Computer Assisted Mechanics and Engineering Sciences, 9: 87–104, 2002.

Distributed collaborative knowledge elicitation

Witold Pedrycz
Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada, T6R 2G7
and
Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland

(Received August 6, 2001)

In this study, we develop an idea of knowledge elicitation realized over a collection of databases. The essence of such elicitation deals with a determination of common structure in databases. Depending upon a way in which databases are accessible abd can collaborate, we distinguish between a vertical and horizontal collaboration. In the first case, the databases deal with objects defined in the same attribute (feature) space. The horizontal collaboration takes place when dealing with the same objects but being defined in different attribute spaces and therefore forming separate databases.

We develop a new clustering architecture supporting the mechanisms of collaboration. It is based on a standard FCM (Fuzzy C-Means) method. When it comes to the horizontal collaboration, the clustering algorithms interact by exchanging information about local partition matrices. In this sense, the required communication links are established at the level of information granules (more specifically, fuzzy sets forming the partition matrices) rather than patterns directly available in the databases. We discuss how this form of collaboration helps meet requirements of data confidentiality. In case of the horizontal collaboration, the method operates at the level of the prototypes formed for each individual database. Numeric examples are used to illustrate the method.

Keywords: fuzzy clustering, collaboration, data confidentiality and security, data interaction, cluster (partition) interaction, vertical (data-based) and horizontal (feature-based) collaboration

1. Introductory comments

Undoubtedly, a distributed nature of data is inherent to most information systems. Various databases are constructed, used and maintained independent from each other. The challenge is to make sense of such distributed data at a global level. This task of data mining calls for a orchestrated effort and implies a highly collaborative nature of search for dependencies in data such that such findings are common to all databases (which is of genuine interest). To shed light on the spectrum of the processing problems and the ensuing collaborative mechanisms, we consider two interesting and general processing scenarios:

- Search for a common structure in databases Within a given organizational structure (company, network of sales offices, etc.), there are several local databases of customers (e.g., each supermarket generates its own database or a sales office maintains a local database of its customers). Generally, we can assume that all databases have the same attributes (features) while each database consists of different objects(patterns). To derive some common relationships that are common to all these databases, we allow the databases to collaborate at the level of the patterns. Quite commonly, we may not be allowed to have access to all databases but eventually but be provided by some general aggregates as illustrated in Fig. 1. Bearing this in mind, we can talk about *vertical* (data based) collaboration of knowledge elicitation.

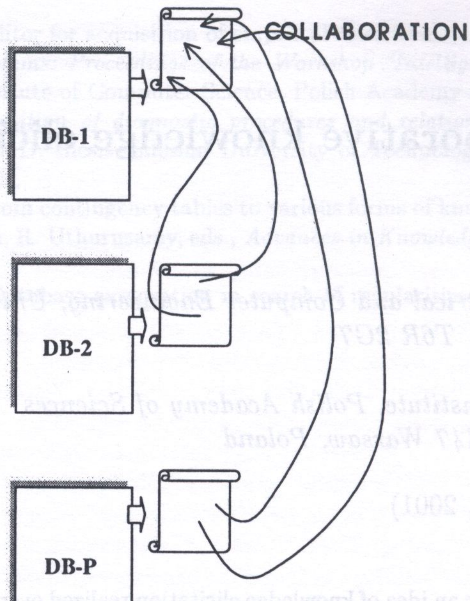

Fig. 1. Vertical collaboration between databases at a local level

- Security issues and discovery of data structures across different datasets. Consider that information about the same group of clients is collected in different databases while an individual company (bank, store, etc.) builds its own database. Because of confidentiality and security requirements, the companies cannot share information about clients in a direct manner. However all of them are interested in deriving some associations that help them learn about clients (namely their profiles and needs). As they are concerned with the same population of clients, we may anticipate that the basic structure of the population of such patterns, in spite of possible minor differences, should hold across all databases. The approach taken in this case would be to build clusters in each database and exchange information at the level of the clusters treated as information granules. In this way the security issues are not compromised while a sound mechanism of collaboration/ interaction between the databases could be established. Graphically, we can envision the situation of such collaboration as shown in Fig. 2. Evidently, in this case we are concerned with a *horizontal* (that is feature-based) collaboration in knowledge elicitation.

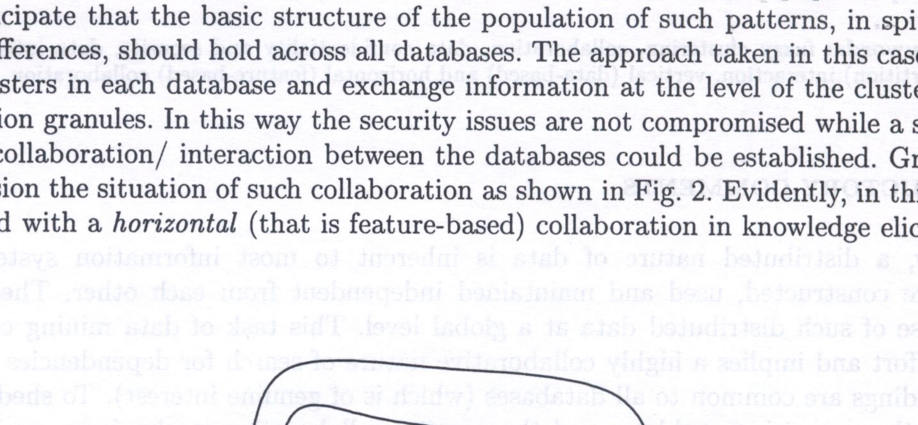

Fig. 2. Collaboration between databases at the level of "local" structures (clusters) discovered there; note that no direct collaboration at the data level is allowed

As knowledge elicitation is inherently user-oriented, we are interested in collaborative clustering as the results of clustering are information granules. In the sequel, this implies a certain type of collaboration as indicated before, namely a vertical collaborative clustering that involves databases involving various objects and horizontal clustering where we are faced with the same objects but characterized by various attributes.

As far as the algorithmic issues are concerned, the underlying idea of collaboration dwells on a well-known Fuzzy C-Means (FCM), cf. [2]. The reader may refer to pertinent details as to the generic method that is used as a canvass of the collaborative schemes developed in the study. In general, we can think of clustering [1, 3, 5, 6, 9, 10] as a vehicle of building information granules. It is also worth stressing that fuzzy clustering arose as a fundamental technique in construction of fuzzy models; refer e.g., to [4, 7, 8, 10–12].

The material is organized as follows. First, we proceed with the horizontal collaborative clustering by introducing all necessary notations, formulating the problem itself and discussing its algorithmic aspects. In the sequel, we use a number of numeric examples to illustrate the method. Second, we concentrate on the vertical clustering following the same scheme as in the first approach.

2. THE ALGORITHM OF HORIZONTAL COLLABORATIVE CLUSTERING

In this section, we introduce all necessary notation, formulate the underlying optimization problem implied by the objective function-based clustering and derive the solution in a form of an iterative scheme.

2.1. Notation

In what follows, we consider P subsets of data located in different spaces (viz. the patterns there are described by different features). As each subset concerns the same patterns (that is each pattern results as a concatenation of the corresponding subpatterns), the number of elements in each subset is the same and equal to N. We are interested in partitioning the data into c fuzzy clusters. The result of clustering completed for each subset of data comes in the form of a partition matrix and a collection of prototypes. We use a bracket notation to identify the specific subset. That is we use the notation $U[ii]$ and $\mathbf{v}[ii]$ to denote the partition matrix and the i-th prototype produced by the clustering realized for the ii-subset of data. Similarly, the dimensionality of the patterns in each subset could be different; to underline this we use a pertinent index, say $n[ii]$. The distance function between the i-th prototype and k-th pattern in the same subset is denoted by $d^2_{ik}[ii]$, $i = 1, 2, \ldots, c$, $k = 1, 2, \ldots, N$.

The objective function guiding the formation of the clusters that is completed for each subset assumes a well-known form as being encountered in the standard FCM algorithm

$$\sum_{k=1}^{N}\sum_{i=1}^{c} u^2_{ik}[ii]\, d^2_{ik}[ii],$$

$ii = 1, 2, \ldots, P$. The collaboration between the subsets is established through a matrix of connections (or interaction coefficients or interactions, for brief), see Fig. 3.

Each entry of the collaborative matrix states describes an intensity of the interaction. In general, $\alpha[ii, kk]$ assumes nonnegative values. The higher the value of the interaction coefficient, the stronger the collaboration between the corresponding subsets. To accommodate the collaboration effect in the optimization process, the objective function is expanded into the form

$$Q[ii] = \sum_{k=1}^{N}\sum_{i=1}^{c} u^2_{ik}[ii]\, d^2_{ik}[ii] + \sum_{\substack{jj=1\\ jj\neq ii}}^{P} \alpha[ii, jj] \sum_{k=1}^{N}\sum_{i=1}^{c} \{u_{ik}[ii] - u_{ik}[jj]\}^2\, d^2_{ik}[ii], \tag{1}$$

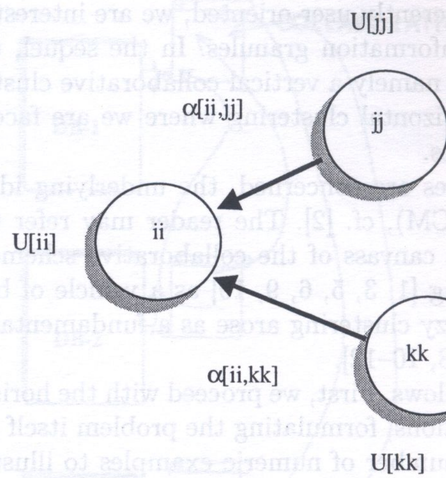

Fig. 3. Collaboration in the clustering scheme represented by the matrix of interactions between the subsets

$ii = 1, 2, \ldots, P$. The role of the second term in the above expression is to make the clustering based on the ii-th subset "aware" of other partitions. It becomes obvious that if the structures in datasets are similar then the differences between the partition matrices tend to be lower.

As usual, we require that the partition matrix satisfies 'standard" requirements of membership grades summing to 1 for each patterns and the membership grades contained in the unit interval. All in all, the collaborative clustering converts into the following family of P optimization problems with membership constraints,

$$\text{Min } Q[ii] \quad \text{subject to} \quad U[ii] \in U,$$

where U is a family of all fuzzy partition matrices, namely

$$U = \left\{ u_{ik}[ii] \in [0,1] \,\middle|\, \sum_{i=1}^{c} u_{ik}[ii] = 1 \text{ for all } k \text{ and } 0 < \sum_{k=1}^{N} u_{ik}[ii] < N \text{ for } i \right\}.$$

The minimization is carried out with respect to the fuzzy partition and the prototypes. This problem and its solution are discussed in detail in the ensuing section.

2.2. Optimization details of the collaborative clustering

The above optimization task splits into two problems, namely a determination of the partition matrix $U[ii]$ and the prototypes $\mathbf{v}_1[ii]$, $\mathbf{v}_2[ii]$, ..., $\mathbf{v}_c[ii]$. These problems are solved separately for each of the collaborating subsets of patterns.

To determine the partition matrix, we exploit a technique of Lagrange multipliers so that the constraint occurring in the problem becomes merged as a part of unconstraint optimization. This leads to the new objective function $V[ii]$,

$$V[ii] = \sum_{i=1}^{c} u_{ik}^2[ii]\, d_{ik}^2[ii] + \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \alpha[ii, jj] \sum_{i=1}^{c} \{u_{ik}[ii] - u_{ik}[jj]\}^2 d_{ik}^2[ii] - \lambda \left(\sum_{i=1}^{c} u_{ik}[ii] - 1 \right), \tag{2}$$

for each $k = 1, 2, \ldots, N$. where λ denotes a Lagrange multiplier. The necessary conditions leading to the local minimum of $V[ii]$ read as follows,

$$\frac{\partial V[ii]}{\partial u_{st}[ii]} = 0, \qquad \frac{\partial V[ii]}{\partial \lambda} = 0, \tag{3}$$

$s = 1, 2, \ldots, c$, $t = 1, 2, \ldots, N$. Solving these with respect to the unknown partition matrix we get

$$u_{st}[ii] = \frac{\varphi_{st}[ii]}{1+\psi[ii]} + \frac{1}{\sum_{j=1}^{c} \frac{d_{st}^2}{d_{jt}^2}} \left[1 - \sum_{j=1}^{c} \frac{\varphi_{jt}[ii]}{1+\psi[ii]} \right] \tag{4}$$

with the two coefficients expressed in the form

$$\varphi_{st}[ii] = \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \alpha[ii, jj]\, u_{st}[jj], \tag{5}$$

$$\psi[ii] = \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \alpha[ii, jj]. \tag{6}$$

In the calculations of the prototypes we confine ourselves to the Euclidean distance between the patterns and the prototypes. The necessary condition for the minimum of the objective function is of the form $\nabla_{\mathbf{v}[ii]} Q = 0$. The details are obvious yet the calculations are somewhat tedious. Finally, the resulting prototypes are equal to

$$v_{st}[ii] = \frac{A_{st}[ii] + C_{st}[ii]}{B_s[ii] + D_s[ii]}, \tag{7}$$

$s = 1, 2, \ldots, c$, $t = 1, 2, \ldots, n[ii]$, $ii = 1, 2, \ldots P$.

The coefficients in the above expression are as follows

$$A_{st}[ii] = \sum_{k=1}^{N} u_{sk}^2[ii]\, x_{kt}[ii], \tag{8}$$

$$B_s[ii] \sum_{k=1}^{N} u_{sk}^2[ii], \tag{9}$$

$$C_{st}[ii] = \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \alpha[ii, jj] \sum_{k=1}^{N} (u_{sk}[ii] - u_{sk}[jj])^2 x_{kt}[ii], \tag{10}$$

$$D_s[ii] = \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \alpha[ii, jj] \sum_{k=1}^{N} (u_{sk}[ii] - u_{sk}[jj])^2 \tag{11}$$

(note that $\mathbf{x}_k[ii]$ denotes a k-th pattern coming from the ii-th subset of patterns).

2.3. The overall clustering scheme

The general clustering scheme consists of two phases:

- generation of clusters without collaboration. This phase involves the use of the FCM algorithm for each subset of data. Obviously, the number of clusters needs to be the same for all datasets. During this phase we seek independently a structure in each subset of data
- collaboration of the clusters. Here we start with the already computed partition matrices, set up the collaboration level (through the values of the interaction coefficients arranged in $\alpha[ii, jj]$) and proceed with a simultaneous optimization of the partition matrices

Moving on to the formal algorithm, the computational details are organized in the following way:

Given: subsets of patterns $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_P$,

Select: distance function, number of clusters c, termination criterion, and collaboration matrix $\alpha[ii, jj]$.

Initiate randomly all partition matrices $U[1]$, $U[2]$, ..., $U[P]$

Phase I For each data

repeat

compute prototypes $\{\mathbf{v}_i[ii]\}$, $i = 1, 2, \ldots, c$, and partition matrices $U[ii]$ for all subsets of patterns

until a termination criterion has been satisfied.

Phase II *repeat*

For the given matrix of collaborative links $\alpha[ii, jj]$ compute prototypes and partition matrices $U[ii]$ using (4) and(7)

until a termination criterion has been satisfied.

The termination criterion relies on the changes to the partition matrices obtained in successive iterations of the clustering method, for instance a Tchebyschev distance could serve as a sound measure of changes in the partition matrices. Subsequently, when this distance is lower than an assumed threshold value ($\varepsilon > 0$), the optimization is terminated.

3. Quantification of the Collaborative Phenomenon of the Clustering

There are two levels of assessing a collaboration effect occurring between the clusters, namely the level of data and the level of information granules (that is fuzzy sets included in the partition matrix). In this latter quantification, we use the results of clustering without any collaboration as a point of reference.

The *level of data* involves a comparison carried out at the level of the numeric representatives of the clustering, that is the prototypes (centroids). The impact of the collaboration is then expressed in the changes of the prototypes occurring as a result of the collaboration.

At the *level of information granules* (partitions and fuzzy sets), the effect of collaboration is expressed in two ways as shown schematically in Fig. 4 where the collaboration involves two datasets (viz. $P = 2$) indicated by **1** and **2**. Similarly, by **1-ref** and **2-ref** we denote the results (partition matrices) resulting from the clustering carried out without any collaboration. First, we express how

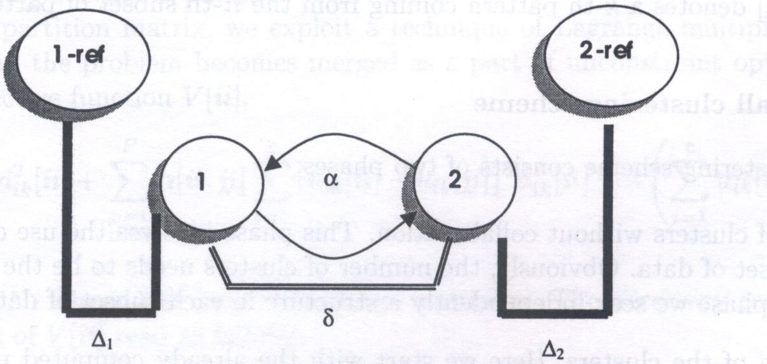

Fig. 4. Two ways of quantification of collaboration at the level of information granules; see a detailed description in text

close the two partition matrices are as a result of the collaboration. The pertinent measure reads as an average distance between the partition matrices $U_1 = [u_{ik}[\mathbf{1}]]$ and $U_2 = [u_{ik}[\mathbf{2}]]$, that is

$$\delta = \frac{1}{N * c} \sum_{k=1}^{N} \sum_{i=1}^{c} |u_{ik}[\mathbf{1}] - u_{ik}[\mathbf{2}]| \,. \tag{12}$$

Evidently, the stronger the collaboration (higher values of α), the lower the values of δ. In this sense, this index helps us translate the collaboration parameters (α) into the effective changes in the membership grades (that are the apparent final result of such interaction). The plot of δ regarded as a function of α is useful in revealing how the collaboration takes place. It tells how much the data subset is susceptible to the collaborative impact coming from the other subsets of patterns. For instance, no changes in the values of δ for increasing values of αs is an indicator of strong differences existing between the structures in the two datasets.

The second criterion takes into consideration the results of clustering obtained without any collaboration and treats this as a reference point. Using such partition matrices, we quantify how far the collaboration affects the results of clustering. For instance for the first data set we have

$$\Delta_1 = \frac{1}{N * c} \sum_{k=1}^{N} \sum_{i=1}^{c} |u_{ik}[\mathbf{1}] - u_{ik}[\mathbf{1\text{-}ref}]| \,. \tag{13}$$

For the second data subset we obtain

$$\Delta_2 = \frac{1}{N * c} \sum_{k=1}^{N} \sum_{i=1}^{c} |u_{ik}[\mathbf{2}] - u_{ik}[\mathbf{2\text{-}ref}]| \,. \tag{14}$$

While the above index exhibit a global character, one can investigate the changes at the level of the individual cluster and patterns. This local behavior of the collaboration is helpful in identifying elements whose membership grades are affected quite significantly as a result of collaboration and those whose structure is compatible across all datasets.

4. Experiments

In the series of numeric experiments, we use a Boston housing data available on the WWW, see `ftp://ftp.ics.uci.edu/pub/machine-learning-databases/housing/` . It consists of 506 patterns describing real estate in the Boston area. There are 14 features describing the patterns. These include crime rate, nitric acid concentration, median value of the house, just to name a few. We distinguish between two subsets of features where the first one can be treated as descriptors of social aspects of the data

A = {per capita crime rate by town, nitric oxides concentration (parts per 10 million), proportion of owner-occupied units built prior to 1940, weighted distances to five Boston employment centers, pupil-teacher ratio by town, % lower status of the population, median value of owner-occupied homes in \$1000's}

and

B = {proportion of residential land zoned for lots over 25,000 sqft, proportion of non-retail business acres per town, Charles River dummy variable (equal to 1 if tract bounds river; 0 otherwise), average number of rooms per dwelling, index of accessibility to radial highways, full-value property-tax rate per \$10,000, $1000(Bk - 0.63)^2$ where Bk is the proportion of blacks by town}

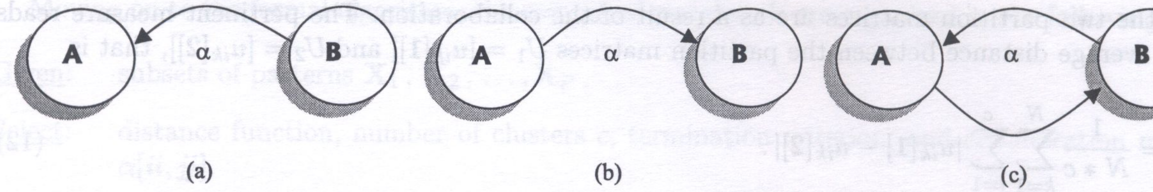

Fig. 5. Scenarios of collaborative clustering used in the experiments

In the following experiments we set up the number of the clusters to be equal to 5, $c = 5$. Several scenarios of collaboration are discussed; see Fig. 5 for its schematic notation. As only two subsets of data are involved, we drop indexes in the collaboration matrix; the meaning of collaboration becomes obvious from the context.

In all experiments we start with clustering that takes place without any collaboration (it was found that the number of iterations equal to 60 was enough to assure no changes to the partition matrices that is the optimization process could be deemed complete). At the next phase the collaboration takes place.

a. There is a collaborative link originating from **B** and affecting **A**. The values of this link (α) are set successively to 0.05, 0.1, 0.5 and 1. The values of the objective function are shown in Fig. 6; as expected the objective function assumes higher values for the increasing levels of collaboration (this is not surprising by noting that the collaboration component contributes additively as a part of this objective function). Noticeable are the drops in the values of the objective function occurring at the beginning of the optimization.

The resulting prototypes change once the collaboration assumes different intensity as shown below:

$\alpha = 0.5$

$$\mathbf{v}_1 = [\ 12.840698 \quad 0.675343 \quad 92.033089 \quad 1.984616 \quad 19.896008 \quad 21.301111 \quad 13.195952\ ]$$
$$\mathbf{v}_2 = [\ 0.291484 \quad 0.437043 \quad 32.688419 \quad 6.474721 \quad 16.880703 \quad 6.680616 \quad 27.994970\ ]$$
$$\mathbf{v}_3 = [\ 0.880803 \quad 0.528872 \quad 68.019592 \quad 3.802059 \quad 18.728121 \quad 12.382829 \quad 21.436878\ ]$$
$$\mathbf{v}_4 = [\ 0.639216 \quad 0.500246 \quad 60.230255 \quad 4.270203 \quad 17.784111 \quad 8.671939 \quad 26.960485\ ]$$
$$\mathbf{v}_5 = [\ 8.554115 \quad 0.661536 \quad 89.209915 \quad 2.277547 \quad 19.955612 \quad 17.302109 \quad 17.229208\ ]$$

$\alpha = 1.0$

$$\mathbf{v}_1 = [\ 13.185606 \quad 0.673498 \quad 91.076195 \quad 2.016629 \quad 19.926291 \quad 20.879652 \quad 13.254309\ ]$$
$$\mathbf{v}_2 = [\ 0.264656 \quad 0.435453 \quad 33.055351 \quad 6.535963 \quad 16.788097 \quad 6.537128 \quad 28.447708\ ]$$
$$\mathbf{v}_3 = [\ 0.759657 \quad 0.527652 \quad 67.316498 \quad 3.848079 \quad 18.747555 \quad 12.533010 \quad 21.190838\ ]$$
$$\mathbf{v}_4 = [\ 0.561581 \quad 0.500323 \quad 59.888401 \quad 4.340404 \quad 17.798523 \quad 8.690695 \quad 26.753359\ ]$$
$$\mathbf{v}_5 = [\ 9.424790 \quad 0.663229 \quad 88.969498 \quad 2.242356 \quad 19.995226 \quad 17.504406 \quad 17.110945\ ]$$

For comparative reasons, the prototypes of the subset A without any collaboration are listed as follows:

$$\mathbf{v}_1 = [\ 11.491062 \quad 0.688633 \quad 94.221016 \quad 1.930663 \quad 19.940283 \quad 21.444845 \quad 13.103884\ ]$$
$$\mathbf{v}_2 = [\ 0.394793 \quad 0.439771 \quad 31.897591 \quad 6.384313 \quad 17.012272 \quad 6.963907 \quad 27.159363\ ]$$
$$\mathbf{v}_3 = [\ 0.860573 \quad 0.489002 \quad 52.550468 \quad 4.605259 \quad 18.520782 \quad 9.673503 \quad 24.041653\ ]$$
$$\mathbf{v}_4 = [\ 1.307117 \quad 0.536930 \quad 75.181625 \quad 3.334807 \quad 17.237040 \quad 9.618564 \quad 27.298693\ ]$$
$$\mathbf{v}_5 = [\ 3.288866 \quad 0.601333 \quad 86.465401 \quad 2.696251 \quad 19.858582 \quad 15.527621 \quad 18.926182\ ]$$

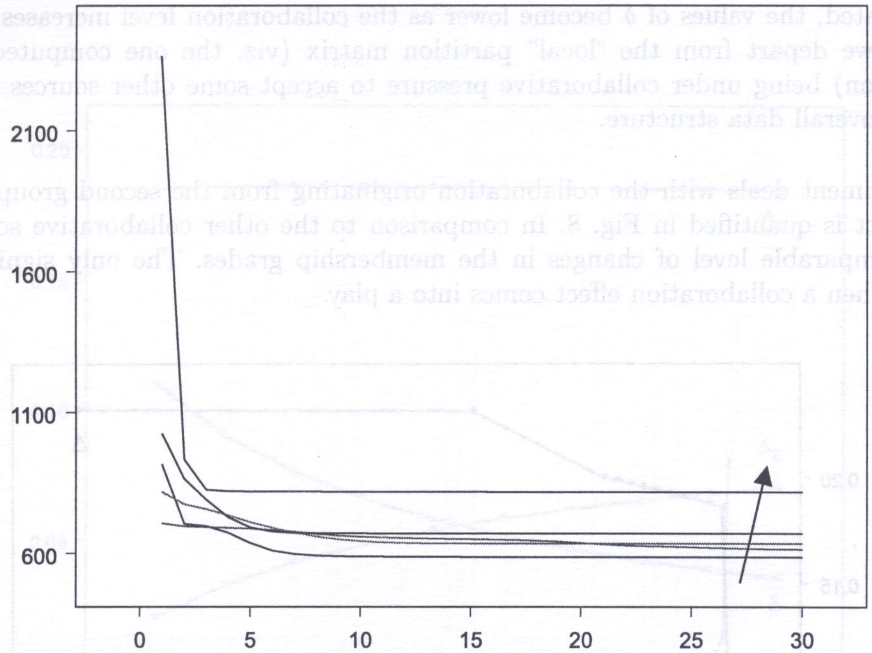

Fig. 6. The values of the objective function in successive iteration steps of the algorithm and selected values of the collaborative link

One can note that higher values of α lead to more evident translations of the prototypes in comparison to their original location when no collaboration took place. The collaboration effect can be quantified in the language of membership functions (partition matrices). Following the notation introduced in Section 3, the values of the indexes δ and Δ_1 are illustrated in Fig. 7.

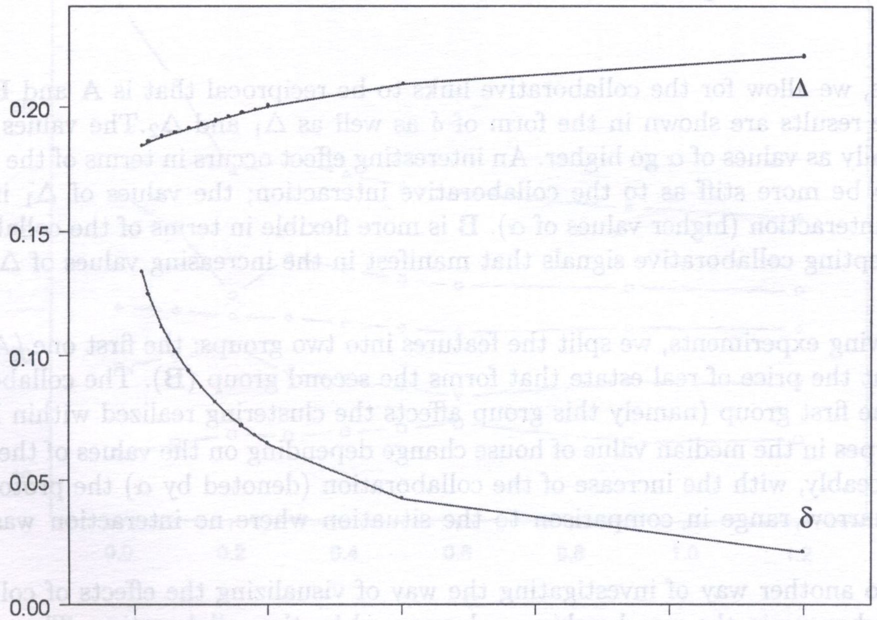

Fig. 7. Values of δ and Δ for selected values of α

As anticipated, the values of δ become lower as the collaboration level increases while Δ_2 gets higher as we depart from the "local" partition matrix (viz. the one computed without any collaboration) being under collaborative pressure to accept some other sources of information about the overall data structure.

b. This experiment deals with the collaboration originating from the second group. The collaboration effect is quantified in Fig. 8. In comparison to the other collaborative scheme, there is a quite comparable level of changes in the membership grades. The only significant jump is reported when a collaboration effect comes into a play.

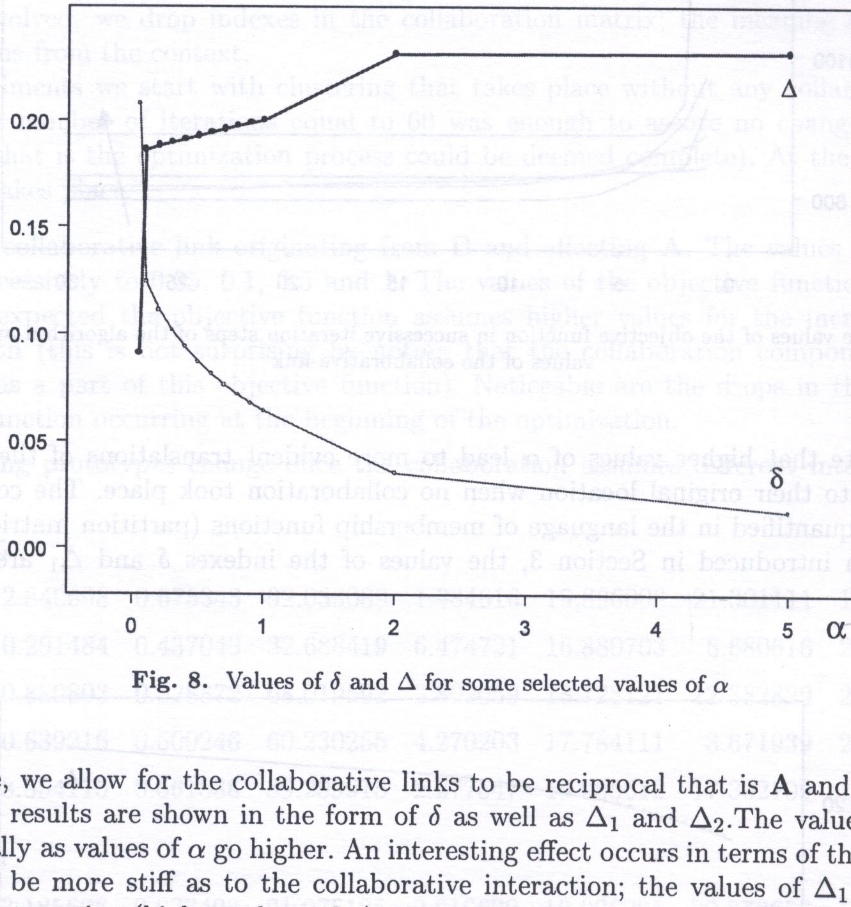

Fig. 8. Values of δ and Δ for some selected values of α

c. In this case, we allow for the collaborative links to be reciprocal that is **A** and **B** interact; see Fig. 9. The results are shown in the form of δ as well as Δ_1 and Δ_2.The values of δ go down monotonically as values of α go higher. An interesting effect occurs in terms of the collaboration: **A** tends to be more stiff as to the collaborative interaction; the values of Δ_1 in spite of the increasing interaction (higher values of α). **B** is more flexible in terms of the collaboration more readily accepting collaborative signals that manifest in the increasing values of Δ_2.

In the following experiments, we split the features into two groups: the first one (**A**) includes all the features but the price of real estate that forms the second group (**B**). The collaborative link is activated by the first group (namely this group affects the clustering realized within **B**).

The prototypes in the median value of house change depending on the values of the collaborative feedback. Noticeably, with the increase of the collaboration (denoted by α) the prototypes tend to occupy more narrow range in comparison to the situation where no interaction was present, see Fig. 10.

There is also another way of investigating the way of visualizing the effects of collaboration by looking at the changes in the membership grades caused by the collaboration. The changes in the membership grades occurring for two levels of collaboration are shown in Fig. 11.

Now we keep changing the number of clusters while retaining the same level of collaboration ($\alpha = 0.5$) to analyze how this affects the changes of δ and Δ.

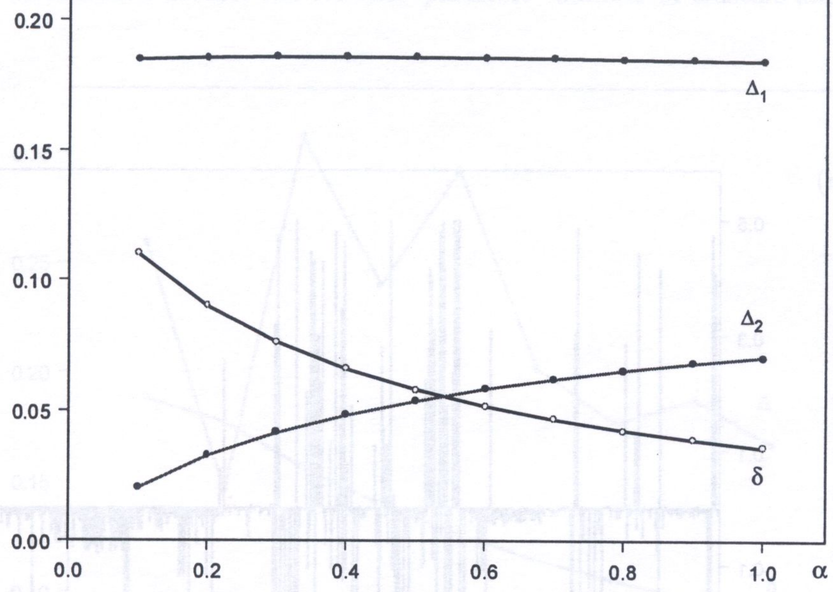

Fig. 9. Values of δ and Δ for selected values of α

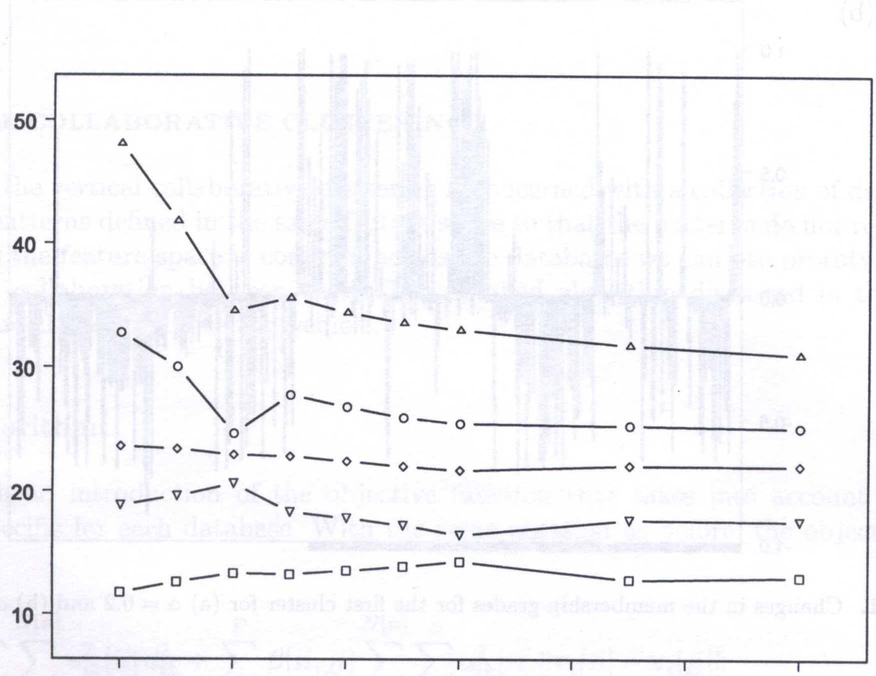

Fig. 10. Prototypes in the median value of real state as a function of the collaboration linkage α

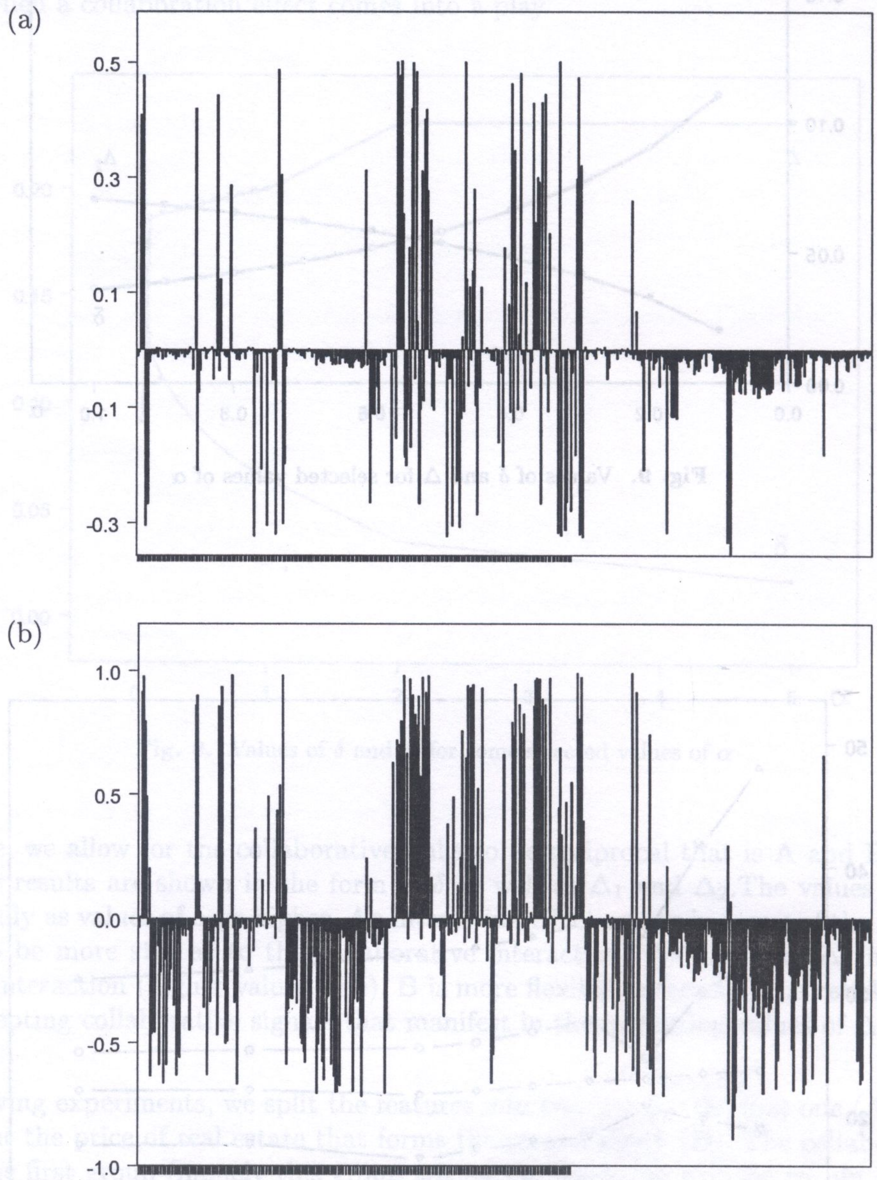

Fig. 11. Changes in the membership grades for the first cluster for (a) $\alpha = 0.2$ and (b) $\alpha = 0.5$

As expected, the values of δ go down with the increasing number of the clusters, Fig. 12. The reason for this trend is obvious: we get more clusters, the individual membership grades go down and the differences become lower. As to the second index, it less monotonic as with the changes of the number of clusters each dataset has its own "plausible" number of clusters and this could vary between them.

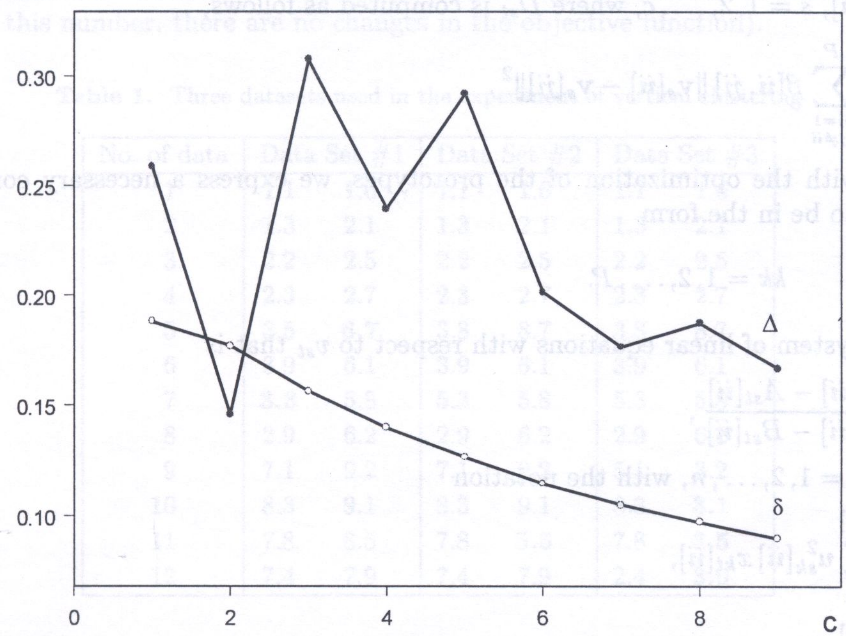

Fig. 12. δ and Δ as functions of the number of the clusters (c)

5. Vertical collaborative clustering

As discussed, the vertical collaborative clustering is concerned with a collection of databases involving different patterns defined in the same feature space so that the patterns do not repeat across the databases. As the feature space is common across the databases we can use prototypes as a means of facilitating collaboration between them. The detailed algorithm discussed in the next section concentrates on this communication vehicle.

5.1. The algorithm

We start with an introduction of the objective function that takes into account the vectors of prototypes specific for each database. With the same notation as before, the objective function is given as

$$Q[ii] = \sum_{i=1}^{c} \sum_{i=1}^{N[ii]} u_{ik}^2[ii]\, d_{ik}^2 + \sum_{\substack{jj=1 \\ jj\neq ii}}^{P} \beta[ii,jj] \sum_{k=1}^{N[ii]} \sum_{i=1}^{c} u_{ik}^2[ii]\, \|\mathbf{v}_i[ii] - \mathbf{v}_i[jj]\| \tag{15}$$

where $\beta[ii, jj]$ (> 0) describes a level of collaboration between the datasets and $\| \ \|$ denotes a distance function between the prototypes. The optimization of (15) is carried out for the partition matrix $U[ii]$ and the prototypes of the clusters $v[ii]$. This implies two separate optimization problems where the first one involving the partition matrix is subject to constraints. Not including all computational details, the final expression governing computations of the partition matrix reads in the form

$$u_{st} = \frac{1}{\sum_{j=1}^{c} \frac{D_{st}^2}{D_{jt}^2}}, \tag{16}$$

$t = 1, 2, \ldots, N[ii]$, $s = 1, 2, \ldots, c$, where D_{st} is computed as follows,

$$D_{st}^2 = d_{st}^2 + \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \beta[ii, jj] \, \|\mathbf{v}_s[ii] - \mathbf{v}_s[jj]\|^2 . \tag{17}$$

Proceeding with the optimization of the prototypes, we express a necessary condition for the minimum of Q to be in the form

$$\frac{\partial Q}{\partial \mathbf{v}_s[kk]} = 0, \qquad kk = 1, 2, \ldots, P. \tag{18}$$

This implies a system of linear equations with respect to v_{st} that is

$$v_{st}[ii] = \frac{F_{st}[ii] - A_{st}[ii]}{C_{st}[ii] - B_{st}[ii]}, \tag{19}$$

$s = 1, 2, \ldots, c$, $t = 1, 2, \ldots, n$, with the notation

$$A_{st}[ii] = \sum_{k=1}^{N[ii]} u_{sk}^2[ii] \, x_{kt}[ii],$$

$$B_{st}[ii] = \sum_{k=1}^{N[ii]} u_{sk}^2[ii],$$

$$C_{st}[ii] = \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \beta[ii, jj] \sum_{k=1}^{N[ii]} u_{sk}^2[ii],$$

$$F_{st}[ii] = \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \beta[ii, jj] \sum_{k=1}^{N[ii]} u_{sk}^2[ii] \, v_{st}[jj],$$

The overall computing scheme can be presented in the following fashion:

Given: subsets of patterns $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_P$ in the same feature space,

Select: distance function, number of clusters c, termination criterion, and collaboration matrix $\beta[ii, jj]$.

Initiate randomly all partition matrices $U[1]$, $U[2], \ldots, U[P]$

Phase I For each data

repeat

compute prototypes $\{\mathbf{v}_i[ii]\}$, $i = 1, 2, \ldots, c$, and partition matrices $U[ii]$ for all subsets of patterns

until a termination criterion has been satisfied.

Phase II *repeat*

For the given matrix of collaborative links $\beta[ii, jj]$ compute prototypes and partition matrices $U[ii]$ using (19) and (16)

until a termination criterion has been satisfied.

6. EXPERIMENTS

To illustrate how the method of this collaborative clustering works, we consider three collections of two-dimensional synthetic data collected in Table 1 where we identify the numbers of data points. The elements different from one data set to another are indicated in boldface. We partition the data into 3 clusters. The number of iterations the clustering algorithm has been run is equal to 15 (practically, at this number, there are no changes in the objective function).

Table 1. Three datasets used in the experiment of vertical clustering

No. of data	Data Set #1		Data Set #2		Data Set #3	
1	1.1	1.6	1.1	1.6	1.1	1.6
2	1.3	2.1	1.3	2.1	1.3	2.1
3	2.2	2.5	2.2	2.5	2.2	2.5
4	2.3	2.7	2.3	2.7	2.3	2.7
5	3.5	**6.7**	3.8	8.7	3.8	8.7
6	3.9	6.1	3.9	6.1	3.9	6.1
7	**3.3**	5.8	5.3	5.8	5.3	5.8
8	2.9	6.2	2.9	6.2	2.9	6.2
9	7.1	9.2	7.1	9.2	5.1	3.2
10	8.3	9.1	8.3	9.1	8.3	**3.1**
11	7.8	8.5	7.8	5.5	7.8	**3.5**
12	7.4	7.9	7.4	7.9	2.4	**3.9**

For comparative reasons, we start with a scenario in which there is no collaboration. The resulting partition matrices and prototypes are listed below:

Partition matrix; first data set			Partition matrix; second data set			Partition matrix; third data set		
0.962885	0.029110	0.008005	0.025856	0.964681	0.009462	0.039863	0.032573	0.927565
0.987977	0.009657	0.002366	0.008925	0.988080	0.002995	0.017405	0.012870	0.969724
0.973869	0.021538	0.004593	0.014625	0.981004	0.004371	0.003631	0.002755	0.993615
0.948994	0.042514	0.008491	0.029609	0.961978	0.008414	0.009680	0.006951	0.983369
0.010942	0.977999	0.011059	0.708906	0.075178	0.215916	0.818173	0.091730	0.090097
0.013902	0.973516	0.012582	0.982872	0.008537	0.008591	0.975737	0.011179	0.013085
0.010637	0.983763	0.005600	0.780051	0.067610	0.152339	0.754797	0.161347	0.083856
0.015808	0.975499	0.008693	0.875552	0.078263	0.046185	0.913704	0.030640	0.055656
0.007394	0.024593	0.968013	0.054116	0.012098	0.933786	0.207793	0.492985	0.299222
0.006575	0.017805	0.975620	0.044340	0.012775	0.942884	0.013797	0.974664	0.011539
0.000675	0.002018	0.997307	0.284482	0.093742	0.621776	0.001866	0.996727	0.001407
0.009464	0.031015	0.959521	0.011821	0.002588	0.985591	0.183367	0.071699	0.744934

The prototypes of the three datasets as tabulated below, show significant differences between them. In particular, the second and third prototype vary a lot across the datasets:

Dataset #1 – prototypes	Dataset #2 – prototypes	Dataset #3 – prototypes
[1.718665 2.222599]	[1.746825 2.253552]	[1.908616 2.492797]
[3.399974 6.196972]	[4.024643 6.495806]	[3.865419 6.565224]
[7.655201 8.676819]	[7.545924 8.293676]	[7.655816 3.344443]

Now, let us set up a collaboration level equal to 1; more specifically, $\beta[ii,jj]=1.0$ for all $ii\neq jj$. The collaboration established in this way results in similar prototypes as quantified in the following table:

Dataset #1 – prototypes	Dataset #2 – prototypes	Dataset #3 – prototypes
[1.865954 2.338975]	[1.869684 2.344301]	[1.944257 2.400455]
[3.720608 6.258747]	[3.834615 6.340519]	[3.858967 6.242465]
[7.444662 7.334947]	[7.412808 7.155983]	[7.358653 6.138383]

Noticeably, the prototypes start exhibiting a strong resemblance across the data that is a visible indicator of the ongoing collaboration. The effect of collaboration driving the prototypes closer for each dataset translates into changes in membership grades of the individual data points. Computationally, the change is taken as the sum of absolute differences taken over all clusters that is

$$\sum_{i=1}^{c}|u_{ik}-u_{ik}(\text{no\_collaboration})|$$

with $u_{ik}(\text{no\_collaboration})$ denoting the membership grade of the k-th pattern in the i-th cluster in case no collaboration is present. This effect of collaboration is shown in Table 2. Immediately, we recognize that some patterns are quite strongly affected by the collaboration. Those are the patterns that are different between datasets. With the increasing values of β-s, the collaboration becomes more vigorous. Subsequently, the values of the changes in the membership grades are shown in Table 2. It can be seen that some of the patterns are heavily affected by the collaboration meaning that at these points the structure are quite distinct and there any reconciliation between them requires a substantial level of effort. These particular patterns are indicated in boldface.

In the sequel, a total change in the membership (Δ) determined as

$$\Delta=\sum_{k=1}^{N}\sum_{i=1}^{c}|u_{ik}-u_{ik}(\text{no\_collaboration})|$$

and now regarded as a function of β is summarized in Fig. 13. Again, there is a strong monotonic relationship between the level of this collaboration and the manifesting changes in the partition matrix; the detailed relationships vary between datasets (groups of data).

Table 2. Changes in the membership grades of the individual data points in three datasets for β=1.0

Pattern	Change in membership		
no.	(first dataset)	(second dataset)	(third dataset)
1	0.032886	0.035287	0.032886
2	0.022677	0.022392	0.022677
3	0.027390	0.014654	0.027390
4	0.043216	0.020496	0.043216
5	0.011151	0.024933	0.011151
6	0.029918	0.012801	0.029918
7	0.064940	**0.186115**	0.064940
8	0.100724	0.056257	0.100724
9	**0.398732**	**0.366174**	**0.398732**
10	**0.323950**	**0.287519**	**0.323950**
11	**0.274974**	**0.302670**	**0.274974**
12	**0.144609**	**0.186651**	**0.144609**

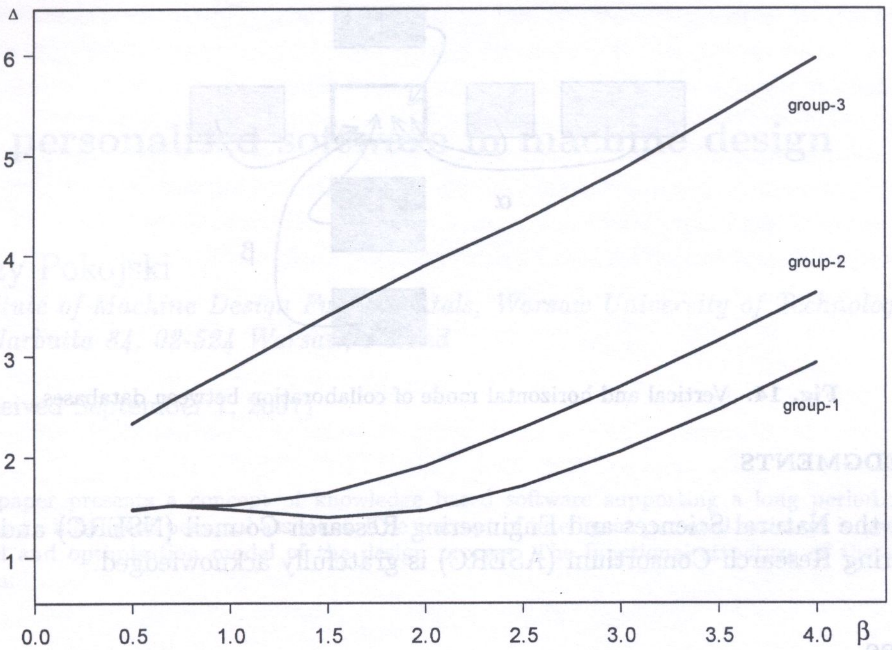

Fig. 13. Δ as a function of β for datasets used in the experiment

7. Conclusions

We have introduced an idea of collaborative processing, in general and collaborative clustering, in particular. It has been shown that a communication and collaboration between separate datasets can be effectively realized at the more abstract level of membership grades (partition matrices) and prototypes. Two types of collaboration (vertical and horizontal) were studied in detail. We provided a complete clustering algorithm by dwelling the method on the standard FCM method. The quantification of the collaboration effect can be realized either at the level of the prototypes or the partition matrices. An interesting expansion of the method discussed here involves a partial (limited) collaboration where not all patterns are available to form a collaborative link. This simply calls for an extra Boolean vector $\mathbf{b} = [b_1, b_2, \ldots, b_N]$ modifying the objective function in the form

$$Q[ii] = \sum_{k=1}^{N} \sum_{i=1}^{c} u_{ik}^2[ii]\, d_{ik}^2[ii] + \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \alpha[ii,jj] \sum_{k=1}^{N} \sum_{i=1}^{c} \{u_{ik}[ii] - u_{ik}[jj]\}^2\, b_k d_{ik}^2[ii]$$

where b_k assumes 1 when the k-th pattern is available for collaboration (otherwise b_k is set to 0).

In general, we can envision collaboration that takes place both at the vertical (data) as well as horizontal (feature) level, see Fig. 14. In terms of the objective function, this approach merges the two methods introduced before.

As a matter of fact, we can put down the following expression to emphasize the collaboration mechanism being in place,

$$U[ii] = F(\mathbf{U}[jj], \mathbf{v}[jj]),$$

where $\mathbf{U}$ and $\mathbf{v}$ are used to here denote the information feedback of the other part of the system (both vertical and horizontal).

The approach presented here could be easily generalized to support more specific ideas such as rule-based systems. In this case, we are concerned with the reconciliation of rules in each subset of data. Obviously, the optimization details need to be refined, as the specificity of the problem requires further in-depth investigations of a number of issues related to rules such as their specificity, consistency and completeness.

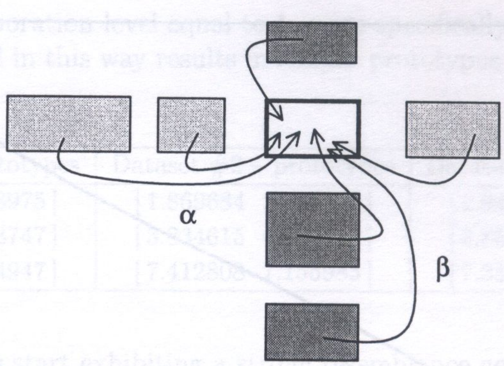

Fig. 14. Vertical and horizontal mode of collaboration between databases

ACKNOWLEDGMENTS

Support from the Natural Sciences and Engineering Research Council (NSERC) and Alberta Software Engineering Research Consortium (ASERC) is gratefully acknowledged.

REFERENCES

[1] M.R. Anderberg. *Cluster Analysis for Applications.* Academic Press, New York, 1973.
[2] J.C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms.* Plenum Press, New York, 1981.
[3] R.N. Dave. Characterization and detection of noise in clustering. *Pattern Recognition Letters*, **12**(11): 657–664, 1991.
[4] M. Delgado, F. Gomez-Skarmeta, F. Martin. A fuzzy clustering-based prototyping for fuzzy rule-based modeling. *IEEE Transactions on Fuzzy Systems*, **5**(2): 223–233, 1997.
[5] R.O. Duda, P.E. Hart, D.G. Stork. *Pattern Classification*, 2nd edition. J. Wiley, New York, 2001.
[6] J.C. Dunn. A fuzzy relative of the ISODFATA process and its use in detecting compact well-separated clusters, *J. of Cybernetics*, **3**(3): 32–57, 1973.
[7] F. Hoppner et al. *Fuzzy Cluster Analysis.* J. Wiley, Chichester, 1999.
[8] A. Kandel. *Fuzzy Mathematical Techniques with Applications.* Addison–Wesley, Reading, MA, 1986.
[9] P.R. Kersten. Fuzzy order statistics and their application to fuzzy clustering. *IEEE Transactions on Fuzzy Systems*, **7**(6): 708–712, 1999.
[10] W.Pedrycz. *Fuzzy Sets Engineering.* CRC Press, Boca Raton, FL, 1995.
[11] W. Pedrycz. Conditional fuzzy clustering in the design of radial basis function neural networks. *IEEE Transactions on Neural Networks*, **9**(4): 601–612, 1998.
[12] M. Setnes. Supervised fuzzy clustering for rule extraction. *IEEE Transactions on Fuzzy Systems*, **8**(4): 416–424, 2000.