Knowledge-Based Clustering in Computational Intelligence

Witold Pedrycz

Department of Electrical & Computer Engineering University of Alberta
Edmonton T6R 2G7 Canada
and Systems Research Institute, Polish Academy of Sciences
Warsaw, Poland

Summary. Clustering is commonly regarded as a synonym of unsupervised learning aimed at the discovery of structure in highly dimensional data. With the evident plethora of existing algorithms, the area offers an outstanding diversity of possible approaches along with their underlying features and potential applications. With the inclusion of fuzzy sets, fuzzy clustering became an integral component of Computational Intelligence (CI) and is now broadly exploited in fuzzy modeling, fuzzy control, pattern recognition, and exploratory data analysis. A lot of pursuits of CI are human-centric in the sense they are either initiated or driven by some domain knowledge or the results generated by the CI constructs are made easily interpretable. In this sense, to follow the tendency of human-centricity so profoundly visible in the CI domain, the very concept of fuzzy clustering needs to be carefully revisited. We propose a certain paradigm shift that brings us to the idea of *knowledge*-based clustering in which the development of information granules – fuzzy sets is governed by the use of data as well as domain knowledge supplied through an interaction with the developers, users and experts. In this study, we elaborate on the concepts and algorithms of knowledge-based clustering by considering the well known scheme of Fuzzy C-Means (FCM) and viewing it as an operational model using which a number of essential developments could be easily explained. The fundamental concepts discussed here involve clustering with domain knowledge articulated through partial supervision and proximity-based knowledge hints. Furthermore we exploit the concepts of collaborative as well as consensus driven clustering. Interesting and useful linkages between information granularity and privacy and security of data are also discussed.

1 Introductory Comments

The human-centric facet of Computational Intelligence (CI) becomes profoundly visible in a significant number of developments. One could mention here system modeling, pattern recognition, and decision-making. In data analysis tasks completed in the setting of the CI, the phenomenon of human-centricity manifests quite vividly in several ways and the needs there are well

Witold Pedrycz: *Knowledge-Based Clustering in Computational Intelligence*, Studies in Computational Intelligence (SCI) **63**, 317–341 (2007)
www.springerlink.com

articulated. First, the results are presented at some suitable level of abstraction secured by the use of information granules. Likewise the semantics of the information granules that are used to organize findings about data is conveyed in the language of fuzzy sets whose interpretation is quite intuitive. In this sense, we envision that the available mechanisms of presentation of results to the end-user are quite effective. Second, the communication with the human at the entry point when the data sets become analyzed is not that well developed. Domain knowledge available there is crucial to the build up of models (say, fuzzy models) and the establishment of their transparency and readability. It is worth stressing that the transparency and accuracy are the two dominant requirements of fuzzy models we are interested in satisfying to the highest possible extent.

The effective two-way communication is a key to the success of CI constructs, especially if we are concerned with the ways how all computing becomes navigated. For instance, the mechanisms of relevance feedback that become more visible in various interactive systems hinge upon the well-established and effective human-centric schemes of processing in which we effectively accept user hints and directives and release results in a highly comprehensible format.

Given the existing algorithms of clustering that are pivotal to the design of information granules (and as such playing an important role in the CI constructs), we become cognizant that the principles guiding processing realized by them need to be augmented. The main quest is to assure that the fuzzy clustering operates not only on data (its data-driven optimization underpinnings are well known) but takes full advantage of various sources of knowledge available when dealing with the problem at hand. In particular, we anticipate that any guidance available from the user could be incorporated as a part of the optimization environment. This point of view as to the unified treatment of data and knowledge in clustering augments the existing principle of data analysis and gives rise to the concept of knowledge-based clustering. The ultimate objective of this study is to introduce and explore various scenarios where knowledge could be seamlessly included into the algorithmic architecture of fuzzy clustering. We discuss several fundamental concepts such as clustering with partial supervision and proximity knowledge hints, collaborative clustering and a consensus mode of clustering.

The organization of the material reflects the main concepts discussed in the study. For the sake of completeness, in Section 3, we study with a brief note on the FCM algorithm by highlighting the main features that make its role visible in the CI domain. Section 3 is devoted to the formulation of the main challenges and spells out a number of open questions. In Section 4, we cover the mechanisms of human-oriented guidance such as partial supervision and proximity-based clustering. Distributed data mining in the unsupervised mode is discussed in Section 5. Collaborative fuzzy clustering is presented in Section 6 where we formulate the problem, discuss privacy aspects linked with information granularity, and present the underlying principles. The vertical mode of collaboration is presented along with the detailed design phases (Section 7). Further we elaborate on consensus based clustering. Concluding comments are covered in Section 9.

2 Fuzzy C-Means (FCM) as an Example of the CI Algorithm of Data Analysis

To make a consistent exposure of the overall material and assure linkages with the ensuing optimization developments, we confine ourselves to one of the objective function based fuzzy clustering. More specifically, we consider a Fuzzy C-Means (FCM) [4] governed by the following objective function

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

where $\mathbf{x}_k$ denotes an multidimensional data point (pattern), $\mathbf{v}_i$ is an i-th prototype and $U = [u_{ik}]$, $i = 1, 2, \ldots, c$; $k = 1, 2, \ldots, N$ is a partition matrix. $\|.\|$ denotes a certain distance function and "m" stands for a fuzzification coefficient; $m > 1.0$. The minimization of (1) is realized with respect to the partition matrix and the prototypes. The optimization scheme and all specific features of the minimization of Q are well reported in the literature, refer for instance to [1, 20]. What is of interest to us here is an observation that fuzzy clustering is inherently associated with the granularity of information. In a nutshell fuzzy clustering leads to the abstraction of data into a format of information granules. Two essential and somewhat orthogonal dimensions of the granulation process are envisioned: (a) numeric realization of the granulation through a collection of the prototypes, and (b) a collection of information granules – fuzzy sets represented by successive rows of the partition matrix. Interestingly enough, there is a direct correspondence between these two representations. Given a collection of prototypes we can determine the entries of the partition matrix. And vice versa, a given partition matrix along with the data gives rise to the prototypes. The interpretability of the results of the FCM is its significant and highly valuable feature of the algorithm. As a collection of fuzzy sets (described by the corresponding rows of the generated partition matrix) offer a holistic view at the structure of data, this feature of the FCM emphasizes its linkages with the main thrusts of Computational Intelligence.

3 Challenges and Open Issues

Indisputably, fuzzy clustering (including FCM) is one of the well-established conceptual and algorithmic avenues that has become an integral, highly visible construct present in numerous modeling pursuits encountered in fuzzy systems, neurofuzzy systems, and Computational Intelligence, in general. Given all of those, arises an obvious question as to the further developments that could support some open issues and anticipated or already existing challenges. They could eventually contribute to the changes of the landscape of this area in the years to come.

While any projection in the rapidly developing areas could be quire risky, there are several challenges which could be quite visible and influential in the buildup and progression of the area in the future.

Knowledge-based orientation of fuzzy clustering. A heavy and visible reliance on numeric data is an evident feature of fuzzy clustering as it could be seen today. There are, however, other important factors one has to take into account when discovering the structure in data. Various sources of knowledge are available from experts, data analysts, users and they come in various formats. The fundamental challenge concerns efficient ways of their incorporation into the clustering schemes, both as a concept and the algorithmic enhancement. This is not a straightforward task given the fact that clustering has to reconcile numeric aspects (data) and knowledge component (human factors). In essence, the knowledge-based orientation of clustering is in line of human-centricity of Computational Intelligence and the development of interaction schemes.

Distributed character of processing This challenge has emerged because of the inherently distributed nature of data. Those tend to be distributed at individual locations (say, sensor networks) and this poses an interesting quest as to the distributed clustering. The centralized mode that is predominant today in fuzzy clustering requires a careful revision. The clustering techniques available nowadays that predominantly revolve around a single, huge and centrally available dataset do require a careful re-visiting and reformulation.

Communication, collaboration and consensus building All of those aspects are associated in one way or another with the distributed nature of data sets. Given the distributed character of data, it is also very likely that they cannot be shared because of the privacy and security restrictions. On the other hand, some collaboration and interaction would be highly desirable given the fact that the structure in some datasets could be quite similar and sharing the knowledge about the discovery of clusters within one dataset with other sites could be beneficial. How to facilitate collaboration and consensus building in data analysis while respecting security requirements becomes an evident challenge.

Each of these challenges comes with a suite of their own quite specific problems that do require a very careful attention both at the conceptual as well as algorithmic level. We have highlighted the list of challenges and in the remainder of this study present some of the possible formulations of the associated problems and look at their solutions. It is needless to say that our proposal points at some direction that deems to be of relevance however does not pretend to offer a complete solution to the problem. Some algorithmic pursuits are also presented as an illustration of some possibilities emerging there.

4 Data and Knowledge in Clustering: Forming a Human-Centric Perspective of Computational Intelligence in Data Analysis

In fuzzy clustering, we are ultimately faced with the problem of optimization driven by data. This clearly emphasizes the role of data in the processes of revealing the structure. While this is the evident and dominant tendency, a shift of this data-oriented paradigm is contemplated in light of the fact that not only the data are essential but any domain knowledge available from users, designers has to play a pivotal role. Considering such domain knowledge as an important and indispensable component of data analysis, it becomes clear that it cast data analysis in some human-centric perspective. To be more descriptive, we may refer to pursuits carried out in this way as a knowledge-based clustering. There are two fundamental issues that need to be addressed in the setting of the knowledge-based clustering: (a) what type of knowledge-based hints could be envisioned, and (b) how they could be incorporated as a part of the optimization (more specifically, what needs to be done with regard to the possible augmentation of the objective function and how the ensuing optimization scheme has to be augmented to efficiently cope with the modified objective function).

5 Fuzzy Clustering and Mechanisms of Human-Oriented Guidance

In this section, we highlight several commonly encountered alternatives that emerge when dealing with domain knowledge and building formal mechanisms which reformulate the underlying objective function. We focus on two formats of domain knowledge being available in this setting that is labeling of some selected data points and assessments of proximity of some pairs of data.

5.1 Mechanisms of Partial Supervision

The effect of partial supervision involves a subset of labeled data, which come with their class membership. These knowledge-based hints have to be included into the objective function and reflect that some patterns have been labeled. In the optimization, we expect that the structure to be discovered conforms to the membership grades already provided for these selected patterns. More descriptively, we can treat the labeled patterns to form a grid of "anchor" points using which we attempt to discover the entire structure in the data set.

Put it differently, such labeled data should help us navigate a process of revealing the structure. The generic objective function shown in the form (1) has to be revisited and expanded so that the structural information (labeled data points) is taken into consideration. While there could be different alternatives possible with this regard, we consider the following additive expansion of the objective function, [20, 21, 22]

$$Q = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^2 \|\mathbf{x}_k - \mathbf{v}_i\|^2 + \alpha \sum_{i=1}^{c}\sum_{k=1}^{N} (u_{ik} - f_{ik} b_k)^2 \|\mathbf{x}_k - \mathbf{v}_i\|^2 \qquad (2)$$

The first term is aimed at the discovery of the structure in data and is the same as in the standard FCM. The second term (weighted by some positive scaling factor α) addresses the effect of partial supervision. It requires careful attention because of the way in which it has been introduced into the objective function and the role it plays during its optimization. There are two essential data structures containing information about the initial labeling process (labeled data points)

- the vector of labels, denoted by $\mathbf{b} = [b_1 b_2 \ldots b_N]^T$. Each pattern $\mathbf{x}_k$ comes with a Boolean indicator: we assign b_k equal to1 if the pattern has been already labeled and $b_k = 0$ otherwise.
- The partition matrix $F = [f_{ik}]$, $i = 1, 2, \ldots, c$; $k = 1, 2, \ldots N$ which contains membership grades assigned to the selected patterns (already identified by the nonzero values of b). If $b_k = 1$ then the corresponding column shows the provided membership grades. If $b_k = 0$ then the entries of the corresponding k-th column of F do not matter; technically we could set them up to zero.

The nonnegative weight factor (α) helps set up a suitable balance between the supervised and unsupervised mode of learning. Apparently when $\alpha = 0$ then we end up with the standard FCM. Likewise if there are no labeled patterns ($\mathbf{b} = 0$) then the objective function reads as

$$Q = (1+\alpha) \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^2 d_{ik}^2 \qquad (3)$$

and becomes nothing but a scaled version of the standard objective function encountered in the FCM optimization process. If the values of α increase significantly, we start discounting any structural aspect of optimization (where properly developed clusters tend to minimize) and rely primarily on the information contained in the labels of the patterns. Subsequently, any departure from the required membership values in F would lead to the significant increase in the values of the objective function.

One could consider a slightly modified version of the objective function

$$Q = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^2 d_{ik}^2 + \alpha \sum_{i=1}^{c}\sum_{k=1}^{N} (u_{ik} - f_{ik})^2 b_k\, d_{ik}^2 \qquad (4)$$

where the labeling vector **b** shows up in a slightly different format. In essence, this function captures the essence of partial supervision. For some slight variations on the issue of partial supervision, the reader may refer to the work by [3, 1, 15, 17, 28].

Once the objective function (2) has been optimized, the resulting entries of the partition matrix U assume the form

$$u_{ik} = \frac{1}{1+\alpha}\left[\frac{1+\alpha\left(1-b_k\sum\limits_{i=1}^{c} f_{ik}\right)}{\sum\limits_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^2} + \alpha f_{ik} b_k\right] \tag{5}$$

For $\alpha = 0$, the formula returns the result produced by the "standard" FCM. Moving on to the computations of the prototypes, the necessary condition for the minimum of Q with respect to the prototypes comes in the form $\frac{\partial Q}{\partial v_{st}} = 0$, $s = 1, 2, .., c$; $t = 1, 2, \ldots, n$. Calculating the respective partial derivatives one obtains

$$\begin{aligned}\frac{\partial Q}{\partial v_{st}} &= \frac{\partial}{\partial v_{st}}\left[\sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^2 \sum_{j=1}^{n}(x_{kj}-v_{ij})^2\right.\\ &\quad \left. +\,\alpha\sum_{i=1}^{c}\sum_{k=1}^{N}(u_{ik}-f_{ik}b_k)^2\sum_{j=1}^{n}(x_{kj}-v_{ij})^2\right] \\ &= \frac{\partial}{\partial v_{st}}\left[\sum_{i=1}^{c}\sum_{k=1}^{N}\left[u_{ik}^2+(u_{ik}-f_{ik}b_k)^2\right]\sum_{j=1}^{n}(x_{kj}-v_{ij})^2\right]\end{aligned} \tag{6}$$

Let us introduce the following shorthand notation

$$\Psi_{ik} = u_{ik}^2 + (u_{ik} - f_{ik}b_k)^2 \tag{7}$$

This leads to the optimality condition of the form

$$\frac{\partial Q}{\partial v_{st}} = 2\sum_{k=1}^{N}\Psi_{sk}(x_{kt} - v_{st}) = 0 \tag{8}$$

and finally we derive the prototypes in the following form

$$\mathbf{v}_s = \frac{\sum\limits_{k=1}^{N}\Psi_{sk}\mathbf{x}_k}{\sum\limits_{k=1}^{N}\Psi_{sk}} \tag{9}$$

5.2 Clustering with Proximity Hints

The concept of proximity is one of the fundamental notions when assessing the mutual dependency between membership occurring two patterns. Consider two patterns with their corresponding columns in the partition matrix denoted by "k" and "l", that is $\mathbf{u}_k$ and $\mathbf{u}_l$, respectively. The proximity between them, $\mathrm{Prox}(\mathbf{u}_k, \mathbf{u}_l)$, is defined in the following form [23, 25]

$$\mathrm{Prox}(\mathbf{u}_k, \mathbf{u}_l) = \sum_{i=1}^{c} \min(u_{ik}, u_{il}) \tag{10}$$

Note that the proximity function is symmetric and returns 1 for the same pattern (k = 1); however this relationship is not transitive. In virtue of the properties of any partition matrix we immediately obtain

$$\mathrm{Prox}(\mathbf{u}_k, \mathbf{u}_1) = \sum_{i=1}^{c} \min(u_{ik}, u_{il}) = \mathrm{Prox}(\mathbf{u}_1, \mathbf{u}_k) \tag{11}$$
$$\mathrm{Prox}(\mathbf{u}_k, \mathbf{u}_k) = \sum_{i=1}^{c} \min(u_{ik}, u_{ik}) = 1$$

Let us illustrate the concept of proximity for $c = 2$. In this case $u_{1k} = 1 - u_{2k}$ so that we can confine ourselves to a single argument. The resulting plot (with the first coordinates of the patterns, u_{1k} and u_{1l}) is included in Figure 1.

The incorporation of the proximity-based knowledge hints leads to the two optimization processes. The first one is the same as captured by the original objective function. In the second one we reconcile the proximity hints

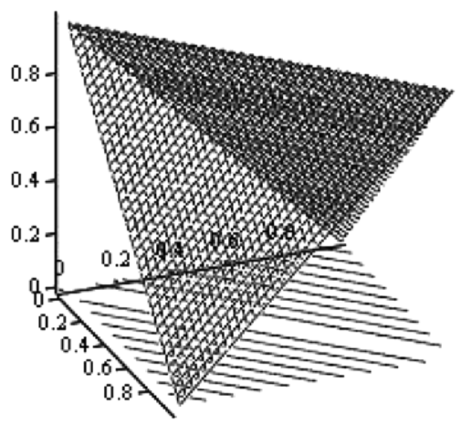

Fig. 1. Proximity function as a function of membership grades encountered in the partition matrix

with the proximity values induced by the partition matrix generated by the generic FCM. Denote the proximity values delivered by the user as $\mathrm{Prox}[k_1, k_2]$ where k_1 and k_2 are the indexes of the data points for which the proximity value is provided. Obviously these hints are given for some pairs of data so to emphasize that we introduce a Boolean predicate $B[k_1, k_2]$ defined in the following manner

$$B[k_1, k_2] = \begin{cases} 1, \text{if the value of } \mathrm{Prox}[k_1, k_2] \text{ has} \\ \quad \text{been specified for the pair } (k_1, k_2) \\ 0, \text{otherwise} \end{cases} \tag{12}$$

Note that for any pair of data, the corresponding induced level of proximity that is associated with the partition matrix produced by the FCM is computed as given by (10). We request that the proximity knowledge-based hints offered by the designer coincide with the induced proximity values implied by the structure revealed by the FCM on the basis of numeric data. Computationally, we express this requirement by computing the expression (which is a sum of distances between the corresponding values of the proximity values)

$$\sum_{k_1} \sum_{k_2} \|\mathrm{Prox}[k_1, k_2] - \sum_{i=1}^{c} \min(u_{ik_1}, u_{ik_2})\|^2 B[k_1, k_2] \tag{13}$$

By making changes to the entries of the partition matrix U, we minimize the value of the expression given above thus arriving at some agreement between the data and the domain knowledge. The optimization activities are then organized into two processes exchanging results as outlined in Figure 2. There are two optimization activities. The first one, being driven by data

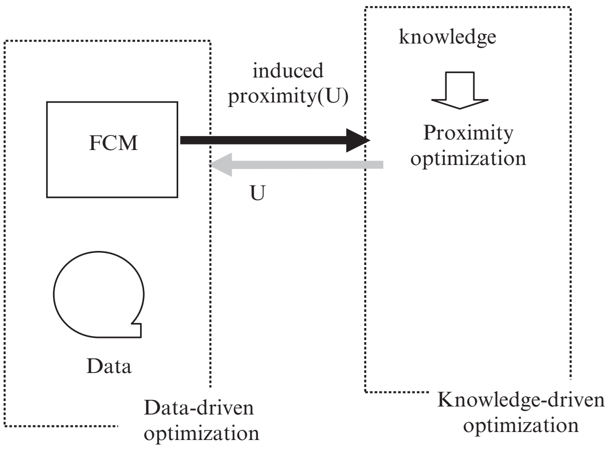

Fig. 2. The optimization data – and knowledge-driven processes of proximity-based fuzzy clustering

produces some partition matrix. The values of this matrix are communicated to the second optimization process driven by the proximity-based knowledge hints. At this stage, the proximity values induced by the partition matrix are compared with the proximities coming as knowledge hints and (13) is minimized giving rise to the new values of the partition matrix U which in turn is communicated to the data driven optimization phase. At this point, this "revised" partition matrix is used to further minimize the objective function following the iterative scheme of the FCM.

6 Distributed Data Mining

Quite commonly we encounter situations where databases are distributed rather than centralized [10, 19, 29]. There are different outlets of the same company and each of them operates independently and collects data about customers populating their independent databases. The data are not available to others. In banking, each branch may run its own database and such databases could be geographically remote from each other. In health institutions, there could be separate datasets with a very limited communication between the individual institutions. In sensor networks (which become quite popular given the nature of various initiatives such as intelligent houses, information highway, etc.), we encounter local databases that operate independently from each other and are inherently distributed. They are also subject to numerous technical constraints (e.g., a fairly limited communication bandwidth, limited power supply, etc) which significantly reduce possible interaction between the datasets. Under these circumstances, the "standard" data mining activities are faced now new challenges that need to be addressed. It becomes apparent that processing all data in a centralized manner cannot be exercised. On the other hand, data mining of each of the individual databases could benefit from availability of findings coming from others. The technical constraints and privacy issues dictate a certain level of interaction. There are two general modes of interaction that is collaborative clustering and consensus clustering both of which are aimed at the data mining realized in the distributed environment. The main difference lies in the level of interaction. The collaborative clustering is positioned at the more active side where the structures are revealed in a more collective manner through some ongoing interaction. The consensus driven clustering is focused on the reconciliation of the findings while there is no active involvement at the stage of constructing clusters.

7 Collaborative Clustering

Given the distributed character of data residing at separate databases, we are ultimately faced with the need for some collaborative activities of data mining. With the distributed character of available data come various issues of privacy, security, limited communication capabilities that have to be carefully investigated. We show that the notion of information granularity that is at heart of fuzzy sets plays a pivotal role in this setting.

7.1 Privacy and Security of Computing Versus Levels of Information Granularity

While the direct access to the numeric data is not allowed because of the privacy constraints [2, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 18, 30, 32, 33] all possible interaction could be realized through some interaction occurring at the higher level of abstraction delivered by information granules. In objective function based fuzzy clustering, there are two important facets of information granulation conveyed by (a) partition matrices and (b) prototypes. Partition matrices are, in essence, a collection of fuzzy sets which reflect the nature of the data. They do not reveal detailed numeric information. In this sense, there is no breach of privacy and partition matrices could be communicated not revealing details about individual data points. Likewise prototypes are reflective of the structure of data and form a summarization of data. Given a prototype, detailed numeric data are hidden behind them and cannot be reconstructed back to the original form of the individual data points. In either case, no numeric data are directly made available.

The level of information granularity [34] is linked with the level of detail and in this sense when changing the level of granularity possible leakage of privacy could occur. For instance, in limit when the number of clusters becomes equal to the number of data points, each prototype is just the data point and not privacy is retained. Obviously, this scenario is quite unrealistic as the structure (the number of clusters) is kept quite condensed when contrasted with all data. The schematic view of privacy offered through information granulation resulting within the process of clustering is illustrated in Figure 3. We note here that the granular constructs (either prototypes or partition matrices) build some granular interfaces.

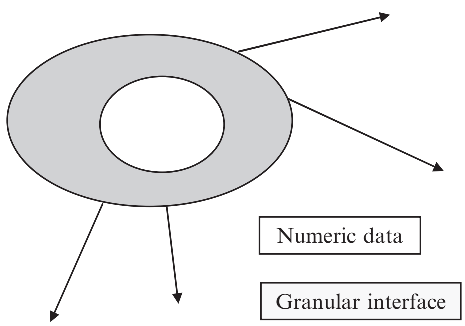

Fig. 3. Granular interface offering secure communication and formed by the results of the fuzzy clustering (partition matrices and prototypes)

7.2 The Underlying Principle of Collaborative Clustering

When dealing with distributed databases we are often interested in a collaborative style of discovery of relationships [24, 25] that could be common to all of the databases. There are a lot of scenarios where such collaborative pursuits could be deemed highly beneficial. We could envision a situation where the databases are located in quite remote locations and given some privacy requirements as well as possible technical constraints we are not allowed to collect (transfer) all data into a single location and run any centralized algorithm of data mining, say clustering. On the other hand, at the level of each database each administrator/analyst involved in its collection, maintenance and other activities could easily appreciate the need for some joint activities of data mining. Schematically, we can envision the overall situation as schematically visualized in Figure 4.

While the collaboration can assume a variety of detailed schemes, the two of them are the most essential. We refer to them as horizontal and vertical modes of collaboration or briefly horizontal and vertical clustering. More descriptively, given are "P" data sets $\mathbf{X}[1]$, $\mathbf{X}[2]$, .. $\mathbf{X}[p]$ where $\mathbf{X}[ii]$ stands for the ii-th dataset (we adhere to the consistent notation of using square brackets to identify a certain data set) in *horizontal* clustering we have the same objects that are described in *different* feature spaces. In other words, these could be the same collection of patients coming with their records built within each medical institution. The schematic illustration of this mode of clustering portrayed in Figure 4 underlines the fact that any possible collaboration occurs at the structural level viz. through the information granules (clusters) built over the data; the clusters are shown in the form of auxiliary interface layer surrounding the data. The net of directed links shows how the collaboration between different data sets takes place. The width of the links emphasizes the fact that an intensity of collaboration could be different depending upon

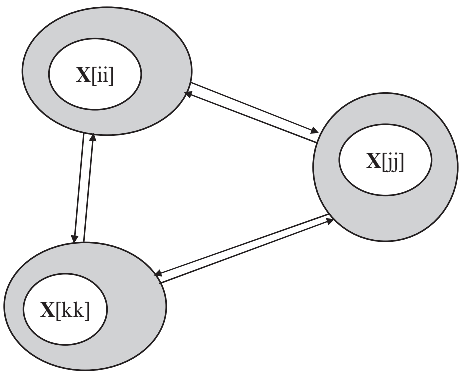

Fig. 4. A scheme of collaborative clustering involving several datasets and interacting at the level of granular interfaces

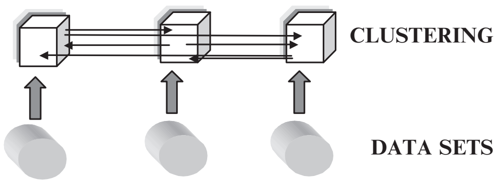

Fig. 5. A general scheme of horizontal clustering; all communication is realized through some granular interface

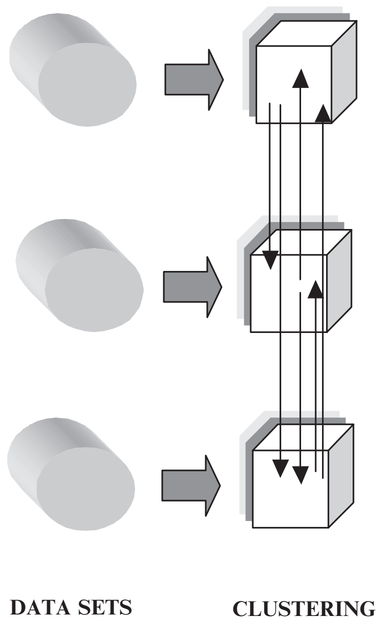

Fig. 6. A general scheme of vertical clustering; note a "stack" of data sets communicating through some layer of granular communication

the dataset being involved and the intension of the collaboration say, a willingness of some organization to accept findings from external sources).

The mode of *vertical* clustering, Figure 6, is complementary to the one already presented. Here the data sets are described in the same feature space but deal with *different* patterns. In other words, we consider that $\mathbf{X}[1]$, $\mathbf{X}[2]$, $\ldots$, $\mathbf{X}[P]$ are defined in the same feature space while each of them consists of different patterns, $\dim(\mathbf{X}[1]) = \dim(\mathbf{X}[2]) = \ldots \dim(\mathbf{X}[P])$ while $\mathbf{X}[ii]$ $\mathbf{X}[jj]$. We can show the data sets as being stack on each other (hence the name of this clustering mode).

Collaboration happens through some mechanisms of interaction. While the algorithmic details are presented in the subsequent section, it is instructive to underline the nature of the possible collaboration.

- in horizontal clustering we deal with the same patterns and different feature spaces. The communication platform one can establish is through the partition matrix. As we have the same objects, this type of collaboration makes sense. The confidentiality of data has not been breached: we do not operate on individual patterns but the resulting information granules (fuzzy relations, that is partition matrices). As this number is far lower than the number of data, the low granularity of these constructs moves us quite far from the original data
- in vertical clustering we are concerned with different patterns but the same feature space. Hence the communication at the level of the prototypes (which are high level representatives of the data) becomes feasible. Again, because of the aggregate nature of the prototypes, the confidentiality requirement has been satisfied.

There are also a number of hybrid models of collaboration where we encounter data sets with possible links of vertical and horizontal collaboration. The collaborative clustering exhibits two important features:

- The databases are distributed and there is no sharing of their content in terms of the individual records. This restriction is caused by some privacy and security concerns. The communication between the databases can be realized at the higher level of abstraction (which prevents us from any sharing of the detailed numeric data).
- Given the existing communication mechanisms, the clustering realized for the individual datasets takes into account the results about the structures of other datasets and *actively* engages them in the determination of the clusters; hence the term of collaborative clustering.

Depending upon the nature of the data located at each database and their mutual characteristics, we distinguish between two main fundamental modes of clustering such as horizontal and vertical clustering.

8 The Vertical Mode of Collaboration – The Main Flow of Processing

Let us start with setting up all necessary notation which will be subsequently used in the main phases of the development scheme. Let consider "P" databases $\mathbf{X}_1$, $\mathbf{X}_2, \ldots, \mathbf{X}_{\mathrm{P}}$ whose elements (data points, patterns) are defined in the same feature space however each of these datasets consists of different data. Schematically, we can portray it in Figure 6. Given the privacy concerns, it becomes evident that sharing the data becomes impossible however as all data points are defined in the same space, communicating at the level of the prototypes becomes feasible. By noting that, we follow the same notation as included in Figure 6. The collections of the prototypes formed at the individual datasets are denoted by $\mathbf{v}_1[\mathrm{ii}]$, $\mathbf{v}_2[\mathrm{ii}]$, …, $\mathbf{v}_{\mathrm{c}}[\mathrm{ii}]$ (the index in the square brackets pertains to the ii-th dataset).

The mode of *vertical* clustering, refer to Figure 6, is complementary to the one already presented. Here the data sets are described in the same feature space but deal with *different* patterns (data points). In other words, we consider that $\mathbf{X}[1], \mathbf{X}[2], \ldots, \mathbf{X}[\mathrm{P}]$ are defined in the same feature space while each of them consists of different patterns, $\dim(\mathbf{X}[1]) = \dim(\mathbf{X}[2]) = \ldots \dim(\mathbf{X}[\mathrm{P}])$ while $\mathbf{X}[\mathrm{ii}]$ $\mathbf{X}[\mathrm{jj}]$. We can show the data sets as being stack on each other (hence the name of this clustering mode).

In the discussion, we make a fundamental assumption about the same number of clusters. Whether this assumption is realistic or not, it still deserves more discussion. Later on we show how to relax this constraint and how this could be handled in an efficient manner.

8.1 The Development of Collaboration

The collaboration in the clustering process deserves a careful treatment. We do not know in advance if the structures emerging (or being discovered) at the level of the individual datasets are somewhat compatible and in this manner supportive of some collaborative activities. It could well be that in some cases the inherent structures of datasets are very different thus preventing from any effective collaboration to occur. The fundamental decision is whether we allow some datasets to collaborate or they should be eliminated from the collaboration from the very beginning. This important decision needs to be made upfront. One of the feasible possibilities would be to exercise some mechanisms of evaluating consistency of the clusters (structure) at site "ii" and some other dataset "jj". Consider that the fuzzy clustering has been completed separately for each dataset. The resulting structures represented by the prototypes are denoted by $\sim\mathbf{v}_1[\mathrm{ii}], \sim\mathbf{v}_2[\mathrm{ii}], \ldots, \sim\mathbf{v}_\mathrm{c}[\mathrm{ii}]$ for the ii-the dataset and $\sim\mathbf{v}_1[\mathrm{jj}], \sim\mathbf{v}_2[\mathrm{jj}], \ldots, \sim\mathbf{v}_\mathrm{c}[\mathrm{jj}]$. Consider the ii-th data set. The equivalent representation of the structure comes in the form of the partition matrix. For the ii-th dataset, the partition matrix is denoted by $\sim\mathrm{U}[\mathrm{ii}]$ whose elements are computed on the basis of the prototypes when using the dataset $\mathbf{X}[\mathrm{ii}]$.

$$\sim\mathrm{u}_{\mathrm{ik}}[\mathrm{ii}] = \frac{1}{\sum\limits_{\mathrm{j}=1}^{\mathrm{c}} \left(\frac{\|\mathbf{x}_\mathrm{k} - \sim\mathbf{v}_\mathrm{i}[\mathrm{ii}]\|}{|\mathbf{x}_\mathrm{k} - \sim\mathbf{v}_\mathrm{j}[\mathrm{ii}]\|} \right)^{2/(\mathrm{m}-1)}} \tag{14}$$

$\mathbf{x}_\mathrm{k} \in \mathbf{X}[\mathrm{ii}]$. The prototypes of the jj-th dataset being available for collaborative purposes when presented to $\mathbf{X}[\mathrm{ii}]$ give rise to the partition matrix $\sim\mathrm{U}[\mathrm{ii}|\mathrm{jj}]$ formed for the elements of $\mathbf{X}[\mathrm{ii}]$ in the standard manner

$$\sim\mathrm{u}_{\mathrm{ik}}[\mathrm{ii}|\mathrm{jj}] = \frac{1}{\sum\limits_{\mathrm{j}=1}^{\mathrm{c}} \left(\frac{\|\mathbf{x}_\mathrm{k} - \sim\mathbf{v}_\mathrm{i}[\mathrm{jj}]\|}{|\mathbf{x}_\mathrm{k} - \sim\mathbf{v}_\mathrm{j}[\mathrm{jj}]\|} \right)^{2/(\mathrm{m}-1)}} \tag{15}$$

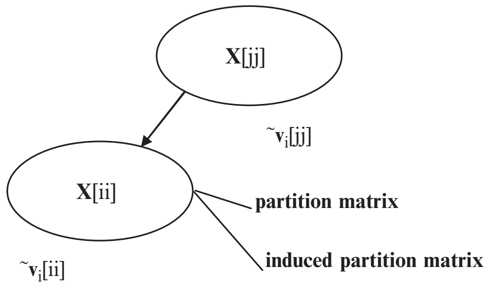

Fig. 7. Statistical verification of possibility of collaboration between datasets "ii" and "jj"

Again the calculations concern the data points of $\mathbf{X}$[ii]. Refer to Figure 7 that highlights the essence of the interaction.

Given the partition matrix ~U[ii] and ~U[ii|jj] (induced partition matrix) we can check whether they are "compatible" meaning that the collaboration between these two datasets could be meaningful. We can test whether the histograms of membership grades of ~U[ii] and ~U[ii|jj] are statistically different (that is there is a statistically significant difference). This could be done using e.g., a standard nonparametric test such as χ^2. If the hypothesis of significant statistical difference between the partition matrices (that is corresponding structures) is not rejected, then we consider that the ii-th dataset can collaborate with the jj-th one. Noticeably, the relationship is not reciprocal so the issue of collaboration of the jj-th dataset with the ii-th needs to be investigated separately.

8.2 The Augmented Objective Function

The "standard" objective function minimized at the level of the ii-th dataset comes in the well-known form of the double sum, $\sum_{i=1}^{c[ii]} \sum_{k=1}^{N[ii]} u_{ik}^m[ii] \|\mathbf{x}_k - \mathbf{v}[ii]\|^2$. Given that the we admit collaboration with the jj-th dataset, in the search for the structure we take advantage of the knowledge of the prototypes representing the jj-th dataset and attempt to make the prototypes $\mathbf{v}_1$[ii], $\mathbf{v}_2$[ii], …, $\mathbf{v}_c$[ii] to be positioned closer to the corresponding prototypes $\mathbf{v}_1$[jj], $\mathbf{v}_2$[jj[, …, $\mathbf{v}_c$[jj]. This request is reflected in the form of the augmented objective function to come in the following format

$$Q[ii] = \sum_{k=1}^{N[ii]} \sum_{i=1}^{c} u_{ik}^2[ii] d_{ik}^2[ii] + \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \beta[ii, jj] \sum_{i=1}^{c} \sum_{k=1}^{N[ii]} u_{ik}^2[ii] \|\mathbf{v}_i[ii] - \mathbf{v}_i[jj]\|^2 \qquad (16)$$

The first component is the same as the one guiding the clustering at the dataset $\mathbf{X}$[ii] while the second part reflects the guidance coming from all other datasets that we identified as potential collaborators (which is done using the χ^2 test described in the previous section). The level of collaboration (which is asymmetric) is guided by the value collaboration coefficient. Its value is chosen on a basis of potential benefits of collaboration. This will be discussed in more detail in the next section. More specifically, β[ii,jj] is a collaboration coefficient supporting an impact coming from the jj-th dataset and affecting the structure to be determined in the ii-th data set. The number of patterns in the ii-th dataset is denoted by N[ii]. We use different letter to distinguish between the horizontal and vertical collaboration. The interpretation of (20) is quite obvious: the first term is the objective function directed towards the search of structure the ii-th dataset while the second articulates the differences between the prototypes (weighted by the partition matrix of the ii-th data set) which have to be made smaller through the refinement of the partition matrix (or effectively the moves of the prototypes in the feature space).

The optimization of Q[ii] involves the determination of the partition matrix U[ii] and the prototypes $\mathbf{v}_i$[ii]. As before we solve the problem for each dataset separately and allow the results interact so that this forms collaboration between the sets. The minimization of the objective function with respect to the partition matrix requires the use of the technique of Lagrange multipliers because of the existence of the standard constraints imposed on the partition matrix. We form an augmented objective function V incorporating the Lagrange multiplier λ and deal with each individual pattern (where $t = 1, 2, \ldots, N[ii]$),

$$V = \sum_{i=1}^{c} u_{it}^2[ii] d_{it}^2[ii] + \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \beta[ii, jj] \sum_{i=1}^{c} u_{it}^2[ii] \|\mathbf{v}_i[ii] - \mathbf{v}_i[jj]\|^2 - \lambda \left(\sum_{i=1}^{c} u_{it} - 1 \right) \tag{17}$$

Taking the derivative of V with respect to u_{st}[ii] and making it zero, we have

$$\frac{\partial V}{\partial u_{st}} = 2u_{st}[ii] d_{st}^2[ii] + 2 \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \beta[ii, jj] u_{st}[ii] \|\mathbf{v}_i[ii] - \mathbf{v}_i[jj]\| - \lambda \tag{18}$$

For notational convenience, let us introduce the shorthand expression

$$D_{ii,jj} = \|\mathbf{v}_i[ii] - \mathbf{v}_i[jj]\|^2 \tag{19}$$

From (18) we derive

$$u_{st}[ii] = \frac{\lambda}{2\left(d_{st}^2[ii] + \sum\limits_{\substack{jj=1 \\ jj \neq ii}}^{P} \beta[ii, jj] D_{ii,jj} \right)} \tag{20}$$

In virtue of the standard normalization condition $\sum_{j=1}^{c} u_{jt}[ii] = 1$ one has

$$\frac{\lambda}{2} = \frac{1}{\sum_{j=1}^{c} \frac{1}{d_{jt}^2[ii] + \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \beta[ii,jj] D_{ii,jj}}} \tag{21}$$

With the following abbreviated notation

$$\varphi[ii] = \sum_{jj \neq ii}^{P} \beta[ii,jj]\, D_{ii,jj} \tag{22}$$

the partition matrix

$$u_{st}[ii] = \frac{1}{\sum_{j=1}^{c} \frac{d_{st}^2[ii] + \varphi[ii]}{d_{jt}^2[ii] + \varphi[ii]}} \tag{23}$$

For the prototypes, we complete calculations of the gradient of Q with respect to the coordinates of the prototype $\mathbf{v}[ii]$ and the solve the following system of equations

$$\frac{\partial Q[ii]}{\partial v_{st}[ii]} = 0,\ \ s = 1, 2, .., c; t = 1, 2, ..n \tag{24}$$

We obtain

$$\frac{\partial Q[ii]}{\partial v_{st}[ii]} = 2\sum_{k=1}^{N} u_{sk}^2[ii](x_{kt} - v_{st}[ii]) + 2\sum_{jj \neq ii}^{P} \beta[ii,jj] \sum_{k=1}^{N} u_{sk}^2[ii](v_{st}[ii] - v_{st}[jj]) = 0 \tag{25}$$

Next

$$v_{st}[ii]\left(\sum_{jj \neq ii}^{P} \beta[ii,jj] \sum_{k=1}^{N[ii]} u_{sk}^2[ii] - \sum_{k=1}^{N[ii]} u_{sk}^2[ii]\right) = \sum_{jj \neq ii}^{P} \beta[ii,jj] \sum_{k=1}^{N[ii]} u_{sk}^2[ii] v_{st}[jj] - \sum_{k=1}^{N[ii]} u_{sk}^2[ii] x_{kt} \tag{26}$$

Finally we get

$$v_{st}[ii] = \frac{\sum_{jj \neq ii}^{P} \beta[ii,jj] \sum_{k=1}^{N[ii]} u_{sk}^2[ii] v_{st}[jj] - 2\sum_{k=1}^{N[ii]} u_{sk}^2[ii] x_{kt}}{\sum_{jj \neq ii}^{P} \beta[ii,jj] \sum_{k=1}^{N[ii]} u_{sk}^2[ii] - \sum_{k=1}^{N[ii]} u_{sk}^2[ii])} \tag{27}$$

An interesting application of vertical clustering occurs when dealing with huge data sets. Instead of clustering them in a single pass, we split them into individual data sets, cluster each of them separately and actively reconcile the results through the collaborative exchange of prototypes.

8.3 The Assessment of the Strength of Collaboration

The choice of a suitable level of collaboration realized between the datasets through clustering denoted by β[ii,jj] deserves attention. Too high values of collaboration coefficient may lead to some instability of collaboration. Too low values of this coefficient may produce a very limited effect of collaboration that could be eventually made almost nonexistent in this manner. Generally speaking, the values of the collaboration coefficient could be asymmetric that is $\beta[\mathrm{ii},\mathrm{jj}] \neq \beta[\mathrm{jj},\mathrm{ii}]$. This is not surprising: we might have a case where at the level of dataset "ii" we are eager to collaborate and quite seriously accept findings coming from what has been discovered at dataset "jj" while the opposite might not be true. As the values of the collaboration coefficients could be different for any pair of datasets, the optimization of their values could be quite demanding and computationally intensive. To alleviate these shortcomings, let us express the coefficient $\beta[\mathrm{ii},\mathrm{jj}]$ as the following product $\beta[\mathrm{ii},\mathrm{jj}] = \omega\ \mathrm{f}(\mathrm{ii},\mathrm{jj})$ meaning that we view it as a function of the specific datasets under collaboration calibrated by some constant ω (>0) whose value does not depend upon the indexes of the datasets. The choice of the function f(ii, jj) can be done in several different ways. In general, we can envision the following intuitive requirement: if the structure revealed at the site of the jj-th dataset is quite different from the one present at the ii-th data set, the level of collaboration could be set up quite low. If there is a high level of agreement between the structure revealed at the jj-th set with what has been found so far at the ii-th dataset, then f(ii, jj) should assume high values. Given these guidelines, we propose the following form of f(ii, jj)

$$\mathrm{f}(\mathrm{ii},\mathrm{jj}) = 1 - \frac{\mathrm{Q}[\mathrm{ii}|\,\mathrm{jj}]}{\mathrm{Q}[\mathrm{ii}] + \mathrm{Q}[\mathrm{ii}|\,\mathrm{jj}]} \tag{28}$$

Here Q[ii] denotes a value of the objective function obtained for clustering without any collaboration (viz. the partition matrix and the prototypes are formed on the basis of optimization realized for $\mathbf{X}$[ii] only). Q[ii|jj] denotes the value of the objective function computed for the prototypes obtained for $\mathbf{X}$[jj] (without any collaboration) and used for data in $\mathbf{X}$[ii]; refer to Figure 8.

In essence, the values of Q[ii] and Q[ii|jj] reflect the compatibility of the structures in the corresponding data sets and in this manner tell us about a possible level of successful collaboration. The prototypes obtained for the dataset "jj" being used to determine the value of the objective function for the ii-th dataset could lead to quite comparable values of the objective function if the structure in $\mathbf{X}$[jj] resembles the structure of $\mathbf{X}$[ii]. In this case we envision

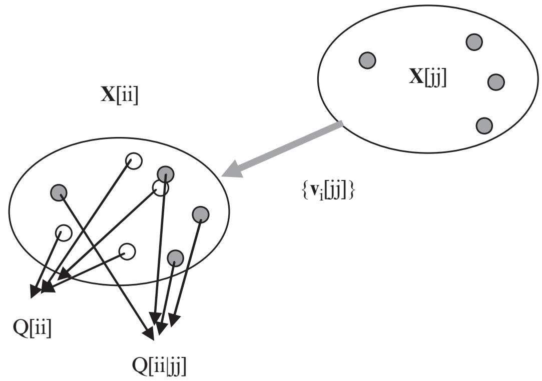

Fig. 8. Computing the values of Q[ii|jj] realized on a basis of the prototypes computed for X[jj]

Q[ii] < Q[ii|jj] yet Q[ii] ≈ Q[ii|jj]. On the other hand, if the structure in **X**[jj] is very different meaning that Q[ii|jj] >> Q[ii], the collaborative impact from what has been established for **X**[jj] could not be very advantageous. If Q[ii|jj] is close to Q[ii], f(ii, jj) approaches 1/2. In the second case, Q[ii|jj] >> Q[ii], the values of f(ii, jj) are close to zero.

Following the process described above, we are left now with a single coefficient (ω) controlling all collaborative activities for all datasets. This is far more practical yet its value needs to be properly selected. Here several alternatives could be sought:

(a) One could monitor the values of the overall objective function (1) during the course of optimization (minimization). The plot of the minimized objective function could be helpful here. The oscillations and a lack of convergence in the successive values of the objective function might strongly suggest that the values of ω are too high (too tight and intensive collaboration) and need to be reduced to assure smooth interaction between the datasets.
(b) We could also look at the differences between the results obtained without collaboration and with collaboration. For instance, a difference between the proximity matrices formed on a basis of the partition matrices constructed for the same dataset **X**[ii] without collaboration and with collaboration could serve as an indicator of the differences between the results. Such differences could be constrained by allowing only for some limited changes caused by the collaboration.

8.4 Dealing with Different Level of Granularity in the Collaboration Process

So far, we have made a strong assumption about the same number of clusters being formed at each individual dataset. This conjuncture could well be valid in many cases (as we consider collaboration realized at the same level of information granularity). It could be also quite inappropriate to made in some other cases. To cope with this problem, we need to move the optimization activities at the higher conceptual level by comparing results of clustering at the level of the proximity matrices. As indicated, when operating at this level of abstraction we are relieved from making any assumption about the uniform level of granularity occurring across all constructs.

9 Consensus–Based Fuzzy Clustering

In contrast to the collaborative clustering in which there is an ongoing active involvement of all datasets and the clustering algorithms running for individual datasets are impacted by the results developed at other sites, consensus-based clustering focuses mainly on the reconciliation of the individually developed structures. In this sense, building consensus is concerned with the formation of structure on the basis of the individual results of clustering developed separately (without any interaction) at the time of running the clustering algorithm. In this section, we are concerned with a collection of clustering methods being run on the same dataset. Hence U[ii], U[jj] stand here for the partition matrices produced by the corresponding clustering method. The essential step is concerned with the determination of some correspondence between the prototypes (partition matrices) formed for by each clustering method. Since there has not been any interaction when building clusters, there are no linkages between them once the clustering has been completed. The determination of this correspondence is an NP complete problem and this limits the feasibility of finding an optimal solution. One way of alleviating this problem is to develop consensus at the level of the partition matrix and the proximity matrices being induced by the partition matrices associated with other data. The use of the proximity matrices helps eliminate the need to identify correspondence between the clusters and handle the cases where there are different numbers of clusters used when running the specific clustering method.

The overall development process of consensus forming is accomplished in the following manner. Given the partition matrix U[ii], U[jj], etc. being developed individually, let us focus on the building consensus focused on U[ii]. Given the information about the structure coming in the form of U[ii] and other partition matrices U[jj], jj $\neq$ ii, the implied consensus-driven partition matrix ~U[ii] comes as a result of forming a sound agreement between the original partition matrix U[ii]. In other words, we would like to make ~U[ii] to be as close as possible to U[ii]. The minimization of the distance of the form $\|U[ii] - {\sim}U[ii]\|^2$ could be a viable optimization alternative. There are some other sources of structural information, see Figure 9. Here, however, we cannot establish a direct relationship between U[ii] (and ~U[ii]) and U[jj] given the reasons outlined before. The difficulties of this nature could be alleviated by

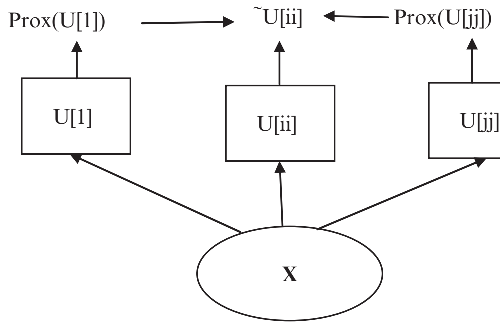

Fig. 9. A development of consensus-based clustering; the consensus building is focused on the partition matrix generated by the ii-th clustering method, U[ii]; here Prox(U[ii]) denotes a matrix of proximity values

considering the corresponding induced proximity matrices, say Prox(U[jj]). It is worth noting that a way in which the proximity matrix has been formed relieves us from the correspondence between the rows of the partition matrices (fuzzy clusters) and the number of clusters. In this sense, we may compare Prox (~U[ii]) and Prox (U[jj]) and searching for consensus by minimizing the distance $\|\text{Prox}(\sim\text{U[ii]})\text{-Prox}(\text{U[jj]})\|^2$. Considering all sources of structural information, the consensus building can be translated into the minimization of the following optimization problem

$$\|\text{U[ii]} - \sim\text{U[ii]}\|^2 + \gamma \sum_{\text{jj} \neq \text{ii}}^{\text{P}} \| \text{Prox}(\text{U[jj]}) - \text{Prox}(\sim\text{U[ii]})\|^2 \qquad (29)$$

The two components are reflective of the two essential sources of information about the structure. The positive weight factor (γ) is aimed at striking a sound compromise between the partition matrix U[ii] associated with the ii-th dataset. The result of the consensus reached for the ii-th method is the fuzzy partition matrix ~U[ii] minimizing the above performance index (29).

10 Concluding Notes

In this study, we emphasized the need for a revision of the paradigm of fuzzy clustering by augmenting it by the mechanisms of domain knowledge into its algorithmic layer. We have presented and discussed the key open issues and associate those to some evident challenges lying ahead in the progression of the discipline. Likewise we showed some pertinent links and outlined some promising avenues of algorithmic developments that might support the design of required conceptual platforms and specific detailed solutions.

We offered a number of algorithmic developments including clustering with partial supervision, collaborative clustering and clustering aimed at building consensus. In all of these we emphasized the role of human-centricity of the clustering framework and a distributed character of the available data. It has been shown how the issues of data privacy and security are alleviated through the use of granular information being inherently associated with the format of results generated in the process of clustering.

Acknowledgments

Support from the Canada Research Chair (CRC) program and the Natural Sciences and Engineering Research Council (NSERC) is gratefully acknowledged.

References

[1] Abonyi, J. and Szeifert, F. (2003). Supervised fuzzy clustering for the identification of fuzzy classifiers, *Pattern Recognition Letters*, 24, 14, 2195–2207.
[2] Agarwal, R. and Srikant, R. (2000). Privacy-preserving data mining. In: *Proc. of the ACM SIGMOD Conference on Management of Data*. ACM Press, New York, May 2000, 439–450.
[3] Bensaid, A. M., Hall, L. O., Bezdek, J. C. and Clarke L. P. (1996). Partially supervised clustering for image segmentation, *Pattern Recognition*, 29, 5, 859–871.
[4] Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, NY.
[5] Claerhout, B. and DeMoor, G.J.E. (2005). Privacy protection for clinical and genomic data: The use of privacy-enhancing techniques in medicine, *Int. Journal of Medical Informatics*, 74, 2–4, 257–265.
[6] Clifton, C. (2000). Using sample size to limit exposure to data mining, *Journal of Computer Security* 8,4, 281–307.
[7] Clifton, C. and Marks, D. (1996). Security and privacy implications of data mining. In: *Workshop on Data Mining and Knowledge Discovery*, Montreal, Canada, 15–19.
[8] Clifton, C. and Thuraisingham, B. (2001). Emerging standards for data mining, *Computer Standards* & *Interfaces*, 23, 3, 187–193.
[9] Coppi, R. and D'Urso, P. (2003). Three-way fuzzy clustering models for LR fuzzy time trajectories, *Computational Statistics* & *Data Analysis*, 43, 2, 149–177.
[10] Da Silva, J. C., Giannella, C., Bhargava, R., Kargupta, H. and Klusch, M. (2005). Distributed data mining and agents, *Engineering Applications of Artificial Intelligence,* 18, 7, 791–807.

[11] Du, W., Zhan, Z. (2002). Building decision tree classifier on private data. In: Clifton, C., Estivill-Castro, V. (Eds.), *IEEE ICDM Workshop on Privacy, Security and Data Mining, Conferences in Research and Practice in Information Technology*, vol. 14, Maebashi City, Japan, ACS, pp. 1–8.

[12] Evfimievski, A., Srikant, R., Agrawal, R. and Gehrke, J. (2004). Privacy preserving mining of association rules, *Information Systems*, 29, 4, 343–364.

[13] Johnsten, T. and Raghavan V.V. (2002). A methodology for hiding knowledge in databases. In: Clifton, C., Estivill-Castro, C. (Eds.), *IEEE ICDM Workshop on Privacy, Security and Data Mining, Conferences in Research and Practice in Information Technology*, vol. 14. Maebashi City, Japan, ACS, pp. 9–17.

[14] Kargupta, H., Kun, L., Datta, S., Ryan, J. and Sivakumar, K. (2003). Homeland security and privacy sensitive data mining from multi-party distributed resources, *Proc.* 12^{th} *IEEE International Conference on Fuzzy Systems, FUZZ '03,*. Volume 2, 25–28 May 2003, vol.2, 1257–1260.

[15] Kersten, P.R. (1996). Including auxiliary information in fuzzy clustering, Proc. *1996 Biennial Conference of the North American Fuzzy Information Processing Society, NAFIPS*, 19–22 June 1996, 221 –224.

[16] Lindell, Y. and Pinkas, B. (2000). Privacy preserving data mining. In: *Lecture Notes in Computer Science*, vol. 1880, 36–54.

[17] Liu, H. and Huang, S.T. (2003). Evolutionary semi-supervised fuzzy clustering, *Pattern Recognition Letters*, 24, 16, 3105–3113.

[18] Merugu, S and Ghosh, J. (2005).A privacy-sensitive approach to distributed clustering, *Pattern Recognition Letters,* 26, 4, 399–410.

[19] Park, B. and Kargupta, H. (2003). Distributed data mining: algorithms, systems, and applications. In: Ye, N. (Ed.), *The Handbook of Data Mining*. Lawrence Erlbaum Associates, N. York, 341–358.

[20] Pedrycz, W. (1985). Algorithms of fuzzy clustering with partial supervision, *Pattern Recognition Letters*, 3, 1985, 13–20.

[21] Pedrycz, W. and Waletzky, J. (1997). Fuzzy clustering with partial supervision, *IEEE Trans. on Systems, Man, and Cybernetics*, 5, 787–795.

[22] Pedrycz, W. and Waletzky, J. (1997). Neural network front-ends in unsupervised learning, *IEEE Trans. on Neural Networks*, 8, 390–401.

[23] Pedrycz, W., Loia, V. and Senatore, S. (2004). P-FCM: A proximity-based clustering, *Fuzzy Sets* & *Systems*, 148, 2004, 21–41.

[24] Pedrycz, W. (2002). Collaborative fuzzy clustering, *Pattern Recognition Letters*, 23, 14, 1675–1686.

[25] Pedrycz, W. (2005). *Knowledge-Based Clustering: From Data to Information Granules*, J. Wiley, N. York.

[26] Pinkas, B. (2002). Cryptographic techniques for privacy-preserving data mining. *ACM SIGKDD Explorations Newsletter* 4, 2, 12–19.

[27] Strehl, A. and Ghosh, J. (2002). Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3, 583–617.

[28] Timm, H., Klawonn, F. and Kruse, R. (2002). An extension of partially supervised fuzzy cluster analysis, *Proc. Annual Meeting of the North American Fuzzy Information Processing Society, NAFIPS 2002*, 27–29 June 2002, 63–68.
[29] Tsoumakas, G., Angelis, L. and Vlahavas, I. (2004). Clustering classifiers for knowledge discovery from physically distributed databases, *Data & Knowledge Engineering*, 49, 3, 223–242.
[30] Verykios, V.S., Bertino, E., Fovino, I.N., Provenza, L.P., Saygin, Y. and Theodoridis Y. (2004). State-of-the-art in privacy preserving data mining. *SIGMOD Record* 33, 1, 50–57.
[31] Wang K., Yu, P.S. and Chakraborty, S. (2004). Bottom-up generalization: a data mining solution to privacy protection, *Proc. 4^{th} IEEE International Conference on Data Mining, ICDM 2004*, 1–4 Nov. 2004, 249–256
[32] Wang, S.L. and Jafari, A. (2005). Using unknowns for hiding sensitive predictive association rules, *Proc. 2005 IEEE International Conference on Information Reuse and Integration*, 223–228.
[33] Wang, E.T., Lee, G. and Lin, Y. T. (2005). A novel method for protecting sensitive knowledge in association rules mining, *Proc. 29^{th} Annual International Computer Software and Applications Conference (COMPSAC 2005)*, vol. 2, 511–516.
[34] Zadeh, L. A. (2005). Toward a generalized theory of uncertainty (GTU) – an outline, *Information Sciences*, 172, 1–2, 1–40.