicating the extent to which it is typical of the whole dataset, *cExc* characterize each identified subgroup indicating the extent to which it is typical of the whole dataset. Thus, like typicality degrees, *cExc* takes into account information on cluster decomposition, but it is closer to *lof* insofar as the it takes as reference all datapoints; it differs from *lof* on the "granularity" point of view and thus provides a different insight on the data.

4 Experimental results

In this section, we illustrate the previous differences between the representativity measure, the local outlying factor and the typicality degrees and we highlight the clustering enrichment provided by *cExc*. We use artificial and real datasets, in one or two dimensions so that graphical representations of the results are possible.

4.1 Computations of *lof* and typicality degrees

Local Outlying Factor *Lof* values depend on the number k of nearest neighbors taken into account to compute the density; for a datapoint x, $lof(x)$ is defined as the maximal value obtained when k varies in a pre-defined interval $[minptsLB,\ \ minptsUB]$ [2]. *minptsLB* corresponds to the minimal number of points a group has to contain so that other objects can be local outliers with respect to that cluster, *minptsUB* is the maximal cardinality of a subgroup for which all points can be potentially outliers. In absence of *a priori* knowledge, we set these values to 4 and 20 for our small one-dimension artificial databases, and 10 and 40 for the larger two-dimension dataset (which in this case respectively correspond to the twentieth and the fifth part of the total number of instances).

Typicality degrees Rifqi [10] defined typicality degrees in a supervised framework for fuzzy data; applying it to unsupervised learning with numerical data requires to adapt the initial definitions. To define data labels, we apply OPCA and use the obtained clusters as labels, although it induces a strong bias in the typicality degrees: clustering aims at extracting compact and well-separated subgroups, i.e. it optimizes a kind of typicality. Therefore, we expect all typicality degrees to be high: the aim of these experiments is to compare the data characterization with that provided by *cExc* and *lof*, rather than exploiting the computed values as such.

As regards the quantities involved in the degree computation, we made the following choices, denoting x and y datapoints, and C_x the class label associated with x:

$$\text{similarity measure}: s(x,y) = \frac{1}{1+\|x-y\|^2}$$

$$\text{dissimilarity measure}: d(x,y) = \frac{1}{\max_{u,v}\|u-v\|}\|x-y\|$$

internal aggregator : mean

thus within-class similarity : $R(x) = \text{mean}\{s(x,y), y \in C_x\}$

and inter-class dissimilarity : $D(x) = \text{mean}\{d(x,y), y \notin C_x\}$

external aggregator : probabilistic t-conorm

thus final typicality degree : $T(x) = R(x) + D(x) - R(x) * D(x)$

4.2 Artificial datasets

One-dimension data Figure 3 and 4 illustrate the obtained values of *cExc*, *lof* and typicality degrees for one-dimension artificial data. The datapoints are represented with stars at ordinate 0, possibly repeated and vertically shifted in case of multi-occurrences; note that the scale changes from one graph to another.

General properties Fig. 3(a, b, c) corresponds to Dataset 1 (defined on fig. 1a) and illustrates the general properties of the three data characterizations presented in the previous section, in terms of outlier handling: *cExc* and *lof* provide similar results and highlight the fact that the lefmost point is not representative, i.e. is an outlier with respect to the whole dataset; indeed its associated values are 18 for *cExc* and 4.5 for *lof*, both are significantly higher than the reference value 1. On the contrary, in terms of typicality degrees, this datapoint appears as maximally typical of its own class. The graphs also show that *lof* values vary for data within natural groups, and take lower values deep inside the cluster; averaging these values give an equivalent to exceptionality coefficients. These comments are valid for all other datasets, we will not repeat them.

Fig. 3(d, e, f) illustrate another effect of the theoretical differences between *cExc*, *lof* and typicality degrees, in terms of chaining effect: the dataset consists in two natural groups linked through a chain of data close one to another. In this case, OPCA resorts to fuzzy c-means to handle the data and identify the two groups; exceptionality coefficients and *lof* values then indicate that the two groups are equivalent (the difference between the two values comes from a slight cardinality difference, and it is not significant). For typicality degrees, it appears that the central data have a low typicality because they are too similar to data belonging to the other class. This difference again is due to the fact that typicality

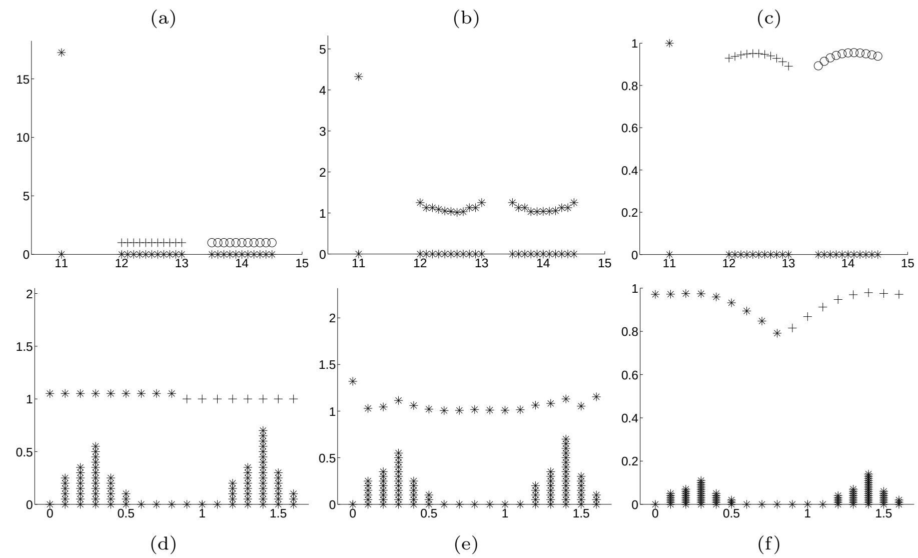

Fig. 3. Values of exceptionality coefficients, *lof*, and typicality degrees for two one-dimensional artificial datasets. The data are represented with ordinates 0, repeated and vertically shifted in case of multiple occurrences, for *cExc* and typicality degrees, the symbols indicate the cluster information. Note that the scale changes from one graph to another.

degrees take as reference the cluster itself and not the whole dataset. Note that *lof* values, which only consider local density values, show a slight difference for the extreme data.

Clustering enrichment The results obtained with Dataset 3, which was presented on fig. 1c, shown on fig. 4(a, b, c), highlight the information provided by representativity measure and how it complements the clustering results: Dataset 3 only differs from Dataset 1 in the number of occurrences of the leftmost datapoint which constitutes an outlier. For Dataset 3, this point gets $cExc = 2.5$ against 18 for Dataset 1 (see fig. 3a), which is far less significant when compared to the reference value 1; thus *cExc* makes it possible to distinguish between the two cases and shows that in the second one, the associated cluster is less to be considered as an outlier. On the contrary, *local outlying factors* and typicality degrees hardly make a difference between Datasets 1 and 3: for *lof*, the leftmost points of the two clusters appear slightly more outlying because their neighborhood contains a denser region; likewise, the typicality degrees are slightly higher because the inter-class dissimilarity increases with the additional points. Both for *lof* and typicality degrees, the differences between Datasets 1 and 3 are not significant: they provide information at another granularity level and do not aim at characterizing the dataset as a whole.

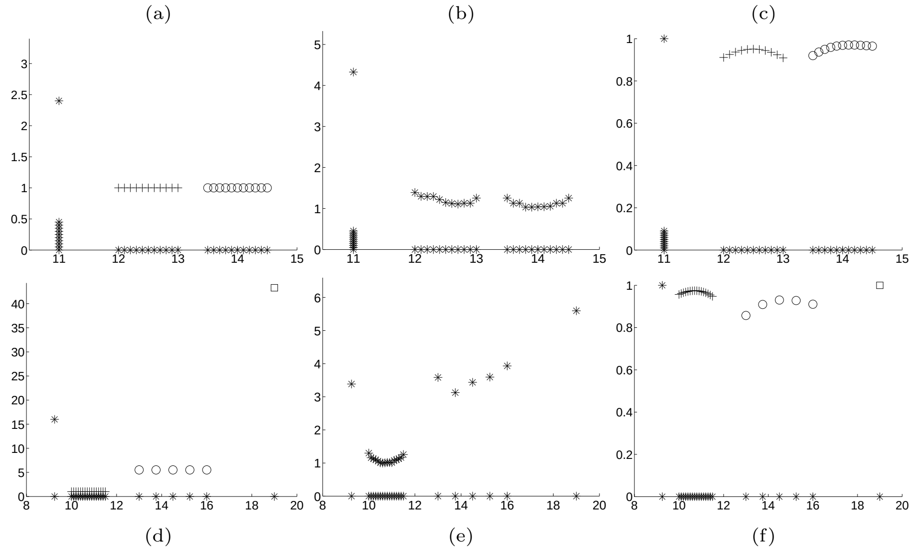

Fig. 4. Values of exceptionality coefficients, *lof*, and typicality degrees for two one-dimensional artificial datasets. Note that the scale changes from one graph to another.

Fig. 4(d, e, f) correspond to Dataset 2 defined on fig. 1b and illustrate a more complex case. Exceptionality coefficients distinguish between the local and the global outliers (resp. leftmost and rightmost datapoints) and show that the local one is more representative of the whole dataset; moreover they indicate that the low density group is less typical of the whole dataset than the denser one but more characteristic than the outliers. Thus they enrich the clustering result and complete the subgroups decomposition.

The characterization provided by *lof* values lead to a less intuitive ranking: the most outlying point is also the rightmost point, but there is no real difference between the local outlier and the average *lof* of the low density group. This is due to the fact that the local density is equivalent in the two cases (with the neighborhood definition implied by the hyperparameter choice, the result is different for other hyperparameter values); *cExc* distinguishes between them because its definition is based on the isolation date, which provides more aggregated information than individual isolation distances. Lastly, this dataset also shows that typicality degrees are sensitive to local density: the low density group has a lower typicality, which is due to the fact that the within-class similarity is lower in this case than for denser groups.

These artificial one-dimension datasets illustrate the possible interpretation of exceptionality coefficients and the differences existing between *cExc*, *local outlying factors* and typicality degrees.

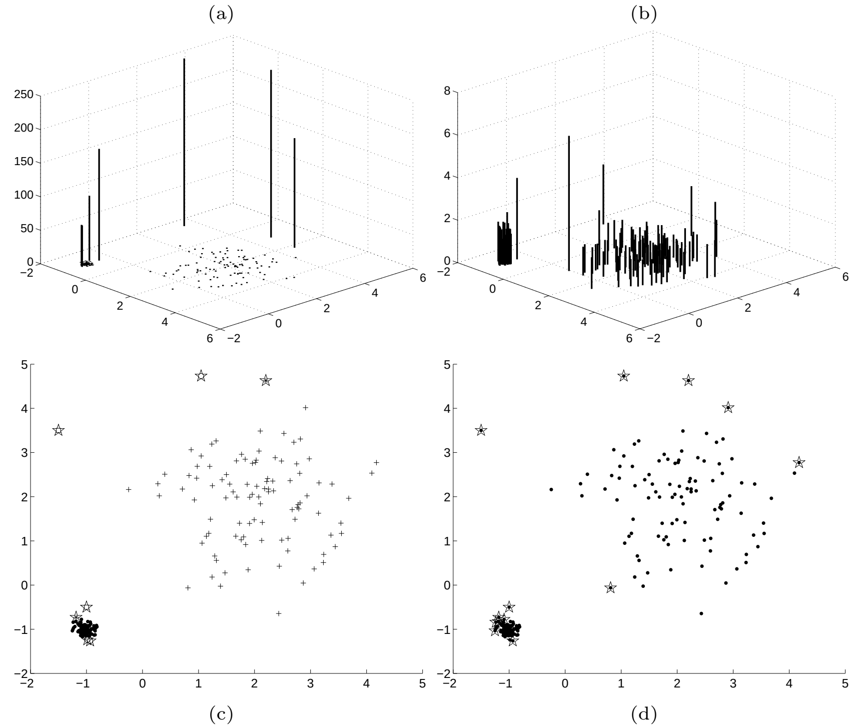

Fig. 5. Two-dimension artificial data handling. (a) exceptionality coefficients; (b) *local outlying factors*; (c) and (d) are the corresponding 2D projections, where stars represent datapoints for which the coefficient values (*cExc* or *lof*, resp.) differ from the average by more than one standard deviation.

Two-dimension artificial dataset Figure 5 illustrates the results obtained by exceptionality coefficients and *local outlying factors* on an artificial database made of two Gaussians with different variances and two outliers with coordinates $(-1, -0.5)$ and $(-1.5, 3.5)$. The 3D plots (graphs (a) and (b)) respectively show the *cExc* and *lof* values for each point, the 2D plots (graphs (c) and (d)) underline the identified outliers, here defined as points for which the coefficient value differs from the average by more than one standard deviation. With this definition, it appears that *cExc* combined to OPCA identifies the two major clusters, which get a maximal representativity coefficients, and the outliers, which get higher exceptionality coefficients, and some other points which indeed are locally isolated. It is to be noted that two points get the same maximal *cExc* value, $cExc = 249.9$, corresponding to the outlier $(-1.5, 3.5)$ and an isolated point $(1.05, 4.73)$, this result appears as relevant and intuitive when looking at the data distribution (see fig. 5c).

Local outlying factors give naturally less information as they tell nothing about natural grouping of similar points but they only indicate the most outlying ones; it appears that the contrast between the maximal and minimal values is much lower than for *cExc*. According to our criterion, *lof* identifies 13 outliers (7 are isolated by *cExc*), which only partially overlap with *cExc* results and have a different ranking: the most outlying point for *lof* is none of the predefined outliers, but a locally isolated point situated between the two major groups: its neighborhood contains points from both clusters, which lead to a density contrast and a high *lof* value. Globally, the obtained results seem difficult to justify, as compared to *cExc*: in the latter case, the information is aggregated at a cluster level, and thus it is already summarized and easier to interpret.

4.3 Real dataset

Figure 6 illustrates the information gain provided by the exceptionality coefficients: it shows the results obtained for two student classes, described by their marks for a specific test. The fuzzy profile may lead to think that the two classes are quite similar, they have the same main structure made of a large middle group, with some smaller groups. The left class appears less homogeneous as it contains two students having difficulty but has a more important best student groups.

The exceptionality coefficients modify this impression and stress the difference between the two classes which do not have the same level: the left one can be globally characterized as having better results. Indeed, in the right case, the best students appear as outliers and the data is mostly described by the middle cluster, whereas on the left case, the exceptions correspond to the lower results and the high ones are more characteristic of the whole data.

5 Conclusion

We propose a technique to enrich a clustering result in order to better characterize a dataset: applying a clustering algorithm which detects simultaneously major trends and marginal behaviors, we define a numerical coefficient that quantifies the extent to which the groups are representative of the whole dataset and this representativity measure gives an interesting insight on the data.

The obtained results show that these coefficients provide a useful information and complete the subgroups decomposition. They can be particularly helpful in natural concept modelling, whose aim is to establish a link between numerical values and semantic knowledge expressed by linguistic terms: usually analysts are asked to label many examples and supervised learning methods are applied to determine the associated characterization. The enriched clustering technique makes it possible to extract automatically the relevant subgroups, represent them with fuzzy subsets and provides additional information necessary for labelling and choosing the most appropriate modifying adverb. Thus an expert is not asked to label all data, but only a reduced number of fuzzy characterizations.

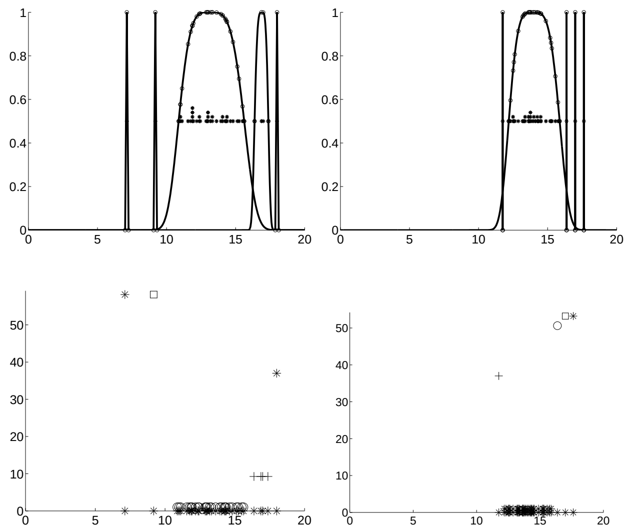

Fig. 6. Real data: evaluation of two students years. First row, profiles obtained with OPCA with Gaussian of order 4 membership functions; second row, exceptionality coefficients.

References

1. V. Barnett and T. Lewis. *Outliers in statistical data.* Wiley and Sons, 1994.
2. M. Breunig, H. Kriegel, R. Ng, and J. Sander. LOF: Identifying density-based local outliers. In *Proc. of ACM SIGMOD*, volume 29, pages 93–104. ACM, 2000.
3. R. Davé. Characterization and detection of noise in clustering. *Pattern Recognition Letters*, 12:657–664, 1991.
4. R. Davé and R. Khrishnapuram. Robust clustering methods: a unified view. *IEEE Transactions on fuzzy systems*, 5(2):270–293, 1997.
5. H. Frigui and R. Krishnapuram. A robust competitive clustering algorithm with applications in computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):450–465, 1999.
6. F. Höppner, F. Klawonn, R. Kruse, and T. Runkler. *Fuzzy Cluster Analysis, Methods for classification, data analysis and image recognition.* Wiley, 2000.
7. A. Jain, M. Murty, and P. Flynn. Data clustering: a review. *ACM Computing survey*, 31(3):264–323, 1999.
8. E. Knorr, R. Ng, and V. Tucakov. Distance based outliers : algorithms and applications. *Very Large Data Bases Journal*, 8(3–4):237–253, 2000.
9. M.J. Lesot and B. Bouchon-Meunier. Extraction de concepts descriptifs avec exceptions par classification non supervise hybride. In *Rencontres Francophones sur la Logique Floue et ses Applications, LFA'03*, Tours, France, 2003.
10. M. Rifqi. *Mesure de comparaison, typicalité et classification d'objets flous : théorie et pratique.* PhD thesis, Université de Paris VI, 1996.
11. C. Saint-Jean and C. Frélicot. An hybrid parametric model for semi-supervised robust clustering. In *Int. Conf. on Recent Developments in Mixture Modelling (MIXTURES)*, Hambourg, Germany, 2001.