International Journal of Automation and Computing 05(1), January 2008, 90-102
DOI: 10.1007/s11633-008-0090-3

Autonomous Clustering Using Rough Set Theory

Charlotte Bean[1] Chandra Kambhampati[2,*]

[1]Warwick Medical School Gibbet Hill Campus, University of Warwick, Coventry CV4 7AL, UK
[2]Department of Computer Science, University of Hull, Hull HU6 7RX, UK

Abstract: This paper proposes a clustering technique that minimizes the need for subjective human intervention and is based on elements of rough set theory (RST). The proposed algorithm is unified in its approach to clustering and makes use of both local and global data properties to obtain clustering solutions. It handles single-type and mixed attribute data sets with ease. The results from three data sets of single and mixed attribute types are used to illustrate the technique and establish its efficiency.

Keywords: Rough set theory (RST), data clustering, knowledge-oriented clustering, autonomous.

1 Introduction

Recent years have seen a rapid growth in the volume and complexity of electronic data being gathered and stored. As a result of this increase, the task of extracting meaningful knowledge in the form of patterns, relationships and groupings to be used in applications such as decision support, prediction and taxonomy has become arduous and essential. Furthermore, the need to discover underlying data structures in mixed attribute data calls for efficient data analysis with minimal human intervention.

Cluster analysis is one such technique that is used to reveal characteristics of underlying patterns in data. It extracts inherent groupings of homogeneous points from heterogeneous data and although there is no agreed benchmark definition for the terms "cluster", "class", and "group", they intuitively describe collections of data points with natural homogeneity. Agglomerative hierarchical clustering and iterative partitional clustering are two major categories of clustering algorithm that may be cast into a single algorithmic framework as shown in Figs. 1 and 2.

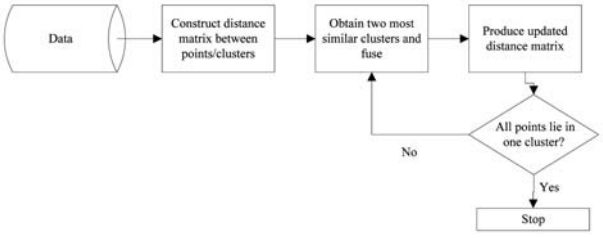

Fig. 1 Agglomerative hierarchical clustering framework

Agglomerative hierarchical clustering[1−10] imposes a hierarchical decomposition on a dataset through the iterative fusion of points and clusters, and a final clustering is determined according to some pre-determined cut-off criterion. Partitional algorithms, including k-means[5,7,8,11−16] and fuzzy c-means (FCM)[17−22], follow an iterative optimization strategy for partitioning a database into a predetermined number of clusters. The process is initialized by defining seed points or an initial partition, and the successive swapping of data points determines a locally optimal partition. The FCM methodology differs only in the sense that points are enabled to have a degree of membership to all clusters. Both categories of clustering technique have advantages and disadvantages. Hierarchical clustering has an obvious benefit in that it does not require the number of clusters to be determined a priori; however there is a trade-off in the need to select a termination point for the algorithm.

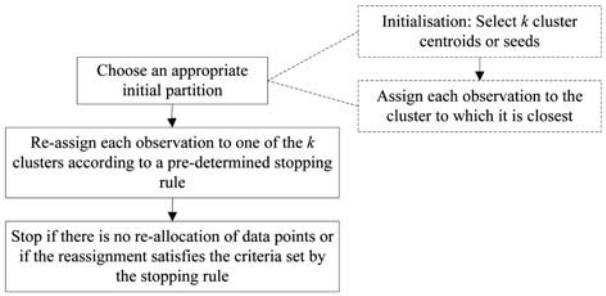

Fig. 2 Iterative partitional clustering framework

Although structurally varied, the two categories of clustering algorithm discussed above share the common property of relying on local data properties to reach an optimal clustering solution, which carries the risk of producing a distorted view of the data structure. Rough set theory (RST), introduced by Pawlak et. al.[23−26], moves away from this local dependence and focuses on the idea of using global data properties to establish similarity between the objects in the form of coarse and representative patterns. The rigorous framework of RST is provided by a well-defined indiscernibility relationship that classifies objects into classes on the basis of perceived differences from an initial knowledge source and its aim is not to perform exploratory data analysis, but to establish similarities that are evident in the raw data. In terms of its role as a set theoretical tool, RST is often compared to fuzzy set theory (FST) with the argument that the two are competing notions. Upon investigating this view, Dubois and Prade[27] suggested that they are in fact mathematical tools with different purposes. Whereas FST deals with the concept of "vagueness" in the boundary of a sub-class of a set, RST focuses on coarseness of knowledge within the set itself. It is this notion of coarseness teamed with the ability to obtain meaningful knowledge from uncertain and incomplete data that makes RST a valuable tool for extracting relationships from real-world data. Since both cluster analysis and RST form data

Manuscript received September 29, 2006; revised December 26, 2006
*Corresponding author. E-mail address: c.kambhampati@hull.ac.uk

groupings, the conceptual link between the techniques is evident[28,29]. However, the fact that cluster analysis is an exploratory tool used to reveal underlying groupings whereas RST imposes a partitioning structure on a dataset suggests that RST provides scope for discovering "possible" data clusterings with a view to assessing them on the basis of global information inherent in the data[30]. Early attempts at combining concepts from both techniques have led to a hierarchical type clustering algorithm called knowledge-oriented clustering[29], in which a modification procedure allows for the simplification of knowledge. This process of knowledge simplification is not incorporated in the traditional hierarchical clustering techniques.

This paper proposes an autonomous methodology for extracting knowledge and relationships from mixed attribute data in the form of coarse clusters that reflect important global properties of the data. The resultant clustering technique is presented as a simple algorithm, and the modified tools from RST are used to form the classes. By virtue of the fact that RST reflects global data properties, the clustering solution is unaffected by local discrepancies. Thus, it has the following advantages: 1) Avoiding the generation of too many small and unrepresentative clusters; 2) Leading to a coarse clustering of the universe. Furthermore, the reliance of the traditional clustering techniques on local optimality paves the way for a number of different clustering solutions and scope for distorted results.

The proposed algorithm also eliminates the need for subjectivity in obtaining a representative number of final clusters and an "optimal" clustering solution is determined according to the convergence of a well-defined accuracy measure. Procedures to determine data partitionings and cluster modifications are developed with an emphasis on minimizing the level of computational complexity to obtain optimal clusters efficiently.

The structure of this paper is organized as follows. Section 2 provides a preliminary overview of RST followed by an introduction to the knowledge-oriented algorithm in Section 3. This section incorporates a break-down of the generic clustering procedure and provides a detailed discussion of the key steps. Section 4 introduces the proposed autonomous knowledge-oriented clustering algorithm. Section 5 provides a detailed demonstration of the algorithm on real and generated data. The paper concludes in Section 6 with a summary and suggestions for future research.

2 Preliminaries

The popularity of RST as a tool for handling uncertainty in data has risen since its introduction in the early 1980s and it has been used successfully in a number of applications such as data mining, knowledge discovery, and decision making[23,28−30]. Its role in knowledge-oriented clustering will become apparent in the next section but a preliminary overview of the main rough set concepts will be given here.

Definition 1. Information system

An information system is defined as a family of sets $\mathbf{A} = (U, A)$ where U is a non-empty universe of objects and A is a finite non-empty set of attributes such that $\forall a \in A$, $a: U \rightarrow V_a$, where V_a is the value set of a.

Definition 2. Decision system

Let U be a finite universe of objects and A a finite set of attributes. A decision system is the family of sets $\mathbf{A} = (U, A \cup \{d\})$ such that $d \notin A$ is a decision attribute and members of A are referred to as condition attributes.

Definition 3. B-indiscernibility relation

Let $\mathbf{A} = (U, A)$ be an information system. Given a set of attributes $B \subseteq A$, classes are formed according to a B-indiscernibility relation.

$$Ind_A(B) = \{x, x' \in U : \forall a \in B, a(x) = a(x')\} \quad (1)$$

which induces a partitioning of the universe U according to the attribute set B. The resultant classes are known as indiscernibility classes $[x]_B$.

The B-indiscernibility relation is a mathematical equivalence relation that partitions U into a finite number of disjoint equivalence (indiscernibility) classes $[x]_B$ as depicted in Fig. 3.

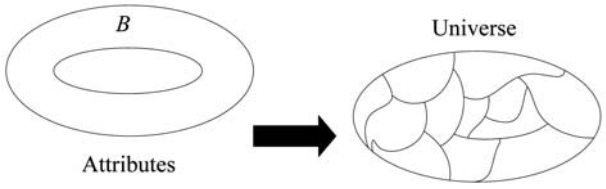

Fig. 3 Partitioning of a universe U

Thus, any set $X \subseteq U$ can be approximated solely on the basis of information in $B \subseteq A$ by constructing a B-lower approximation and B-upper approximation of X defined respectively as:

Definition 4. B-lower and upper approximations

Let $\mathbf{A} = (U, A)$ be an information system and $Ind_A(B)$ an indiscernibility relation placed on universe U with respect to the attribute set $B \subseteq A$. For a given set $X \subseteq U$ a B-lower approximation of X is defined as

$$\underline{B}X = \{x : [x]_B \subseteq X\} \quad (2)$$

and a B-upper approximation of X is defined as

$$\overline{B}X = \{x : [x]_B \cap X \neq \emptyset\}. \quad (3)$$

The lower approximation consists of objects that definitely belong to X and the upper approximation contains objects that possibly belong to X. Consequently, X is classified as a rough set if its B-boundary region, $BN_B(X) = \overline{B}X - \underline{B}X$, is non-empty. In other words, there is a region of uncertainty regarding set membership. This uncertainty may be quantified for individual points x by assessing the degree of overlap between the indiscernibility class $[x]_B$ and the rough set X. In this manner, classifications maintain a global sense of knowledge.

3 Knowledge-oriented clustering: Generic framework

The generic clustering framework shown in Figs. 1 and 2 illustrates how points are traditionally assigned to clusters according to the two categories of clustering. Although the two procedures are distinct in both their algorithmic construction and the premise upon which final clusters are obtained, both of them rely on the local data properties to refine clustering formations. In the context of hierarchical clustering, this is achieved with the calculation of distances between clusters whereas the use of local optimality in partitional clustering leads to the final clustering solution. Without doubt, the two techniques have achieved

success in a range of applications[5,7,10,16] but by extracting selected useful properties of the algorithms and teaming them with tools from RST, it is possible to overcome some of the drawbacks associated with these traditional methods through the use of knowledge-oriented (K-O) clustering. The algorithmic framework of K-O clustering is similar to that of agglomerative hierarchical clustering as shown in Fig. 1. However, the main clustering tool is a form of indiscernibility relation taken from RST. In using a simple algorithmic framework, K-O clustering is computationally efficient. Furthermore, the combined use of tools from hierarchical clustering and RST allows clusters to be formed using both local and global properties of the data. Fig. 4 shows the used algorithm.

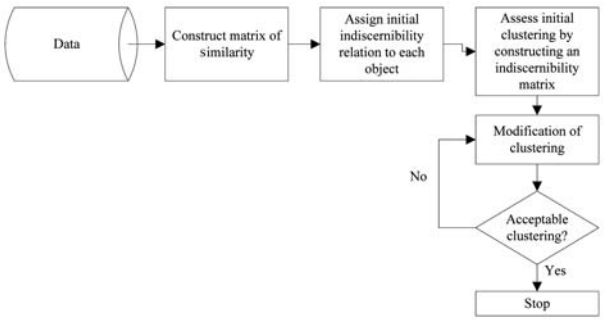

Fig. 4 The knowledge-oriented clustering algorithm

The above algorithm includes the following steps:

Step 1. Construct a matrix of similarities $S = \{s(x_i, x_j)\}$ between all pairs of objects.

Step 2. Assign an initial indiscernibility relation R_i to each object in the universe. Pool information to obtain an initial clustering U/R.

Step 3. Construct an indiscernibility matrix $\Gamma = \{\gamma(x_i, x_j)\}$ to assess the clustering U/R.

Step 4. Modify clustering according to a modified indiscernibility relation R_i^{mod} to gain a modified clustering U/R_{mod}.

Step 5. Repeat Steps 3 and 4 until a stable clustering is obtained.

The notions of similarity and indiscernibility will be introduced and discussed in Section 3.1. Section 3.2 details the idea of initial clustering in a generic knowledge-oriented clustering framework.

3.1 Similarity and indiscernibility

Knowledge-oriented clustering is under-pinned by the construction of two key symmetric matrices; similarity $S = \{s(x_i, x_j)\}$ and indiscernibility $\Gamma = \{\gamma(x_i, x_j)\}$. They respectively control the local and global extraction of knowledge used to obtain and modify clustering formations. The similarity matrix S is calculated once in the initialization stage of Step 1 of the algorithm whereas the indiscernibility matrix Γ is updated iteratively (see Step 3) until convergence to a final clustering solution is achieved. The recalculation of the indiscernibility matrix at each iteration reflects updated global knowledge of the data whereas the single similarity matrix displays inherent local distances between points.

The local properties of points depend on how similar they are to each other, thus the form of the similarity matrix S is dependent on the distance measure chosen to determine similarity $s(x_i, x_j)$ between pairs of objects. Most clustering algorithms are designed to deal solely with numerical attributes, however, many of the data collected consists of a mixture of both numerical and categorical attributes (e.g. medical data sets). Thus, there is a need for a measure which can take into account the mixed nature of the data. A combined similarity measure of the following form is suggested:

$$s(x_i, x_j) = \frac{k_{\text{num}}}{k}\left(1 - \frac{s_{\text{num}}(x_i, x_j)}{\max\limits_{i,j} s_{\text{num}}(x_i, x_j)}\right) + \frac{k_{\text{cat}}}{k}\left(1 - \frac{s_{\text{cat}}(x_i, x_j)}{\max\limits_{i,j} s_{\text{cat}}(x_i, x_j)}\right) \quad (4)$$

where x_i, x_j are objects in a universe U, $k = k_{\text{num}} + k_{\text{cat}}$ is the total number of attributes, s_{num} is the similarity measure for numerical data, and s_{cat} is the similarity measure for categorical data and also is essentially the Hamming distance.

The Hamming distance is an appropriate s_{cat} measure[31] but the choice of a suitable s_{num} measure is more difficult due to the nature of the data and the wide selection of possible measures. The Euclidean distance measure is well-established and popular, being used in a variety of statistical analyzes, and is a special case of the Minkowski metric. However, for the purpose of clustering, the fact that the Euclidean distance is scale-invariant can lead to distorted results. Although this can be rectified by standardizing the data, it should be remembered that this process can itself effect the clustering solution. Another alternative is to use the Mahalanobis distance. This measure takes into account the covariance structure of the attributes and acknowledges the fact that significant correlations between attributes may influence the final result. Again, this cannot be applied in all circumstances since it relies on the assumptions of normality and homoscedacity in the attributes. It has been suggested by Manly[32] that the Penrose measure is a more appropriate replacement for the Mahalanobis distance when dealing with data sets that have less than 100 degrees of freedom. In summary, the choice of an appropriate s_{num} measure is reliant on a number of factors including the size and application of the data as well as statistical properties and it must be chosen accordingly to the conditions of the given clustering problem.

Global knowledge of the data is represented as the proportion of points that regard each pair of points in the universe to be indiscernible. The information is displayed in the indiscernibility (or "gamma") matrix Γ which is constructed in Step 3 of the algorithm to assess a given clustering formation and induce modification if necessary. Its entries $\gamma(x_i, x_j)$ represent an indiscernibility degree[31] between each pair of objects x_i and x_j, such that $0 \leq \gamma(x_i, x_j) \leq 1$. The resultant indiscernibility matrix is defined as follows.

Definition 5. Indiscernibility matrix

Let $\mathbf{A} = (U, A)$ be an information system with non-empty finite universe $U = \{x_1, x_2, \cdots, x_n\}$ and attribute set $A = \{a_1, a_2, \cdots, a_k\}$. For a given clustering of the universe, the indiscernibility matrix $\Gamma = \{\gamma(x_i, x_j)\}$ represents the global proportion of objects that regard each pair of objects in the universe to be indiscernible, where the indiscernibility degree $\gamma(x_i, x_j)$ for each pair of objects is given

by

$$\gamma(x_i, x_j) = \frac{\sum_{k=1}^{|U|} \gamma_k^{\text{indis}}(x_i, x_j)}{\sum_{k=1}^{|U|} \gamma_k^{\text{indis}}(x_i, x_j) + \sum_{k=1}^{|U|} \gamma_k^{\text{dis}}(x_i, x_j)} \quad (5)$$

where

$$\gamma_k^{\text{indis}}(x_i, x_j) = \begin{cases} 1, & \text{if } x_k R_k x_i \text{ and } x_k R_k x_j \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

and

$$\gamma_k^{\text{dis}}(x_i, x_j) = \begin{cases} 1, & \text{if not}(x_i R_k x_j) \\ 0, & \text{otherwise} \end{cases} . \quad (7)$$

It should be noted that the notion of indiscernibility in this context is more general than the form outlined in Definition 3 and no longer satisfies every property of an equivalence relation. Definition 3 defines objects to be indiscernible if they possess identical attribute values, whereas the general form of indiscernibility (see Definition 6) allows objects to be regarded as indiscernible if their similarity value $s(x_i, x_j)$ exceeds some pre-determined threshold. With this idea, the relation R_k represents well-defined indiscernibility relation used to partition the universe into classes. $\gamma_k^{\text{indis}}(x_i, x_j)$ assesses indiscernibility between x_i and x_j. It takes the value 1 if x_i, x_j, and x_k all lie in the same indiscernibility class according to the relation R_k. The inclusion of object x_k acknowledges the fact that similarity is measured locally with respect to this point. Conversely, $\gamma_k^{\text{dis}}(x_i, x_j)$ is equal to 1 if x_i and x_j are discernible with respect to R_k. In other words, they do not lie in the same indiscernibility class according to relation R_k.

The success of knowledge-oriented clustering hinges on the information obtained from the similarity and indiscernibility matrices. In the first instance (Step 1), the similarity matrix S draws out local properties of the data in the form of raw distances between points. Since this knowledge forms the basis of the initial indiscernibility relation R_k used to gain a first partitioning of the universe, and since the initial partitioning should be optimal in the sense that a meaningful and representative clustering of the data is ultimately attainable, the selection of an appropriate similarity measure is crucial. On the other hand, the indiscernibility matrix Γ, calculated in Step 3 of the algorithm, displays global knowledge about the positioning of points in the universe which is then used to modify a given clustering into coarser and more meaningful clusters.

3.2 Initial clustering of the data

After initializing the knowledge-oriented clustering algorithm with the calculation of the similarity matrix S, it is necessary to obtain an initial clustering of the universe (see Step 2). This step is dependent on the local knowledge displayed in the similarity matrix and provides a quick overview of the clustering structure of the data, which can be later modified to form definitive clusters. The initial clustering should in some sense represent a best possible first clustering. However, it should be noted that this notion of optimality does not necessarily imply the initial clustering with the least number of clusters. Since clusters may be subsequently joined but not re-partitioned, it does not increase the computational burden to obtain a high number of initial clusters.

The initial clustering of the data is governed by key threshold parameters which must be chosen to ensure a true reflection of inherent clustering properties. Failure to do so will lead to a distorted final clustering. Specifically, a set of initial threshold values $\{Th_i\}_{i=1}^n$ is selected to correspond to a set of initial indiscernibility relations $\{R_i\}_{i=1}^n$ which are assigned to each object in the universe. The above modified form of indiscernibility allows two points to belong to the same indiscernibility class if the similarity value exceeds a pre-determined threshold.

Definition 6. Initial indiscernibility relation

Let $\mathbf{A} = (U, A)$ be an information system with non-empty finite universe $U = \{x_1, x_2, \cdots, x_n\}$ and attribute set $A = \{a_1, a_2, \cdots, a_k\}$. An initial indiscernibility relation R_i is assigned to each object in the universe as follows:

$$R_i = \{(x_i, x_j) \in U \times U : s(x_i, x_j) \geq Th_i, \quad j = 1, 2, \cdots, n\} \quad (8)$$

where $s(\cdot\ ,\ \cdot)$ is the similarity measure between two objects and Th_i is a derived initial threshold value for object x_i.

R_i induces a partition U/R_i of the universe for all $i = 1, \cdots, n$. Those objects that are similar to x_i ($P_i = \{x_j : x_i R_i x_j\}$) and those objects that are not similar to x_i ($U - P_i = \{x_j : \text{not}\,(x_i R_i x_j)\}$). After obtaining the initial set of partitions, $\{U/R_i : i = 1, 2, \cdots, n\}$, the information is pooled to obtain an overall initial partitioning of the universe U/R, referred to as the initial clustering. The way in which the partitionings $\{U/R_i\}_{i=1}^n$ are formed and, thus, the formation of the initial clustering U/R is highly dependent on the choice of the thresholds $\{Th_i\}_{i=1}^n$. Hirano and Tsumoto[31] made an attempt to set these initial threshold values autonomously using the notion of gradient level similarity. This was achieved by applying a form of Gaussian smoothing to their chosen similarity function to obtain derivative values. Threshold values were selected to correspond to comparably large similarity decreases. However, not only is this technique computationally intensive, but the notion of using interpolation to obtain derivative values results in a high degree of error, particularly in small data sets. A method to overcome these drawbacks in setting the initial threshold values is suggested in Section 4.

4 Knowledge-oriented clustering with autonomy

Knowledge-oriented clustering algorithms can be framed within a generic algorithmic framework shown in Fig. 4, but the efficiency and optimality of the algorithm is dependent on the selection of individual threshold parameters. Not only is this relevant to the initial clustering of the universe, but it is also true in the modification stages of the algorithm (see Step 4) where further threshold values determine updated partitionings of the universe. However, whereas traditional hierarchical clustering algorithms[1–10] rely on subjectivity to determine parameters, it is desirable to develop a set of well-defined procedures for setting the required thresholds autonomously at each stage of the knowledge-oriented clustering algorithm. Thus, the same (or a highly similar) clustering solution is ensured upon applying the algorithm through independent means to the same data. This section details such procedures within the generic framework outlined in Section 3. Section 4.1 introduces a method for obtaining a set of initial threshold values $\{Th_i\}_{i=1}^n$ which will lead to an optimal initial clus-

tering of the universe. The notion of cluster modification is discussed in Sections 4.2 and 4.3.

4.1 Autonomous initial clustering of the data

The initial clustering of the universe is a crucial stage in the knowledge-oriented clustering procedure. If it is done in an incorrect manner, the subsequent clusterings will not fully reflect inherent data structures, which will lead to a distorted and meaningless final clustering. Since the initial partitioning is achieved by imposing initial indiscernibility relations (8) on the data, which are themselves dependent upon selected threshold values $\{Th_i\}_{i=1}^n$, it is the setting of these thresholds that holds the key to a meaningful clustering of the universe. A method is suggested here to determine the initial thresholds autonomously while maintaining the key goal of computational efficiency.

In a physical sense, the centre of gravity (CoG) is an imaginary point around which the centre of an object′s weight lies. Using this idea, points in a plane can be separated into two classes by a line upon which their CoG lies. For two distinct and equally weighted clusters of points, the line will lie mid-way between them. Naturally, when the distinction between clusters becomes more ambiguous, the line will move up or down to reflect this change. In the K-O clustering algorithm, the initial threshold values take on this role of partitioning the objects into two classes. The closer the points lie to the object in question, the "higher" the threshold line is expected to be. In other words, a sensible positioning of the initial threshold line is the line upon which the CoG of the points lies. This is referred to as the CoG line as shown in Fig. 5.

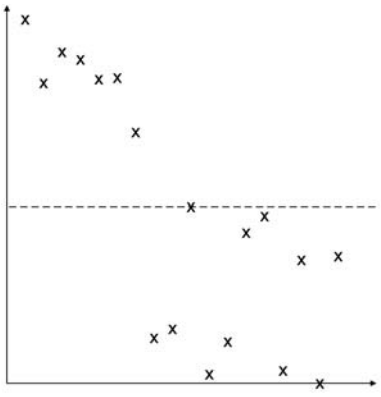

Fig. 5 Centre of gravity line

The CoG line of a set of points in the plane is positioned such that the sum of all perpendicular distances from the points to this line is zero. These calculations may be weighted if the CoG line is seemingly distorted by outlying points. Following this method, an initial threshold Th_i corresponding to the object x_i may be obtained by selecting the similarity value $s(x_i, x_k)$, $k = 1, 2, \cdots, n$, which minimizes the following sum of differences

$$\left| \sum_{j=1}^{n} (s(x_i, x_j) - ws(x_i, x_k)) \right|, \ i = 1, 2, \cdots, n \qquad (9)$$

where w is a weighting value that is usually set to 1 but may be set to 2 to raise the CoG line if necessary. This procedure produces a set of initial threshold values corresponding to each object in the universe from which the initial partitionings may be obtained. This information is then pooled to obtain the initial clustering of the universe U/R.

4.2 Assessment and modification of clusters

As mentioned earlier, the algorithm in the initial step will consist of a relatively high number of clusters. This is a result of the way in which the initial indiscernibility relations partition the universe. Specifically, each initial indiscernibility relation R_i imposes a partitioning of the universe U/R_i consisting of two classes. High numbers of initial clusters occur if the relation R_i disagrees on which pairs of points should belong to the same class. For example, for a given information system, if relation R_i places objects x_i and x_j in different classes, they will automatically belong to different clusters in the initial clustering; even if every other indiscernibility relation places them in the same class. This may be rectified in the later steps of the algorithm using global modification which alters this and, thus, the need for a high number of clusters. The global modification of any given clustering is controlled by the indiscernibility matrix $\Gamma = \{\gamma(x_i, x_j)\}$ introduced in Section 3. Its entries $\gamma(x_i, x_j)$ assess the indiscernibility degree between each pair of objects in the universe, and determine what proportion of the initial indiscernibility relations regard the two points to be indiscernible. In this way, the indiscernibility degree between two objects overlooks local discrepancies between equivalence relations. Modification to the given clustering is then performed using a modified indiscernibility relation as defined below:

Definition 7. Modified indiscernibility relation

Let $\mathbf{A} = (U, A)$ be an information system with non-empty finite universe $U = \{x_1, x_2, \cdots, x_n\}$ and attribute set $A = \{a_1, a_2, \cdots, a_k\}$. Suppose that U/R is a given clustering of the universe. The clustering is modified according to the indiscernibility relation:

$$R_i^{\text{mod}} = \{(x_i, x_j) \in U \times U : \gamma(x_i, x_j) \geq Th_\gamma, \ j = 1, \cdots, n\} \qquad (10)$$

where Th_γ is a pre-determined gamma threshold value.

In performing modification, a given clustering U/R is adapted to gain a coarser and more meaningful clustering of the universe U/R_{mod}. As with the initial thresholds, the choice of the gamma threshold value at each modification step will directly influence the final clustering obtained. Therefore, this value is chosen carefully. In a previous work, the gamma value was effectively hand-picked with a view to assessing the validity of the obtained clusterings and allowing for re-selection of an appropriate value if necessary[29]. This method provides good clusterings and be in keeping with the desire to maintain a high degree of autonomy and computational efficiency in the algorithm; it is preferable and less cumbersome to select the gamma threshold value autonomously according to some pre-determined accuracy criterion. A method for achieving this based on a defined clustering accuracy measure is suggested in Section 4.3.

4.3 Autonomous selection of gamma thresholds in cluster modification

The aim of knowledge-oriented clustering is to use both local and global knowledge to determine the partitioning of a given data set which, in some sense, represents an "accurate" clustering of the universe. Thus, it is possible to

assess the accuracy of a given clustering numerically as a linear combination of two distinct accuracy measures, acc_{within} and acc_{between}. They represent within and between-clusters accuracy, respectively (see Definitions 8 and 9). The within-clusters accuracy acc_{within} determines the degree of homogeneity within clusters for a given clustering formation. It is calculated as the mean (with respect to the number of clusters K) of the set of standard deviations of the unique similarity values corresponding to the objects in each cluster. For consistency, the trivial case of similarity between a point and itself is included. The result is modified to reduce the occurrence of too many clusters containing just one point (i.e., one point clusters). Between-clusters accuracy (acc_{between}) is taken as the mean of the minimum distances between each cluster, where the set of appropriate distances has been reduced to exclude distances between clusters lying at extreme ends of the clustering space. The aim is to obtain a clustering which reflects a high degree of homogeneity within the clusters and the opposite between the clusters. Because of the nature of the similarity value (4), a lower value of acc represents a more accurate clustering.

Definition 8. Let $U/R = \{C_1, C_2, \cdots, C_K\}$ be a clustering of the universe U. If a given cluster C_k, $k \in \{1, 2, \cdots, K\}$, contains m objects $\{x_1, x_2, \cdots, x_m\}$, define the function $A(C_k)$:

$$A(C_k) = \sqrt{\frac{\sum_{j>i}^{m}\sum_{i=1}^{m}[s(x_i, x_j) - \mu_{C_k}]^2}{m}} \tag{11}$$

where $s(x_i, x_j)$ represents the similarity between objects x_i and x_j, and μ_{C_k} is the mean of the similarity values in cluster C_k. The within clusters accuracy for the clustering U/R is defined as

$$acc_{\text{within}}(U/R) = \frac{\sum_{k=1}^{K} A(C_k)}{K} \cdot P^2 \tag{12}$$

where P is the number of clusters with cardinality 1.

Definition 9. Let $U/R = \{C_1, C_2, \cdots, C_K\}$ be a clustering of the universe U. Let $d(C_i, C_j)$ be the minimal distance between the clusters C_i and C_j, and it is calculated as the maximum similarity value between points in each cluster for the similarity measure defined in (4). Define

$$X = \frac{2\sum_{i=1}^{K-1}\sum_{j=i+1}^{K} d(C_i, C_j)}{K(K-1)}, \quad K > 1 \tag{13}$$

and let $B = \{d(C_i, C_j) : d(C_i, C_j) \geq X\}$. The between-clusters accuracy for the clustering U/R is defined as

$$acc_{\text{between}}(U/R) = \mu(B) \tag{14}$$

where μ represents the mean value of the set B.

Using Definitions 8 and 9, a gamma threshold value can be chosen autonomously according to the following Proposition.

Proposition. If U/R is a given clustering of the universe U and $\{Th_{\gamma_i}\}_{i=1}^{N}$ is a pre-determined set of possible gamma thresholds, then the threshold Th_γ used to achieve the modified clustering U/R_{mod} is chosen from the set $\{Th_{\gamma_i}\}_{i=1}^{N}$ to correspond to the minimum accuracy value

$$\min_{U/R_{\gamma_i}} acc(U/R_{\gamma_i}) = \min_{U/R_{\gamma_i}} \{0.1 acc_{\text{within}}(U/R_{\gamma_i}) + 0.9 acc_{\text{between}}(U/R_{\gamma_i})\} \tag{15}$$

where $\{U/R_{\gamma_i}\}_{i=1}^{N}$ are the partitionings generated by the values $\{Th_{\gamma_i}\}_{i=1}^{N}$, respectively.

The modification process is iterated until convergence to a value of the stable acc is achieved, at which the corresponding clustering is deemed to be the final and optimal clustering of the universe with respect to the defined accuracy value (15).

5 Experimental results

In this section, three data sets are clustered using the above algorithm. In the first instance, knowledge-oriented clustering with autonomy is used to cluster a small test data set. In Section 5.1 a step-by-step break-down of the procedure, which corresponds to the generic algorithmic framework shown in Fig. 4, is given. The food nutrient data, available in the Agriculture Yearbook[33], is clustered in Section 5.2 as a practical demonstration of the algorithm. Section 5.3 concludes with an illustration of knowledge-oriented clustering on a small mixed attribute data set.

5.1 Laboratory generated data results

A small test data set, consisting of 18 objects and 2 continuous attributes listed in Table 1, was generated in the department to verify the functionality of the autonomous knowledge-oriented clustering algorithm.

Table 1 Clustering data

Object	Attribute 1	Attribute 2	Object	Attribute 1	Attribute 2
x_1	0.05	0.13	x_{10}	0.40	0.54
x_2	0.06	0.32	x_{11}	0.72	0.90
x_3	0.11	0.21	x_{12}	0.74	0.74
x_4	0.16	0.10	x_{13}	0.47	0.57
x_5	0.19	0.25	x_{14}	0.49	0.50
x_6	0.23	0.13	x_{15}	0.76	0.83
x_7	0.06	0.47	x_{16}	0.61	0.55
x_8	0.68	0.80	x_{17}	0.84	0.80
x_9	0.69	0.74	x_{18}	0.52	0.60

The data set is sufficiently small to enable the workings of the algorithm to be described in an explicit manner, while the clear clustering structure shown in Fig. 6 highlights the data as a suitable candidate for any clustering procedure. Through visual analysis of the data plot, three clusters seem apparent. However, upon applying K-O clustering to the data, a result of four clusters is achieved as shown in Fig. 7. This suggests that the use of global modification draws out the inherent global data properties which remain concealed in a locally-dependent algorithm. To outline the detailed process of K-O clustering, a summary of the step-by-step procedure for the data in Table 1, as stated in Fig. 4, is provided below. Upper triangular forms of the symmetric similarity matrix (see Table A1 of Appendix) and indiscernibility matrices (see Tables A2 and A3 of Appendix) for all stages in the algorithm are provided in Appendix.

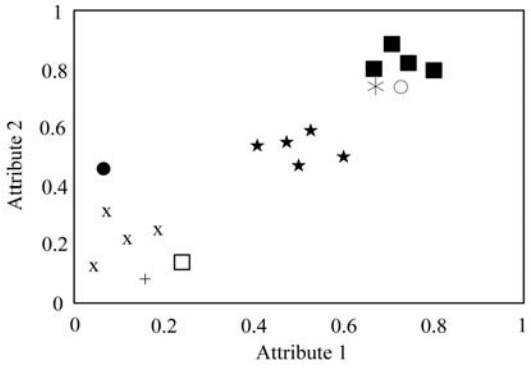

Fig. 6 Initial clusters

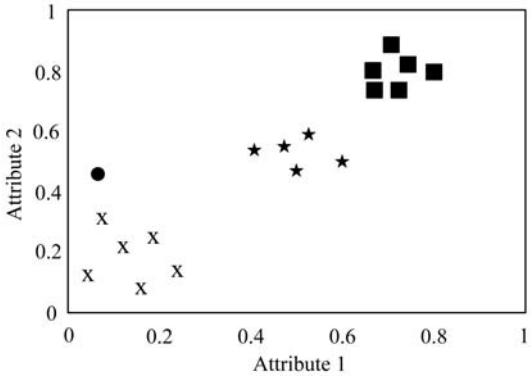

Fig. 7 Final clusters

Step 1. Construct matrix of similarities between all pairs of objects.

The Euclidean distance was selected as an appropriate s_{num} measure for this data and similarity between objects x_1 and x_2, i.e., $s(x_1, x_2)$, and objects x_1 and x_{18}, i.e., $s(x_1, x_{18})$, are calculated as

$$s(x_1, x_2) = 1 - \frac{\sqrt{(0.05 - 0.06)^2 + (0.13 - 0.32)^2}}{1.0359} = 0.81632$$

$$s(x_1, x_{18}) = 1 - \frac{\sqrt{(0.05 - 0.52)^2 + (0.13 - 0.6)^2}}{1.0359} = 0.35833$$

where $\max\limits_{i,j} s_{\text{num}}(x_i, x_j) = 1.0359$. Because of the nature of the similarity measure (4), similarity values closer to 1 indicate a greater similarity between the objects. The complete similarity matrix is displayed in Table A1 of Appendix.

Step 2. Assign an initial indiscernibility relation R_i to each object in the universe and pool the information to obtain an initial clustering U/R.

Initial threshold values Th_i were assigned to each object in the universe using the centre of gravity method (9) with $w = 2$. The results for the objects x_1 and x_{18} are displayed below.

$$U/R_1 = \{\{x_1, x_2, x_3, x_4, x_5, x_6\}, \{x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}\}\}$$
$$\vdots$$
$$U/R_{18} = \{\{x_9, x_{10}, x_{13}, x_{14}, x_{16}, x_{18}\}, \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_{11}, x_{12}, x_{15}, x_{17}\}\}$$

where $Th_1 = 0.81632$ and $Th_{18} = 0.7874$. When pooling the individual partitionings, the initial partitioning of the universe U/R produced 8 clusters as shown in Fig. 6.

$$U/R = \{\{x_1, x_2, x_3, x_5\}, \{x_4\}, \{x_6\}, \{x_7\}, \{x_8, x_{11}, x_{15}, x_{17}\}, \{x_9\}, \{x_{10}, x_{13}, x_{14}, x_{16}, x_{18}\}, \{x_{12}\}\}.$$

Step 3. Construct an indiscernibility matrix to assess the clustering U/R.

Using (5), the indiscernibility degrees between the object x_1 and various other objects are shown as follows.

$$\gamma(x_1, x_2) = 1,$$
$$\gamma(x_1, x_4) = 0.85714,$$
$$\gamma(x_1, x_7) = 0.42857.$$

These results indicate that 100% of the relations assign the objects x_1 and x_2 to the same class whereas only 42.86% of the relations would place x_1 and x_7 together.

Step 4. Modify clustering according to a modified indiscernibility relation R_i^{mod} to gain a modified clustering U/R_{mod}.

After calculating the complete gamma matrix, the initial clustering was modified with $Th_\gamma = 0.5$. Two examples of the individual modified partitionings are shown below followed by the modified clustering of the universe U/R_{mod}:

$$U/R_1^{\text{mod}} = \{\{x_1, x_2, x_3, x_4, x_5, x_6\}, \{x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}\}\}$$
$$\vdots$$
$$U/R_{18}^{\text{mod}} = \{\{x_{10}, x_{13}, x_{14}, x_{16}, x_{18}\}, \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{11}, x_{12}, x_{15}, x_{17}\}\}$$

$$U/R_{\text{mod}} = \{\{x_1, x_2, x_3, x_4, x_5, x_6\}, \{x_7\}, \{x_8, x_9, x_{11}, x_{12}, x_{15}, x_{17}\}, \{x_{10}, x_{13}, x_{14}, x_{16}, x_{18}\}\}.$$

Step 5. Repeat Steps 3 and 4 until a stable clustering is obtained.

For the data given in Table 1, convergence to the final solution was obtained after just one iteration. Fig. 7 shows the resulting clusters.

5.2 Practical clustering demonstration: Food nutrient data

The second data set to be considered is a real-world application. The food nutrient data available in the Agriculture Yearbook[33] have been clustered here using K-O clustering both with and without autonomy[28]. Tables 2 and 3 list the results. This classical clustering data set consists of 27 objects, including different types of meat, fish and foul and 5 attributes (i.e., food-calories, protein, fat, calcium and iron), as shown in Table A4 of Appendix. Protein and iron were found to be superfluous to the clustering[8]. Thus, for the purpose of visualizing the final clusters, the results obtained using 3 attributes, food-calories, fat and calcium, will be discussed.

Table 2 Autonomous clustering results for food nutrient data (Initial number of clusters is 17.)

Iteration	Th_γ	Clusters	acc
1	0.3	11	0.7473
2	0.5	7	0.7173

Table 3 Non-autonomous clustering results for food nutrient data (Initial number of clusters is 16, and $Th_{std} = 0.11$.)

Iteration	Th_γ	Clusters	acc
1	0.5	14	0.7625
2	0.4	9	0.7575
3	0.1	5	0.7452

Tables 2 and 3 display the results of K-O clustering with and without autonomy, respectively. The autonomous algorithm converges after 2 iterations to a final solution of 7 clusters (see Table 2) and the algorithm without autonomy converged after 3 iterations to a solution of 5 clusters (see Table 3). Since the algorithmic framework of knowledge-oriented clustering is similar to that of hierarchical clustering shown in Fig. 4, these results are compared in Table 4 to those obtained by using four traditional agglomerative hierarchical clustering techniques, namely complete-linkage, single-linkage, average-linkage and Ward′s method where the numbers indicate cluster membership. Although the two knowledge-oriented methods led to different final solutions, the similarities between the resulting clusters far out-weigh the differences. Thus, we suggest that both versions of the K-O algorithm have identified the salient features of the data. Furthermore, autonomous K-O clustering is operated with minimal subjectivity which guarantees consistent results when it is applied to the same data by different users. In contrast, the different methods within the agglomerative hierarchical clustering category produce different solutions on the same data. A cross-section of the similarity and gamma values calculated throughout the procedure is provided in Tables 5–7. The corresponding five numbered objects are shown in Fig. 8, where the Euclidean distance was chosen as the s_{num} measure.

Table 4 Comparison of clustering results for food data

Object	Food item	K-O with autonomy	K-O without autonomy	Complete-linkage & Ward′s	Single-linkage	Average-linkage
1	Braised beef	1	1	1	1	1
2	Hamburger	5	5	2	1	1
3	Roast beef	7	1	1	7	1
4	Beef steak	1	1	1	1	1
5	Canned beef	2	2	2	2	2
6	Broiled chicken	3	2	3	2	2
7	Canned chicken	2	2	2	2	2
8	Beef heart	2	2	2	2	2
9	Roast lamb leg	5	5	2	1	1
10	Roast lamb shoulder	1	1	2	1	1
11	Smoked ham	1	1	1	1	1
12	Roast pork	1	1	1	1	1
13	Simmered pork	1	1	1	1	1
14	Beef tongue	2	2	2	2	2
15	Veal cutlet	2	2	2	2	2
16	Baked bluefish	3	2	3	2	2
17	Raw clams	3	3	3	3	3
18	Canned clams	3	3	3	3	3
19	Canned crabmeat	3	2	3	2	2
20	Fried haddock	2	2	3	2	2
21	Broiled mackerel	2	2	2	2	2
22	Canned mackerel	6	3	3	6	3
23	Fried perch	2	2	2	2	2
24	Canned salmon	6	3	3	6	3
25	Canned sardines	4	4	4	4	4
26	Canned tuna	2	2	2	2	2
27	Canned shrimp	3	2	3	5	3

Table 5 Similarity values for food nutrient data

$s(x_i, x_j)$	x_4	x_{10}	x_{22}	x_{24}	x_{25}
x_4	1	0.8263	0.3863	0.3150	0.0586
x_{10}	0.8263	1	0.5208	0.4578	0.1286
x_{22}	0.3863	0.5208	1	0.9186	0.5124
x_{24}	0.3150	0.4578	0.9186	1	0.5008
x_{25}	0.0586	0.1286	0.5124	0.5008	1

Table 6 Gamma values at iteration 1 for food nutrient data

$\gamma(x_i, x_j)$	x_4	x_{10}	x_{22}	x_{24}	x_{25}
x_4	1	0.6667	0	0	0
x_{10}	0.6667	1	0	0	0
x_{22}	0	0	1	1	0.1429
x_{24}	0	0	1	1	0.1429
x_{25}	0	0	0.1429	0.1429	1

Table 7 Gamma values at iteration 2 for food nutrient data

$\gamma(x_i, x_j)$	x_4	x_{10}	x_{22}	x_{24}	x_{25}
x_4	1	0.9231	0	0	0
x_{10}	0.9231	1	0	0	0
x_{22}	0	0	1	1	0
x_{24}	0	0	1	1	0
x_{25}	0	0	0	0	1

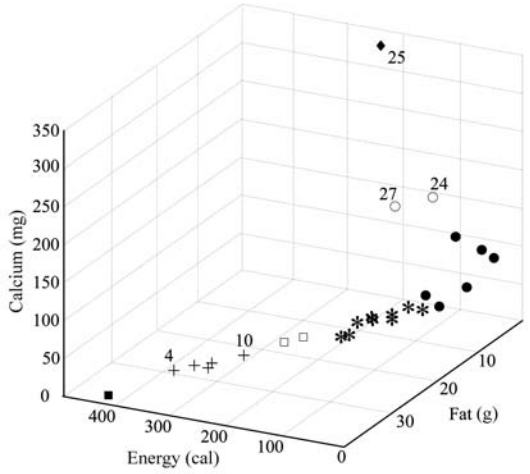

Fig. 8 Final clusters of the food nutrient data

5.3 Mixed attribute data

To establish the effectiveness of the autonomous knowledge-oriented clustering algorithm on a mixed attribute data set, the small data set shown in Table 8 has been clustered. It consists of 9 objects and 4 attributes (2 continuous attributes and 2 categorical attributes), and was originally used by Hirano and Tsumoto[31].

Table 8 Mixed attribute data set

Object	Attribute 1	Attribute 2	Attribute 3	Attribute 4
x_1	0.0	0.0	Round	Small
x_2	0.1	0.0	Round	Small
x_3	0.0	0.1	Round	Small
x_4	0.1	0.1	Round	Small
x_5	0.15	0.15	Square	Small
x_6	0.3	0.3	Square	Large
x_7	0.4	0.3	Square	Large
x_8	0.3	0.4	Square	Large
x_9	0.4	0.4	Square	Large

The similarity matrix S shown in Table A5 of Appendix was calculated using the Euclidean distance as an appropriate s_{num} measure and the Hamming distance as the s_{cat} measure. Using the centre of gravity method with $w = 1$, the following initial indiscernibility relations were obtained and led to an initial clustering U/R of four clusters.

$$U/R_1, U/R_2, U/R_3, U/R_4 = \{\{x_1, x_2, x_3, x_4, x_5\}, \{x_6, x_7, x_8, x_9\}\}$$

$$U/R_5 = \{\{x_2, x_3, x_4, x_5\}, \{x_1, x_6, x_7, x_8, x_9\}\}$$

$$U/R_6, U/R_7, U/R_8, U/R_9 = \{\{x_1, x_2, x_3, x_4\}, \{x_5, x_6, x_7, x_8, x_9\}\}$$

$$U/R = \{\{x_1\}, \{x_2, x_3, x_4\}, \{x_5\}, \{x_6, x_7, x_8, x_9\}\}.$$

The algorithm converged with $Th_\gamma = 0.2$ after just one iteration to a final solution of three clusters, $U/R_{\text{mod}} = \{\{x_1, x_2, x_3, x_4\}, \{x_5\}, \{x_6, x_7, x_8, x_9\}\}$. The complete gamma matrix is displayed in Table A6 of Appendix. In contrast to the result obtained by Hirano and Tsumoto[31], autonomous knowledge-oriented clustering has placed point x_5 into a cluster on its own, resulting in three rather than two final clusters. However, both the raw data (see Table 8) and the indiscernibility matrix (see Table A6 in Appendix) exhibit a degree of ambiguity surrounding the placement of this point. This suggests that the autonomous K-O clustering has exhibited a greater sensitivity to the inherent data knowledge by maintaining a one point cluster containing the point x_5.

6 Conclusions

Cluster analysis is an important exploratory technique for discovering patterns and underlying structures in data. The aim of clustering is to partition a data set into classes such that within-class homogeneity is high and between-class homogeneity weak. However, standard clustering techniques, including agglomerative hierarchical algorithms, k-means clustering and fuzzy c-means clustering, carry a number of inherent problems that directly influence the clustering solution. In all cases, a high degree of subjectivity is required to obtain an "optimal" clustering solution. This results in a non-unified approach to clustering, allowing for different clusters to be obtained when a given technique is applied to the same data by different people. This puts the optimality of any given solution under scrutiny in terms of

how well it really reflects true underlying data structures. Furthermore, the standard techniques generally focus on the clustering of single-type attribute data sets (e.g. continuous attributes) and are unable to cope easily with mixed attribute data. In terms of clustering applications, such as medical data, this is a major disadvantage.

To overcome these problems, this paper has proposed an autonomous knowledge-oriented clustering algorithm. The algorithmic framework forms clusters autonomously according to some pre-defined accuracy measure. In this way, the technique is standardized in the sense that multiple applications of the algorithm to the same data by different people will guarantee the same clustering solution. The algorithm handles mixed attribute data with ease and is such that no modification to the algorithm is needed to move between data sets of different attribute types.

It should be noted that the convergence of the algorithm to an "optimal" solution is governed by the similarity and indiscernibility matrices, which represent local and global knowledge, respectively. In other words, teaming with the algorithm′s standardized approach gives knowledge-oriented clustering the edge over other techniques. By incorporating global knowledge into the procedure, a coarse and representative clustering of the universe is obtained efficiently.

It is demonstrated that the use of global modification draws out the important data properties, which remain hidden in the standard clustering algorithms, and leads to a representative clustering. It is hypothesized that the knowledge-oriented clustering procedure may be used to extract "optimal" and non-ambiguous rules for a decision support system[28]. It remains a further work to assess the performance of the algorithm in situations of high ambiguity where clusters lie particularly close or are, indeed, overlapping.

Appendix

Tables A1–A4 are listed as follows.

Table A1 Similarity matrix for laboratory generated data

The values of the elements in the matrix																	
1	0.816	0.904	0.890	0.822	0.826	0.672	0.112	0.147	0.480	0.015	0.111	0.413	0.445	0.038	0.324	0	0.358
	1	0.883	0.767	0.858	0.754	0.855	0.243	0.270	0.610	0.152	0.228	0.536	0.550	0.164	0.425	0.116	0.480
		1	0.883	0.9137	0.861	0.744	0.208	0.242	0.576	0.111	0.205	0.509	0.539	0.133	0.416	0.094	0.454
			1	0.852	0.927	0.630	0.158	0.198	0.516	0.057	0.166	0.456	0.499	0.088	0.386	0.058	0.405
				1	0.878	0.753	0.289	0.324	0.654	0.190	0.289	0.590	0.623	0.215	0.502	0.178	0.536
					1	0.633	0.221	0.262	0.572	0.119	0.232	0.516	0.563	0.152	0.453	0.125	0.467
						1	0.322	0.338	0.665	0.240	0.294	0.593	0.584	0.240	0.464	0.182	0.539
							1	0.941	0.631	0.896	0.918	0.700	0.657	0.918	0.749	0.846	0.753
								1	0.660	0.843	0.952	0.732	0.698	0.890	0.801	0.844	0.787
									1	0.535	0.619	0.927	0.905	0.554	0.797	0.507	0.871
										1	0.844	0.600	0.555	0.922	0.646	0.849	0.652
											1	0.692	0.665	0.911	0.778	0.887	0.748
												1	0.930	0.624	0.864	0.580	0.944
													1	0.588	0.875	0.555	0.899
														1	0.693	0.918	0.679
															1	0.672	0.901
																1	0.636
																	1

Table A2 Indiscernibility matrix at iteration 1 for laboratory generated data

The values of the elements in the matrix																	
1	1	1	0.857	1	0.714	0.429	0	0	0	0	0	0	0	0	0	0	0
	1	1	0.857	1	0.714	0.429	0	0	0	0	0	0	0	0	0	0	0
		1	0.857	1	0.714	0.429	0	0	0	0	0	0	0	0	0	0	0
			1	0.857	0.833	0.286	0	0	0	0	0	0	0	0	0	0	0
				1	0.714	0.429	0	0	0	0	0	0	0	0	0	0	0
					1	0.143	0	0	0	0	0	0	0	0	0	0	0
						1	0	0	0	0	0	0	0	0	0	0	0
							1	0.6	0	1	0.857	0	0	1	0	1	0
								1	0.364	0.6	0.7	0.364	0.364	0.6	0.364	0.6	0.364
									1	0	0.091	1	1	0	1	0	1
										1	0.857	0	0	1	0	1	0
											1	0.091	0.091	0.857	0.091	0.857	0.091
												1	1	0	1	0	1
													1	0	1	0	1
														1	0	1	0
															1	0	1
																1	0
																	1

Table A3 Indiscernibility matrix at iteration 2 for laboratory generated data

The values of the elements in the matrix																	
1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
		1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
			1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
				1	1	0	0	0	0	0	0	0	0	0	0	0	0
					1	0	0	0	0	0	0	0	0	0	0	0	0
						1	0	0	0	0	0	0	0	0	0	0	0
							1	1	0	1	1	0	0	1	0	1	0
								1	0	1	1	0	0	1	0	1	0
									1	0	0	1	1	0	1	0	1
										1	1	0	0	1	0	1	0
											1	0	0	1	0	1	0
												1	1	0	1	0	1
													1	0	1	0	1
														1	0	1	0
															1	0	1
																1	0

Table A4 Food nutrient data

Object	Food item	Calories	Protein	Fat	Calcium	Iron
1	Braised beef	340	20	28	9	2.6
2	Hamburger	245	21	17	9	2.7
3	Roast beef	420	15	39	7	2.0
4	Beef steak	375	19	32	9	2.6
5	Canned beef	180	22	10	17	3.7
6	Broiled chicken	115	20	3	8	1.4
7	Canned chicken	170	25	7	12	1.5
8	Beef heart	160	26	5	14	5.9
9	Roast lamb leg	265	20	20	9	2.6
10	Roast lamb shoulder	300	18	25	9	2.3
11	Smoked ham	340	20	28	9	2.5
12	Roast pork	340	19	29	9	2.5
13	Simmered pork	355	19	30	9	2.4
14	Beef tongue	205	18	14	7	2.5
15	Veal cutlet	185	23	9	9	2.7
16	Baked bluefish	135	22	4	25	0.6
17	Raw clams	70	11	1	82	6.0
18	Canned clams	45	7	1	74	5.4
19	Canned crabmeat	90	14	2	38	0.8
20	Fried haddock	135	16	5	15	0.5
21	Broiled mackerel	200	19	13	5	1.0
22	Canned mackerel	155	16	9	157	1.8
23	Fried perch	195	16	11	14	1.3
24	Canned salmon	120	17	5	159	0.7
25	Canned sardines	180	22	9	367	2.5
26	Canned tuna	170	25	7	7	1.2
27	Canned shrimp	110	23	1	98	2.6

Table A5 Similarity matrix for mixed attribute data

The value of the elements in the matrix								
1	0.9116	0.9116	0.8750	0.5625	0.1250	0.0581	0.0581	0
	1	0.8750	0.9116	0.6102	0.1813	0.1250	0.1047	0.0581
		1	0.9116	0.6102	0.1813	0.1047	0.1250	0.0581
			1	0.6875	0.2500	0.1813	0.1813	0.1250
				1	0.5625	0.4923	0.4923	0.4375
					1	0.9116	0.9116	0.8750
						1	0.8750	0.9116
							1	0.9116
								1

Table A6 Indiscernibility matrix for mixed attribute data

The values of the elements in the matrix								
1	0.8	0.8	0.8	0.4444	0	0	0	0
	1	1	1	0.5556	0	0	0	0
		1	1	0.5556	0	0	0	0
			1	0.5556	0	0	0	0
				1	0.4444	0.4444	0.4444	0.4444
					1	1	1	1
						1	1	1
							1	1
								1

References

[1] T. Sorensen. A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to the Analyses of the Vegetation on Danish Commons. *Biologiske Skrifter*, vol. 5, no. 4, pp. 1–34, 1948.

[2] P. Sneath. The Application of Computers to Taxonomy. *Journal of General Microbiology*, vol. 17, no. 1, pp. 201–226, 1957.

[3] R. R. Sokal, P. H. A. Sneath. *Principles of Numerical Taxonomy*, W. H. Freeman, San Francisco, USA, 1963.

[4] J. H. Ward. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, 1963.

[5] M. R. Anderberg. *Cluster Analysis for Applications*, Academic Press, New York, USA, 1973.

[6] M. S. Aldenderfer, R. K. Blashfield. *Cluster Analysis*, Sage University Paper, Newbury Park, USA, 1984.

[7] B. S. Everitt. *Cluster Analysis*, Edward Arnold, Cambridge, UK, 1993.

[8] S. Sharma. *Applied Multivariate Techniques*, John Wiley & Sons, New York, USA, 1996.

[9] A. K. Jain, M. N. Murty, P. J. Flynn. Data Clustering: A review. *ACM Computing Surveys*, vol. 31, no.3, pp. 264–323, 1999.

[10] R. R. Yegar. Intelligent Control of the Hierarchical Clustering Process. *IEEE Transactions on Systems, Man, and Cybenetics-Part B*, vol. 30, no. 6, pp. 835–845, 2000.

[11] E. W. Forgey. Cluster Analysis of Multivariate Data: Efficiency Versus Interpretability of Classifications. *Biometrics*, vol. 21, no. 3, pp. 768–769, 1965.

[12] R. C. Jancey. Multidimensional Group Analysis. *Australian Journal of Botany*, vol. 14, no. 1, pp. 127–130, 1966.

[13] J. B. MacQueen. Some Methods of Classification and Analysis of Multivariate Observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, USA, vol. 1, pp. 281–297, 1967.

[14] G. H. Ball, D. J. Hall. ISODATA, a Novel Method of Data Analysis and Pattern Classification, Technical Report AD 699616, Stanford Research Institute, Menlo Park, USA, 1965.

[15] F. H. C. Marriott. Optimization Methods of Cluster Analysis. *Biometrika*, vol. 69, no. 2, pp. 417–421, 1982.

[16] S. Z. Selim, M. A. Ismail. K-means Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality. *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 6, no. 1, pp. 81–87, 1984.

[17] J. C. Dunn. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact, Well Separated Clusters. *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973.

[18] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithm*, Plenum Press, New York, USA, 1981.

[19] A. K. Jain, R. C. Dubes. *Algorithms for Clustering Data*, Prentice-Hall, USA, 1988.

[20] M. S. Kamel, S. Z. Selim. New Algorithms for Solving the Fuzzy Clustering Problem. *Pattern Recognition*, vol. 27, no. 3, pp. 421–428, 1994.

[21] J. S. R. Jang, C. T. Sun, E. Mizutani. *Neuro-fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Prentice-Hall, USA, 1996.

[22] F. Höppner, F. Klawonn, R. Kruse, T. Runkler. *Fuzzy Cluster Analysis*. Wiley & Sons, Chichester, England, 1999.

[23] Z. Pawlak. Rough Sets. *International Journal of Information and Computer Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[24] Z. Pawlak. *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic, Dordrecht, Holland, 1991.

[25] A. Skowron, C. Rauszer. The Discernibility Matrices and Functions in Information Systems. *Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory*, R. Slowinski (ed.), Kluwer Academic, Dordrecht, Holland, pp. 331–362, 1992.

[26] J. Komorowski, Z. Pawlak, L. Polkowski, A. Skowron. Rough Sets: A Tutorial. *Rough Fuzzy Hybridization: A New Method for Decision Making*, S. Pal, A. Skowron (eds.), Springer, Berlin, Germany, 1998.

[27] D. Dubois, H. Prade. Rough Fuzzy Sets and Fuzzy Rough Sets. *International Jounal of General Systems*, vol. 17, no. 2, pp. 191–209, 1989.

[28] C. L. Bean, C. Kambhampati. Knowledge-oriented Clustering for Decision Support. In *Proceedings of IEEE International Joint Conference on Neural Networks*, Portland, Oregon, USA, vol. 4, pp. 3244–3249, 2003.

[29] T. Okuzaki, S. Hirano, S. Kobashi, Y. Hata, Y. Takahashi. A Rough Set Based Clustering Method by Knowledge Combination. *IEICE Transactions on Information and Systems*, vol. 85, no. 12, pp. 1898–1908, 2002.

[30] C. L. Bean, C. Kambhampati, S. Rajasekharan. A Rough Set Solution to a Fuzzy Set Problem. In *Proceedings of IEEE International Conference on Fuzzy Systems, World Congress in Computational Intelligence*, Honolulu, Hawaii, vol. 1, pp. 18–23, 2002.

[31] S. Hirano, S. Tsumoto. A Knowledge-oriented Clustering Technique Based on Rough Sets. In *Proceedings of 25th IEEE International Conference on Computer and Software Applications*, Chicago, USA, pp. 632–637, 2001.

[32] B. J. F. Manly. *Multivariate Statistical Methods, A Primer*, Chapman & Hall, New York, USA, 2000.

[33] J. A. Hartigan. *Clustering Algorithms*, John Wiley & Sons, New York, USA, 1975.

Charlotte Bean received her Master of mathematics (MMath) degree from the University of Hull, UK, in 1999 and the Ph. D. degree in computer science from the same university in 2004. She is currently a research fellow in the Medical School at the University of Warwick, UK.

Her research interests include data analysis, the theoretical development of techniques and algorithms used in statistical modeling and data mining, especially using tools such as rough set theory, FST, and artificial neural networks.

Chandra Kambhampati received his Ph. D. degree from City University, London, for his dissertation on algorithms for optimizing control in 1988. He hold positions in the Department of Cybernetics, at the University of Reading, and the Department of Chemical and Process Engineering, at the University of Newcastle Upon Tyne. He is currently a reader in the Department of Computer Science at the University of Hull. He also heads the Neural, Emergent and Agent Technologies (NEAT) Group, and has a number of doctoral and graduate students, and research associates. He has authored and co-authored over 100 papers on optimization, adaptive optimization, optimal control, neural networks, and fuzzy logic for control and robotics. He is a member of IEEE and IEE.

His research interests include fault tolerant control, networked control systems, signal processing, neural networks, and multi-agent systems.