International Journal of Approximate Reasoning 55 (2014) 238–258

Contents lists available at SciVerse ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Qualitative and quantitative combinations of crisp and rough clustering schemes using dominance relations

Pawan Lingras [a,*], Min Chen [b], Duoqian Miao [b]

[a] Department of Mathematics & Computing Science, Saint Mary's University, Halifax, Nova Scotia B3H 3C3, Canada
[b] School of Electronics and Information Engineering, Tongji University, Shanghai 201804, PR China

ARTICLE INFO

Article history:
Available online 7 June 2013

Keywords:
Crisp clustering
Rough clustering
Dominance relations
Preference relations
Qualitative reasoning
Combination of clustering schemes

ABSTRACT

Due to their unsupervised learning nature, analyzing the semantics of clustering schemes can be difficult. Qualitative information such as preference relations may be useful in semantic analysis of clustering process. This paper describes a framework based on preference or dominance relations that helps us qualitatively analyze a clustering scheme. This qualitative interpretation is shown to be useful for combining clustering schemes that are based on different criteria. The qualitative combination can be used to analyze its quantitative counterpart and can also be used instead of the quantitative combination. The paper further extends the framework to accommodate rough set based clustering. The usefulness of the approach is illustrated using a synthetic retail database.

1. Introduction

Clustering categorizes a set of objects into a number of groups based on their similarity with each other. Usually, the similarity is measured on a quantifiable distance between these objects. For example, visitors to a retail store could be grouped based on their spending. In many practical applications, the distance measure can provide a certain ranking among the objects as well as the clusters. This ranking can be represented by a preference or dominance relation. Fishburn [1] described a qualitative decision theory based on the notion of preference relations. Decision theoretic rough sets proposed by Yao and studied by a number of researchers provide a segue from rough set theory to the preference relations [2–5,4,6–12]. Dominance relations, which can also be described using strict preference and indifference relations, have been used in the context of rough set classification by a number of researchers [13–16]. This paper describes the use of dominance relations for analysis and enhancement of clustering techniques. A framework with strict preference and indifference relations is proposed to capture qualitative information in a crisp clustering scheme. The proposed preference/dominance based framework is used to axiomatize the clustering process. It is also used for combination of clustering schemes based on different attributes. Continuing with our example of retail customers, let us say that we want to cluster them using their loyalty. Since loyalty is a subjective measure, we will use their frequency of visits as a surrogate for loyalty. Now we have two clustering schemes for the retail store visitors: one based on spending and another based on visits (loyalty). We may want to combine these two clustering schemes to get a better understanding of the value of the customers. Traditionally, such combination is achieved by combining the two sets of attributes and re-cluster the objects. It is possible to assign different weights to these attributes as an indication of their relative importance. We will term this traditional combination as the quantitative approach. The dominance relation based framework provides us with a qualitative combination method. The qualitative combination can be useful as an alternative to the quantitative approach as it can reduce the emphasis on numeric distances. It can also be used

* Corresponding author.
E-mail addresses: pawan@cs.smu.ca (P. Lingras), minchen2008halifax@yahoo.com (M. Chen), miaoduoqian@163.com (D. Miao).

0888-613X/$ - see front matter
http://dx.doi.org/10.1016/j.ijar.2013.05.007

in knowledge networks, where the attribute values may not be transportable for privacy or efficiency considerations [17]. The qualitative combination can also help us in understanding the axiomatic behavior of quantitative combination.

In addition to clearly identifiable groups of objects, it is possible that a data set may consist of several objects that lie on the fringes. The conventional clustering techniques mandate that such objects belong to precisely one cluster. Such a requirement is found to be too restrictive in many data mining applications. In practice, an object may display characteristics of different clusters. In such cases, an object should belong to more than one cluster, and as a result, cluster boundaries necessarily overlap. Fuzzy set representation of clusters, using algorithms such as fuzzy c-means, make it possible for an object to belong to multiple clusters with a degree of membership between 0 and 1 [18–24]. Masson and Denoeux have extended the fuzzy clustering using the belief function theory to propose evidential clustering [22,25,26]. The evidential clustering was also shown to produce lower and upper bounds similar to the rough set theory.

In some cases, the fuzzy degree of membership may be too descriptive for interpreting clustering results. Rough set based clustering provides a solution that is less restrictive than conventional clustering and less descriptive than fuzzy clustering. Peters et al. provide a comparison of fuzzy and rough clustering [27].

Rough set theory has made substantial progress as a classification tool in data mining [28,29]. The basic concept of representing a set as lower and upper bounds can be used in a broader context such as clustering. Clustering in relation to rough set theory is attracting increasing interest among researchers [30–36]. Lingras [37] described how a rough set theoretic classification scheme can be represented using a rough set genome. In subsequent publications [38,39], modifications of K-means and *Kohonen Self-Organizing Maps* (SOMs) were proposed to create intervals of clusters based on rough set theory. Yao [3] described a more generic usage of the lower and upper bounds using the interval set algebra. Yao et al. [9] studied the implications of interval set algebra for cluster analysis. Recently, a number of research efforts relate the unsupervised rough clustering with decision theoretic rough set framework [40–44].

The preference/dominance relations based framework can be easily extended to accommodate rough clustering. The paper describes the extended framework and qualitative combination of rough clustering. The qualitative combination is again compared with its quantitative counter part. The treatment presented in this paper provides further insight into rough clustering and combination of rough clustering schemes.

All the theoretical development in this study is applied to a synthetic retail dataset to illustrate its applications in data mining.

Section 2 provides a review of crisp and soft clustering, followed by the definitions of preference, dominance, and indifference relations, as well as an overview of dominance based rough set theory. The notion of qualitative crisp clustering is introduced in Section 3. The qualitative combination of crisp clustering is proposed in Section 4. A comparison of qualitative and quantitative combination of crisp clustering is presented in Section 5. Section 6 extends the notion of qualitative clustering to rough set theory. The qualitative combination of rough clustering is derived in Section 7, followed by a comparative study of qualitative and quantitative combination of rough clustering in Section 8. Section 9 summarizes the contributions of the paper and concludes the study.

2. Review of literature

In this section, we will review crisp and soft clustering, followed by the definitions of the dominance and preference relations. We will then briefly discuss how dominance relations are used in rough set based classification. The section will further describe the use of preference relations in a more general decision theoretic framework and its relationship with rough sets. This section will end with a study of how decision theoretic framework has been applied to unsupervised clustering problems.

2.1. Crisp and soft clustering

Clustering is a popular set of data mining techniques that uses the principle of unsupervised learning to group similar pattern. One can broadly categorize clustering techniques into crisp and soft clustering methods. In crisp clustering, a pattern can belong to one and only one cluster leading to mutually exclusive clusters. Soft clustering techniques make it possible for a pattern to belong to multiple and overlapping clusters. In this subsection, we will provide a brief overview of crisp clustering, followed by soft clustering based on fuzzy set and belief function theories. Since rough clustering is central to the present study, it is discussed in greater detail in the following subsection. Peters et al. [27] discuss the complementary features of rough and fuzzy clustering in relation to other soft clustering methods.

2.1.1. Crisp clustering

Let $X = \{\vec{x}_1, \ldots, \vec{x}_n\}$ be a set of objects represented as vectors. We want to group these objects into k clusters. The resulting clustering scheme cs partitions X into clusters $cs = \{\vec{c}_1, \ldots, \vec{c}_k\}$. Each cluster $\vec{c}_t$ $(1 \leqslant t \leqslant k)$ is represented by the centroid vector of the cluster. For notational convenience, we will use $\vec{c}_t$ to represent the centroid vector of the cluster and c_t to denot the set of objects that belong to the cluster depending on the context.

K-means clustering is one of the most popular statistical clustering techniques [45,46]. The objective is to assign n objects to k clusters. The process begins by randomly choosing k objects as the centroids of the k clusters. The objects are assigned to one of the k clusters based on the minimum value of the distance $d(\vec{x}_l, \vec{c}_t)$ between the object vector $\vec{x}_l$ and the cluster vector $\vec{c}_t$. The distance $d(\vec{x}_l, \vec{c}_t)$ can be the standard Euclidean distance.

After the assignment of all the objects to various clusters, the new centroid vectors of the clusters are calculated as:

$$\vec{c}_t = \frac{\sum_{\vec{x}_l \in c_t} \vec{x}_l}{\mid c_t \mid}, \quad \text{where } 1 \leqslant t \leqslant k.$$

Here $\mid c_t \mid$ is a cardinality of cluster c_t.The process stops when the centroids of clusters stabilize, i.e. the centroid vectors from the previous iteration are identical to those generated in the current iteration.

2.1.2. Fuzzy c-means algorithm

The fuzzy c-means (FCM) was originally proposed by Dunn in 1973 [18] and improved by Bezdek in 1981 [20]. It was one of the first soft clustering techniques that allowed objects to belong to two or more clusters with a degree of belonging to clusters. Fuzzy c-means minimizes the objective function:

$$\sum_{i=1}^{n} \sum_{j=1}^{k} u_{ij}^{m} \, d(\vec{x}_i, \vec{c}_j), \quad 1 < m < \infty \tag{1}$$

where n is the number of objects and each object is a d-dimensional vector. A parameter m is any real number greater than 1, u_{ij} is the degree of membership of the i-th object $(\vec{x}_i)$ in the cluster j, and $d(\vec{x}_i, \vec{c}_j)$ is the Euclidean distance between an object and a cluster center c_j.

The degree of membership given by a matrix $\vec{u}$ for objects on the edge of a cluster, may have a lesser degree than objects in the center of a cluster. But, the sum of these coefficients for any given object x_i is defined to be 1.

$$\sum_{j=1}^{k} \; u_{ij} = 1 \quad \forall i. \tag{2}$$

In fuzzy c-means, the centroid of a cluster is obtained by average of all objects, weighted by their degree of membership to a cluster:

$$\vec{c}_j = \frac{\sum_{i=1}^{n} \; u_{ij}^{m} \; \vec{x}_i}{\sum_{i=1}^{n} u_{ij}^{m}}. \tag{3}$$

FCM is an iterative algorithm that terminates if

$$max\left(\left|u_{ij}^{t+1} - u_{ij}^{t}\right|\right) < \delta \tag{4}$$

where δ is a termination criterion between 0 and 1, and t is the iteration step. Pal et al. [23] proposed a variation of fuzzy c-means algorithm called possibilistic c-means by relaxing the restriction given by Eq. (2).

2.1.3. Evidential c-means algorithm

Evidence theory based on belief functions formed the basis of a soft clustering algorithm technique proposed by Denoeux and Masson [22], which builds on the fuzzy c-means algorithm. The original proposal assigned basic belief mass to only k clusters from the set of cluster cs, which was further generalized by the authors by assigning basic belief mass to the power set 2^{cs} of clusters resulting in 2^k possible assignments of basic belief mass. This generalized algorithm is called evidential c-means ECM [25]. The 2^{cs} subsets of cs in the ECM algorithm model all situations ranging from complete ignorance to full certainty concerning the membership of an object.

The ECM algorithm first obtains a credal partition followed by a separate treatment of an empty set of the credal partition in order to obtain basic belief assignment (bba) for all 2^k classes. The similarity between an object and a cluster is measured using the Euclidean metric. In order to obtain the final solution matrix, the problem is represented as an unconstrained optimization problem and solved using an iterative algorithm.

Hence, for a dataset that has 3 clusters, ECM generates $2^3 = 8$ different bba values m_i for an object $\vec{x}_i$. These eight values correspond to the knowledge regarding the class membership of the object $\vec{x}_i$. If a bba value of an object $\vec{x}_i$ for a cluster $\vec{c}_j$ is 1, then we can say that cluster assignment of $\vec{x}_i$ is known with certainty. For any other case, we have partial knowledge about cluster membership of an object $(0 < bba < 1)$; no knowledge about cluster membership of an object (based on the bba for the entire set of clusters cs), and outlier characteristics of an object (based on bba of the empty set $\emptyset$). Masson and Denoeux [26] have further extended the evidential soft clustering by proposing an ensemble clustering under the belief function framework. The representation of no knowledge and identification of outliers is one of the major strengths of the clustering based on belief function theory.

2.2. Rough clustering

Due to space limitations, some familiarity with rough set theory is assumed [29]. Rough sets were originally proposed using equivalence relations. However, it is possible to define a pair of lower and upper bounds $(lower(c_t), upper(c_t))$ or a rough set for every set $c_t \subseteq X$ as long as the properties specified by Pawlak [29] are satisfied. In addition to $(lower(c_t), upper(c_t))$, we will also use the term boundary region of c_t denoted by $bnd(c_t) = upper(c_t) - lower(c_t)$. Yao et al. [2] described various generalizations of rough sets by relaxing the assumptions of an underlying equivalence relation. Such a trend towards generalization is also evident in rough mereology proposed by Polkowski and Skowron [47] and the use of information granules in a distributed environment by Skowron and Stepaniuk. The present study uses such a generalized view of rough sets. If one adopts a more restrictive view of rough set theory, the rough sets developed in this paper may have to be looked upon as interval sets.

Let us consider a hypothetical classification scheme

$$X/p = \{c_1, c_2, \ldots, c_k\} \tag{5}$$

that partitions the set X based on an equivalence relation p. Let us assume due to insufficient knowledge that it is not possible to precisely describe the sets c_t, $1 \leqslant t \leqslant k$, in the partition. Based on the available information, however, it is possible to define each set $c_t \in X/p$ using its lower $lower(c_t)$ and upper $upper(c_t)$ bounds.

We are considering the upper and lower bounds of only a few subsets of X. Therefore, it is not possible to verify all the properties of the rough sets [29]. However, the family of upper and lower bounds of $c_t \in X/p$ are required to follow some of the basic rough set properties such as:

(P1) An object $\vec{x}$ can be part of at most one lower bound;

(P2) $\vec{x} \in lower(c_t) \Rightarrow \vec{x} \in upper(c_t)$;

(P3) An object $\vec{x}$ is not part of any lower bound $\Leftrightarrow$ $\vec{x}$ belongs to two or more upper bounds.

Property (P1) emphasizes the fact that a lower bound is included in a set. If two sets are mutually exclusive, their lower bounds should not overlap. Property (P2) confirms the fact that the lower bound is contained in the upper bound. Property (P3) is applicable to the objects in the boundary regions, which are defined as the differences between upper and lower bounds. The exact membership of objects in the boundary region is ambiguous. Therefore, property (P3) states that an object cannot belong to only a single boundary region. Property (P3) is specific to the rough sets interpretation of lower and upper bounds. Yao [3] provided a comprehensive generalization of the lower and upper bounds using the interval set algebra. Yao et al. [9] further explored the implications of interval set algebra for cluster analysis. While the approach proposed in this paper uses the traditional rough set model, future work will extend the proposed qualitative clustering based on interval set cluster analysis.

Note that (P1)–(P3) are not necessarily independent or complete. However, enumerating them will be helpful later in understanding the rough set adaptation of evolutionary, neural, and statistical clustering methods. In the context of decision-theoretic rough set model, Yao and Zhao [12] provide a more detailed discussion on the important properties of rough sets and positive, boundary, and negative regions.

Incorporating rough sets into K-means clustering requires the addition of the concept of lower and upper bounds. Calculation of the centroids of clusters from conventional K-means needs to be modified to include the effects of these bounds. The modified centroid calculations for rough sets are then given by: if $lower(\vec{c}) \neq \emptyset$ and $bnd(\vec{c}) = \emptyset$

$$\vec{c} = \frac{\sum_{\vec{x} \in lower(\vec{c})} \vec{x}}{|lower(\vec{c})|}$$

else if $lower(\vec{c}) = \emptyset$ and $bnd(\vec{c}) \neq \emptyset$

$$\vec{c} = \frac{\sum_{\vec{x} \in bnd(\vec{c})} \vec{x}}{|bnd(\vec{c})|}$$

else

$$\vec{c} = w_{lower} \times \frac{\sum_{\vec{x} \in lower(\vec{c})} \vec{x}}{|lower(\vec{c})|} + w_{upper} \times \frac{\sum_{\vec{x} \in bnd(\vec{c})} \vec{x}}{|bnd(\vec{c})|}. \tag{6}$$

The parameters w_{lower} and w_{upper} correspond to the relative importance of lower and upper bounds, and $w_{lower} + w_{upper} = 1$. If the upper bound of each cluster were equal to its lower bound, the clusters would be conventional clusters. Therefore, the boundary region $bnd(\vec{c})$ will be empty, and the second term in the equation will be ignored. Thus, Eq. (6) will reduce to conventional centroid calculations.

The next step in the modification of the K-means algorithms for rough sets is to design criteria to determine whether an object belongs to the upper or lower bound of a cluster given as follows. For each object vector $\vec{x}$, let $d(\vec{x}, \vec{c}_j)$ be the distance between itself and the centroid of cluster $\vec{c}_j$. Let $d(\vec{x}, \vec{c}_i) = \min_{1 \leqslant j \leqslant k} d(\vec{x}, \vec{c}_j)$. The ratio $d(\vec{x}, \vec{c}_i)/d(\vec{x}, \vec{c}_j)$, $1 \leqslant i, j \leqslant k$, is used to determine the membership of $\vec{x}$. Let $T = \{j : d(\vec{x}, \vec{c}_i)/d(\vec{x}, \vec{c}_j) \leqslant threshold \text{ and } i \neq j\}$.

1. If $T \neq \emptyset, \vec{x} \in upper(c_i)$ and $\vec{x} \in upper(c_j), \forall j \in T$. Furthermore, $\vec{x}$ is not part of any lower bound. The above criterion guarantees that property (P3) is satisfied.
2. Otherwise, if $T = \emptyset, \vec{x} \in lower(c_i)$. In addition, by property (P3), $\vec{x} \in upper(c_i)$.

Rough clustering is gaining increasing attention from researchers. The rough K-means approach, in particular, has been a subject of further research. Peters [35] discussed various deficiencies of Lingras and West's original proposal [38]. The first set of independently suggested alternatives by Peters are similar to the Eq. (6). Peters also suggests the use of ratios of distances as opposed to differences between distances similar to those used in the rough set based Kohonen algorithm described in [39]. The use of ratios is a better solution than differences. The differences vary based on the values in input vectors. The ratios, on the other hand, are not susceptible to the input values. Peters [35] has proposed additional significant modifications to rough K-means that improve the algorithm in a number of aspects. The refined rough K-means algorithm simplifies the calculations of the centroid by ensuring that lower bound of every cluster has at least one object. It also improves the quality of clusters as clusters with empty lower bound have a limited basis for its existence. Peters tested the refined rough K-means for various datasets. The experiments were used to analyze the convergence, dependency on the initial cluster assignment, study of Davies–Boulden index, and to show that the boundary region can be interpreted as a security zone as opposed to the unambiguous assignments of objects to clusters in conventional clustering. Despite the refinements, Peters concluded that there are additional areas in which the rough K-means needs further improvement, namely in terms of selection of parameters.

2.3. Preference based decision theory

Fishburn's [1] seminal work in 1970 on preference based decision theory forms the basis of our work. Fishburn suggested that the qualitative notion of preferences plays a significant part in our decision making. He provided rigorous axiomatic preference structures and their numerical representation. The axioms or conditions on preference structures provide consistent and coherent decision making. These axioms also enable a structured and simplified reasoning process.

Based on certain criteria, it may be possible to partially order set X of objects using a set of dominance or preference relations [8] defined as follows:

Strict preference relation: Let $\vec{x}_i$ and $\vec{x}_j$ be two vectors. If $\vec{x}_j$ is worse than $\vec{x}_i$, then $\vec{x}_i \succ \vec{x}_j$. Assuming that $\vec{x}_i \neq \vec{x}_j \neq \vec{x}_l$, the strict preference relation obeys the following properties:
Transitivity: $(\vec{x}_i \succ \vec{x}_j) \wedge (\vec{x}_j \succ \vec{x}_l) \Rightarrow \vec{x}_i \succ \vec{x}_l$.
Asymmetry: $\vec{x}_i \succ \vec{x}_j \Rightarrow \neg(\vec{x}_j \succ \vec{x}_i)$.

Indifference relation: Let $\vec{x}_i$ and $\vec{x}_j$ be two vectors. If $\vec{x}_i$ and $\vec{x}_j$ are similar, then $\vec{x}_i \sim \vec{x}_j$. Assuming that $\vec{x}_i \neq \vec{x}_j \neq \vec{x}_l$, the indifference relation obeys the following properties:
Transitivity: $(\vec{x}_i \sim \vec{x}_j) \wedge (\vec{x}_j \sim \vec{x}_l) \Rightarrow \vec{x}_i \sim \vec{x}_l$.
Symmetry: $\vec{x}_i \sim \vec{x}_j \Rightarrow \vec{x}_j \sim \vec{x}_i$.
Reflexivity: $\vec{x}_i \sim \vec{x}_i$.

Dominance relation: Let $\vec{x}_i$ and $\vec{x}_j$ be two vectors. If $\vec{x}_i$ is no worse than $\vec{x}_j$, then $\vec{x}_i \succeq \vec{x}_j$. The dominance relation is transitive: $(\vec{x}_i \succeq \vec{x}_j) \wedge (\vec{x}_j \succeq \vec{x}_l) \Rightarrow \vec{x}_i \succeq \vec{x}_l$.

The three relations defined above are related to each other as follows:
$\vec{x}_i \succeq \vec{x}_j \Leftrightarrow \neg(\vec{x}_j \succ \vec{x}_i)$,
$\vec{x}_i \succeq \vec{x}_j \wedge \vec{x}_j \succeq \vec{x}_i \Leftrightarrow \vec{x}_j \sim \vec{x}_i$,
$\vec{x}_i \succ \vec{x}_j \vee \vec{x}_i \sim \vec{x}_j \Leftrightarrow \vec{x}_j \succeq \vec{x}_i$.

In dominance based rough sets, attributes with totally ordered domains are called criteria [13–16]. The objects are ordered based on the decision class they belong to. Greco et al. [14–16] assumed monotonicity constraints, i.e. a higher evaluation of an object on an attribute, with other evaluations being fixed, should not decrease its assignment to the class. Greco et al. [14–16] claim that the ordinal classification problem with monotonicity constraints is similar to classification problem in machine learning with two additional constraints. The first constraint is that there is an ordinal scale on each attribute and between class indices. The second constraint stipulates that the expected class index increases with increasing values on attributes. Despite the monotonic data, we could have a situation where an object $\vec{x}_i$ may not be worse than another object $\vec{x}_j$ on all attributes, but, $\vec{x}_i$ is assigned to a worse class than $\vec{x}_j$, violating the monotonicity. Greco et al. [14–16] used rough set theory to deal with such inconsistency by proposing Dominance-based Rough Set Approach (DRSA). Dominance based rough set theory is attracting research interest. DRSA has been used successfully in a number of applications such as bankruptcy risk prediction, breast cancer diagnosis, house pricing, credit rating, and liver disorder diagnosis [14–16]. Chen and Tzeng [13] proposed an efficient implementation of DRSA. Combination of dominance relations from multiple criteria is one of the important features of the recent research using dominance relation. The present study uses some of the concepts proposed by Greco et al. [14–16] and Chen and Tzeng [13] for analyzing crisp and rough clustering.

2.4. Decision theoretic rough sets

Yao et al. [6] first proposed the notion of decision theoretic rough set model in 1990. Yao and Wong [7] further formalized the decision theoretic framework in the context of rough set theory. The generalized rough set models [2] in combination with the earlier decision theoretic approach led to a more generalized decision theoretic rough set framework [4,5,11,12]. The generalized framework can exist in the absence of an equivalence relation that was the basis of the original rough set proposal. The framework used the abstract notion of positive, negative, and boundary regions from rough set theory to describe a three-way decision theory. The three decisions consist of acceptance of, rejection of, or non-commitment to a hypothesis. The most general case uses a preference relation. The conventional rough sets based on equivalence relations can be formulated using the preference relations. Similarly, the decision theoretic rough sets are described using preference relations.

This special issue describes further developments in decision theoretic rough set. Jia et al. address the issue of optimization in decision theoretic rough set [48], which is further explored by Shao et al. [49] through complexity reduction in generating decision rules. Li and Yang [50] focus on axiomatic characteristics of probabilistic rough set. Liu et al. [51] broaden the traditional classification in rough set theory by using logistic regression based on decision theory. Qian et al. [52] explore the decision theoretic rough set from granular computing perspective. Azam and Yao [53] incorporate the game theoretic aspects, an important domain of decision theory, in rough set context. Feature selection is one of the major applications of rough set theory, which is studied using cost considerations by Min et al. [54]. The issue of incomplete data in probabilistic rough set is addressed by Grzymala-Busse et al. [55].

The formulation of three-way decision theory is based on binary classification. Lingras et al. [41] generalized the three-way decision making approach to multi-class problem and used it to measure the quality of unsupervised clustering [40,42]. A more rigorous treatment of decision theoretic framework for multi-class rough set can be found in [10], as well as in this special issue by Zhou [56]. Since this paper focuses on the unsupervised clustering problem rather than supervised classification problem, we will use category as a more general term to denote either the classes or clusters. When the three-way decision theoretic approach is extended to k category problem, we are faced with $2^k - 1$ decisions.

2.5. Decision theory and clustering

As mentioned before, Lingras used the multi-class $(2^k - 1)$-way decision theoretic model for studying rough clustering within a decision theoretic framework [40]. They used the decision theoretic framework to propose a cost-benefit analysis for evaluating cluster quality of crisp and rough clustering [42]. They described how the cost-benefit proposal can be applied for making decisions about clustering of customers for promotional campaigns in a retail store. The implications of using decision theoretic rough sets with rough clustering is also addressed in this special issue by Li et al. [57]. Yu et al. in a series of publications [43,44,58] studied the problem of soft clustering within the decision theoretic framework. Yu et al. [58] applied decision theoretic rough set theory to agglomerative hierarchical clustering, which is further refined in this special issue [59]. The hierarchical agglomerative clustering starts with the number of clusters to be the same as the number of objects and successively merges those clusters that will result in the least degradation of quality. Yu et al.'s approach suggested a decision theoretic rough set based cluster quality index that stopped the hierarchical agglomeration to provide the optimal number of clustering. Their approach eliminates the need to predetermine the number of clusters. However, it can be computationally expensive for large datasets.

Yu and Wang [44] used a probability function and the three-way decision theory to create soft/overlapping clusters. For an object, the acceptance decision for a cluster i will result in the inclusion of the object to the lower bound of the cluster. The non-commitment decision leads to the object being a part of the upper bound of cluster, and the rejection decision means that the object does not belong to the upper bound of the cluster. The probability function used by Yu and Wang was density based. However, their approach can be applied with probability functions based on any other criteria.

The concept of automated or parameter-less is further extended by Yu and Chu [43] in a knowledge-oriented clustering, where they provide a more efficient version of hierarchical agglomerative clustering by starting with finer granules of clusters and merging them using a cluster quality measure based on decision theoretic rough sets.

3. Qualitative crisp clustering

We will be dealing with clustering schemes based on multiple criteria. Therefore, we will use subscripts to distinguish them from each other. Each clustering scheme cs_v partitions X into k_v clusters, i.e. $cs_v = \{\vec{c}_{v1}, \ldots, \vec{c}_{vk_v}\}$. As mentioned earlier, $\vec{c}_{vt}$ represents the centroid vector of the cluster and c_{vt} represents the set of objects that belong to the cluster.

Unlike classification, clustering does not have any known decision classes. So the value of decision class cannot be used in creating the set of dominance relations. However, if the clustering is based on monolithic criteria such as visit pattern for customers to a retail store, one can look at the resulting clusters and provide a partial order among the clusters in cs_v. Number of visits can be used as an indication of the customer loyalty. The criteria may be simple such as total visits or a little more complicated based on monthly visits distribution such as 50th, 85th, and 95th percentile monthly visit values [40]. In either case, one can identify some clusters that contain more loyal customers than others.

In crisp clustering, we will use only the strict preference and indifference relations to specify the partial ordering between objects and cluster centroids. All the objects will be similar to exactly one centroid. That is for each object $\vec{x}_t$, there exists one and only one $\vec{c}_{vi}$ such that $\vec{x}_t \sim_v \vec{c}_{vi}$. For the two centroids $\vec{c}_{vi}$ and $\vec{c}_{vj}$ there are three possibilities:

- $\vec{c}_{vi} \succ_v \vec{c}_{vj}$.
- $\vec{c}_{vj} \succ_v \vec{c}_{vi}$.
- There is no relation between $\vec{c}_{vi}$ and $\vec{c}_{vj}$. We will write this as: $\vec{c}_{vi} ?_v \vec{c}_{vj}$.

Relationship between individual pair of objects and centroid vectors can be inferred by the transitivity axioms for strict preference and indifference relations as well as the following axiom:

$$\begin{aligned} &\vec{x}_s \sim_v \vec{c}_{vi} \wedge \vec{x}_t \sim_v \vec{c}_{vj} \wedge \vec{c}_{vi} \succ_v \vec{c}_{vj}, \\ &\Updownarrow \\ &\vec{x}_s \succ_v \vec{x}_t \wedge \vec{x}_s \succ_v \vec{c}_{vj} \wedge \vec{c}_{vi} \succ_v \vec{x}_t. \end{aligned} \tag{7}$$

If one cannot infer any relationship between two objects $\vec{x}_s$ and $\vec{x}_t$, we assume that there is no relation, i.e. $\vec{x}_s ?_v \vec{x}_t$.

The dominance relations do not fully capture the extent of dominance of one object over another. Therefore, we introduce the notion of degree of dominance. In order to understand the notion of degree of dominance, we need to extend the notations of dominance relations.

Let cs_v be a clustering scheme. Let $\vec{x}_{t_1}, \vec{x}_{t_1+1}, \ldots, \vec{x}_{t_1+t_2}$ be $t_1 + t_2 + 1$ vectors representing objects or cluster centroids. Let us assume that

$$\vec{x}_{t_1} \succ_v \vec{x}_{t_1+1}, \vec{x}_{t_1+1} \succ_v \vec{x}_{t_1+2}, \ldots, \vec{x}_{t_1+t_2-1} \succ_v \vec{x}_{t_1+t_2}.$$

We will abbreviate such ordering as follows:

$$\vec{x}_{t_1} \succ_v, \ldots, \succ_v \vec{x}_{t_1+t_2}.$$

Degree of the dominance relation: If there exist at most d ($d \geqslant 0$) vectors between $\vec{x}_i$ and $\vec{x}_j$ so that $\vec{x}_i \succ_v \vec{x}_1 \succ_v, \ldots, \succ_v \vec{x}_d \succ_v \vec{x}_j$, then the degree of the dominance relation under the clustering scheme cs_v is $d + 1$. It can be expressed as follows:

$$\vec{x}_i \succ_v \vec{x}_1 \succ_v, \ldots, \succ_v \vec{x}_d \succ_v \vec{x}_j \ (d \geqslant 0) \rightarrow degree(\vec{x}_i \succ_v \vec{x}_j) = d + 1.$$

If $\neg(\vec{x}_i \succ_v \vec{x}_j)$ then $degree(\vec{x}_i \succ_v \vec{x}_j)$ is defined as follows:

- If $\vec{x}_j \succ_v \vec{x}_i$, then $degree(\vec{x}_i \succ_v \vec{x}_j) = -degree(\vec{x}_j \succ_v \vec{x}_i)$.
- If $\vec{x}_i \sim_v \vec{x}_j$, then $degree(\vec{x}_i \sim_v \vec{x}_j) = 0$.
- If no relation exists between $\vec{x}_i$ and $\vec{x}_j$, i.e. $\vec{x}_i ?_v \vec{x}_j$, then $degree(\vec{x}_i \succ_v \vec{x}_j) = 0$.

We can derive the following useful properties for degrees of dominance:

- If $\vec{x}_i \in \vec{c}_{vt_1}$ and $\vec{x}_j \in \vec{c}_{vt_2}$, then $degree(\vec{x}_i \succ_v \vec{x}_j) = degree(\vec{c}_{vt_1} \succ_v \vec{c}_{vt_2})$.
- If $\vec{x}_i \succ_v \vec{x}_1 \succ_v, \ldots, \succ_v \vec{x}_d \succ_v \vec{x}_j$, then $degree(\vec{x}_i \succ_v \vec{x}_j) = degree(\vec{x}_i \succ_v \vec{x}_h) + degree(\vec{x}_h \succ_v \vec{x}_j)$ $(1 \leqslant h \leqslant d)$.
- If $\vec{x}_i \succ_v \vec{x}_j$, then $degree(\vec{x}_i \succ_v \vec{x}_j) > 0$.
- If $\vec{x}_j \succ_v \vec{x}_i$, then $degree(\vec{x}_i \succ_v \vec{x}_j) < 0$.

Example 1. Let $X = \{\vec{x}_1, \ldots, \vec{x}_{18}\}$ be customers to a retail store. We first cluster the customers based on their visit pattern as a surrogate for loyalty, resulting in a clustering scheme cs_v. Fig. 1 shows that cs_v partitions the objects into two clusters denoted as $cs_v = \{\vec{c}_{v1}, \vec{c}_{v2}\}$. Clusters are separated by dashed line. The customers in cluster c_{v1} visit more frequently than those in c_{v2}. Therefore, we get $\vec{c}_{v1} \succ_v \vec{c}_{v2}$.

Let us also cluster these customers based on their spending habits as shown in Fig. 2. The resulting spending based clustering scheme cs_s partitions the objects into four clusters denoted as $cs_s = \{\vec{c}_{s1}, \vec{c}_{s2}, \vec{c}_{s3}, \vec{c}_{s4}\}$. We can create a strict preference order based on spending as $\vec{c}_{s1} \succ_s \vec{c}_{s2} \succ_s \vec{c}_{s3} \succ_s \vec{c}_{s4}$. Please note that while the objects are partially ordered, it is assumed that there is a total order among clusters.

Table 1 shows the relations between objects under the clustering scheme based on visits. Table 2 shows the relations between objects under the clustering scheme based on spending.

Table 3 and Table 4 show the degrees of the dominance in cs_v and cs_s, respectively.

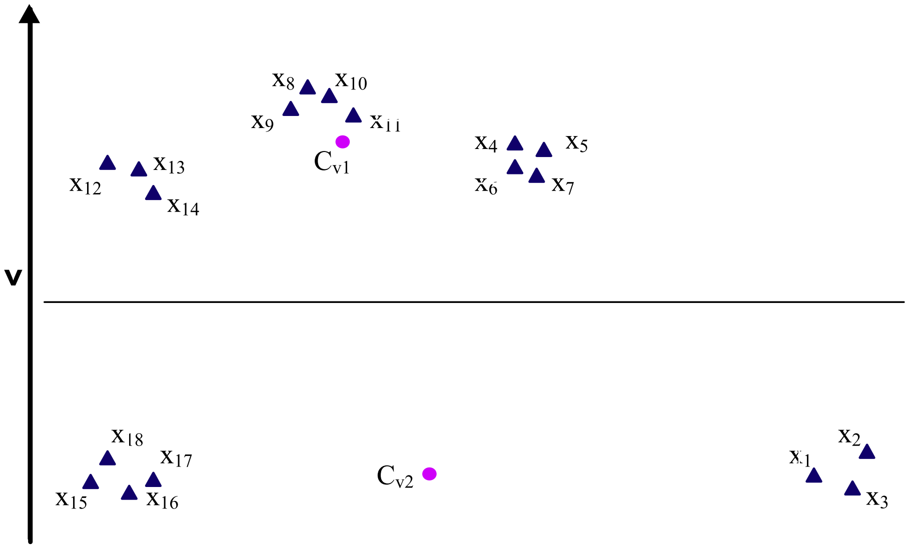

Fig. 1. The partition of objects by cs_v.

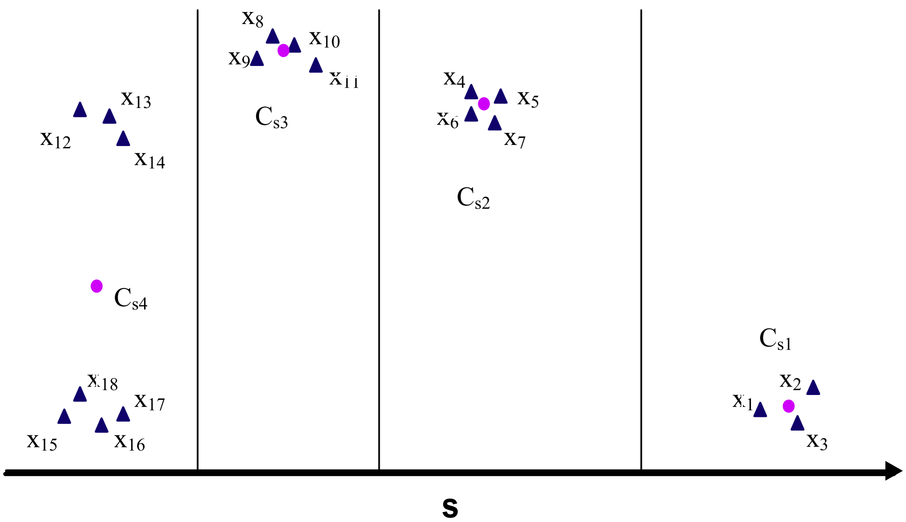

Fig. 2. The partition of objects by cs_s.

Table 1
$\vec{x}_i \succ_v \vec{x}_j$.

$\vec{c}_{vj}$ / $\vec{c}_{vi}$		$\vec{c}_{v1}$	$\vec{c}_{v2}$
	$\vec{x}_j$ / $\vec{x}_i$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7, \vec{x}_8, \vec{x}_9$ $\vec{x}_{10}, \vec{x}_{11}, \vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$ $\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$
$\vec{c}_{v1}$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7, \vec{x}_8, \vec{x}_9$ $\vec{x}_{10}, \vec{x}_{11}, \vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	$\sim_v$	$\succ_v$
$\vec{c}_{v2}$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$ $\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$	–	$\sim_v$

4. Qualitative combination of crisp clustering scheme

Different clustering schemes are usually based on different criteria. For example, customers for a retail store may be clustered separately on their spending potential and their loyalty. Actual spending patterns can be used as a measure of their spending potential. Their frequency of visits can be used as a surrogate measure of their loyalty. In order to group these customers on both the criteria, we will have to combine the two clustering schemes. This section proposes a rule for qualitative combination of clustering scheme. In the next section, we will analyze corresponding quantitative combination.

Table 2
$\vec{x}_i \succ_s \vec{x}_j$.

$\vec{c}_{si}$ \ $\vec{c}_{sj}$		$\vec{c}_{s1}$	$\vec{c}_{s2}$	$\vec{c}_{s3}$	$\vec{c}_{s4}$
	$\vec{x}_i$ \ $\vec{x}_j$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	$\vec{x}_4, \vec{x}_5$ $\vec{x}_6, \vec{x}_7$	$\vec{x}_8, \vec{x}_9$ $\vec{x}_{10}, \vec{x}_{11}$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$ $\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$
$\vec{c}_{s1}$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	$\sim_s$	$\succ_s$	$\succ_s$	$\succ_s$
$\vec{c}_{s2}$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$	–	$\sim_s$	$\succ_s$	$\succ_s$
$\vec{c}_{s3}$	$\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$	–	–	$\sim_s$	$\succ_s$
$\vec{c}_{s4}$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$ $\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$	–	–	–	$\sim_s$

Table 3
$degree(\vec{x}_i \succ_v \vec{x}_j)$.

$\vec{c}_{vi}$ \ $\vec{c}_{vj}$		$\vec{c}_{v1}$	$\vec{c}_{v2}$
	$\vec{x}_i$ \ $\vec{x}_j$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7, \vec{x}_8, \vec{x}_9$ $\vec{x}_{10}, \vec{x}_{11}, \vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$ $\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$
$\vec{c}_{v1}$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7, \vec{x}_8, \vec{x}_9$ $\vec{x}_{10}, \vec{x}_{11}, \vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	0	1
$\vec{c}_{v2}$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$ $\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$	−1	0

Table 4
$degree(\vec{x}_i \succ_s \vec{x}_j)$.

$\vec{c}_{si}$ \ $\vec{c}_{sj}$		$\vec{c}_{s1}$	$\vec{c}_{s2}$	$\vec{c}_{s3}$	$\vec{c}_{s4}$
	$\vec{x}_i$ \ $\vec{x}_j$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	$\vec{x}_4, \vec{x}_5$ $\vec{x}_6, \vec{x}_7$	$\vec{x}_8, \vec{x}_9$ $\vec{x}_{10}, \vec{x}_{11}$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$ $\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$
$\vec{c}_{s1}$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	0	1	2	3
$\vec{c}_{s2}$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$	−1	0	1	2
$\vec{c}_{s3}$	$\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$	−2	−1	0	1
$\vec{c}_{s4}$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$ $\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$	−3	−2	−1	0

The qualitative combination can be efficiently implemented by first taking the intersection of the clusters from original clustering schemes.

Let cs_v and cs_s be two clustering schemes.
$cs_v = \{\vec{c}_{v1}, \ldots, \vec{c}_{vk_v}\}$,
$cs_s = \{\vec{c}_{s1}, \ldots, \vec{c}_{sk_s}\}$,

k_v and k_s are the number of clusters for cs_v and cs_s, respectively. Combination of clustering schemes cs_v and cs_s will be given as:

$$cs_{v\oplus s} = \{\vec{c}_{vi} \cap \vec{c}_{sj} \mid \vec{c}_{vi} \cap \vec{c}_{sj} \neq \emptyset\}.$$

Let $k_{v\oplus s}$ be the number of clusters for $cs_{v\oplus s}$. It can be easily verified that

$$max(k_v, k_s) \leqslant k_{v\oplus s} \leqslant k_v * k_s.$$

The combined clusters will be formed by groups of similar objects given by the combined indifference relation, $\sim_{v\oplus s}$ defined as follows:

If $\vec{x}_t, \vec{x}_r \in \vec{c}_i \cap \vec{c}_j$, then $\vec{x}_t \sim_{v\oplus s} \vec{x}_r$.

We also need to calculate the dominance relation and their degrees of dominance. In order to calculate the dominance, we define the concept of weight that will make it possible for us to attach different importance to different clustering schemes.

Weight: Let $\succ_v$ and $\succ_s$ be two relations. Without loss of generality, we will assume that $\succ_v$ is more important than $\succ_s$. The weight between $\succ_v$ and $\succ_s$ will be an integer value greater than or equal to 1 denoted by $\frac{\succ_v}{\succ_s}$. We can also specify the inverse as $\frac{\succ_s}{\succ_v} = \frac{1}{\frac{\succ_v}{\succ_s}}$.

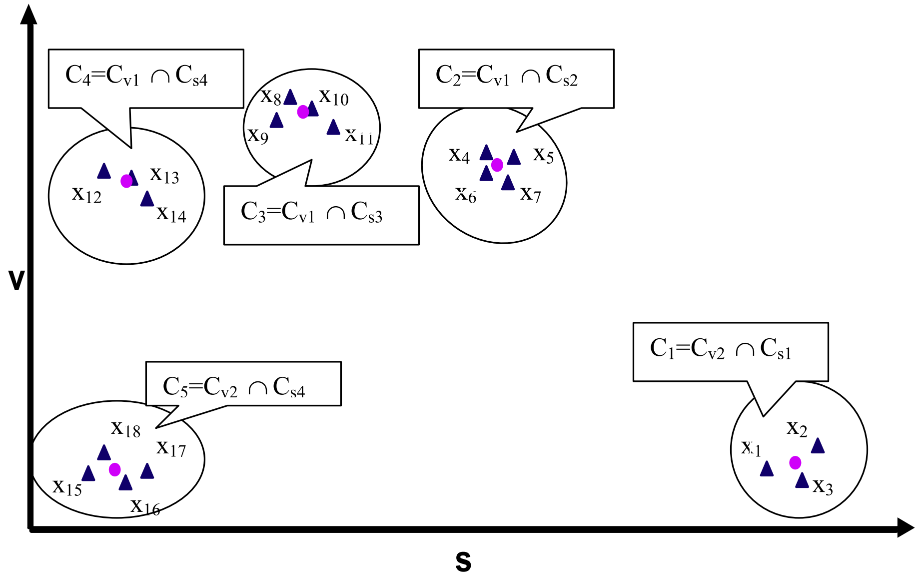

Fig. 3. The partition of objects with visits and spending.

Table 5
$degree(\vec{x}_i \succ_{v\oplus s} \vec{x}_j)$ when $\frac{\succ_v}{\succ_s} = 1$.

$\vec{c}_i$ \ $\vec{c}_j$		$\vec{c}_1$	$\vec{c}_2$	$\vec{c}_3$	$\vec{c}_4$	$\vec{c}_5$
	$\vec{x}_i$ \ $\vec{x}_j$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$	$\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	$\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$
$\vec{c}_1$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	0	0	1	2	3
$\vec{c}_2$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$	0	0	1	2	3
$\vec{c}_3$	$\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$	−1	−1	0	1	2
$\vec{c}_4$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	−2	−2	−1	0	1
$\vec{c}_5$	$\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$	−3	−3	−2	−1	0

Assuming $\frac{\succ_v}{\succ_s} \geqslant 1$, we define the degree of dominance for the combined clustering scheme as:

$$degree(\vec{c}_{vi} \cap \vec{c}_{st} \succ_{v\oplus s} \vec{c}_{vj} \cap \vec{c}_{sr}) = \frac{\succ_v}{\succ_s} * degree(\vec{c}_{vi} \succ_v \vec{c}_{vj}) + degree(\vec{c}_{st} \succ_s \vec{c}_{sr}).$$

Let $\vec{c}_i, \vec{c}_j \in c_{v\oplus s}$. The degree defined above will then be used to define the strict preference relation between $\vec{c}_i$ and $\vec{c}_j$ as:

- If $degree(\vec{c}_i \succ_{v\oplus s} \vec{c}_j) > 0$ then $\vec{c}_i \succ_{v\oplus s} \vec{c}_j$
- else if $degree(\vec{c}_i \succ_{v\oplus s} \vec{c}_j) < 0$ then $\vec{c}_j \succ_{v\oplus s} \vec{c}_i$
- else if $\neg(\vec{c}_i \sim_{v\oplus s} \vec{c}_j)$ and $degree(\vec{c}_i \succ_{v\oplus s} \vec{c}_j) = 0$ there is no relation between $\vec{c}_i$ and $\vec{c}_j$, i.e. $\vec{c}_i ?_{v\oplus s} \vec{c}_j$.

Relationship between individual pair of objects in two different combined clusters can be inferred by the transitivity axioms for strict preference and indifference relations as well as axiom 7 that defines the transitivity between $\sim$ and $\succ$.

Example 2. We use the same dataset presented in Example 1. We want to group the customers based on a combination of loyalty and spending. It is obvious that the combined clustering scheme will have five clusters. As can be seen from Fig. 3, five clusters $C = \{\vec{c}_1, \ldots, \vec{c}_5\}$ are represented by centroid vectors outlined by solid lines. Table 5, Table 7, and Table 9 present the degrees of dominance for the combined clustering schemes, $cs_{v\oplus s}$, by using $\frac{\succ_v}{\succ_s} = 1$, $\frac{\succ_v}{\succ_s} = 3$ and $\frac{\succ_s}{\succ_v} = 3$, respectively. Table 6, Table 8 and Table 10 show the corresponding dominance relationships.

5. Qualitative and quantitative combination of crisp clustering

The previous section discussed qualitative combination of two clustering schemes. In this section, we discuss the conventional quantitative alternative used for clustering based on multicriteria, and study the resulting set of dominance relations. The conventional solution to combine two criteria is to multiply the numeric values of the two criteria with different weights depending on their relative importance. The weighted numeric values are then clustered using an algorithm such as K-means.

Table 6
$\vec{x}_i \succ_{v\oplus s} \vec{x}_j$ when $\frac{\succ_v}{\succ_s} = 1$.

$\vec{c}_i$ \ $\vec{c}_j$		$\vec{c}_1$	$\vec{c}_2$	$\vec{c}_3$	$\vec{c}_4$	$\vec{c}_5$
	$\vec{x}_i$ \ $\vec{x}_j$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$	$\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	$\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$
$\vec{c}_1$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	$\sim_{v\oplus s}$	$?_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_2$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$	$?_{v\oplus s}$	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_3$	$\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$	–	–	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_4$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	–	–	–	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_5$	$\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$	–	–	–	–	$\sim_{v\oplus s}$

Table 7
$degree(\vec{x}_i \succ_{v\oplus s} \vec{x}_j)$ when $\frac{\succ_v}{\succ_s} = 3$.

$\vec{c}_i$ \ $\vec{c}_j$		$\vec{c}_1$	$\vec{c}_2$	$\vec{c}_3$	$\vec{c}_4$	$\vec{c}_5$
	$\vec{x}_i$ \ $\vec{x}_j$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$	$\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	$\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$
$\vec{c}_1$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	0	−2	−1	0	3
$\vec{c}_2$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$	2	0	1	2	5
$\vec{c}_3$	$\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$	1	−1	0	1	4
$\vec{c}_4$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	0	−2	−1	0	3
$\vec{c}_5$	$\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$	−3	−5	−4	−3	0

Table 8
$\vec{x}_i \succ_{v\oplus s} \vec{x}_j$ when $\frac{\succ_v}{\succ_s} = 3$.

$\vec{c}_i$ \ $\vec{c}_j$		$\vec{c}_1$	$\vec{c}_2$	$\vec{c}_3$	$\vec{c}_4$	$\vec{c}_5$
	$\vec{x}_i$ \ $\vec{x}_j$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$	$\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	$\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$
$\vec{c}_1$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	$\sim_{v\oplus s}$	–	–	$?_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_2$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$	$\succ_{v\oplus s}$	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_3$	$\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$	$\succ_{v\oplus s}$	–	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_4$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	$?_{v\oplus s}$	–	–	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_5$	$\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$	–	–	–	–	$\sim_{v\oplus s}$

Table 9
$degree(\vec{x}_i \succ_{v\oplus s} \vec{x}_j)$ when $\frac{\succ_s}{\succ_v} = 3$.

$\vec{c}_i$ \ $\vec{c}_j$		$\vec{c}_1$	$\vec{c}_2$	$\vec{c}_3$	$\vec{c}_4$	$\vec{c}_5$
	$\vec{x}_i$ \ $\vec{x}_j$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$	$\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	$\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$
$\vec{c}_1$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	0	2	5	8	9
$\vec{c}_2$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$	−2	0	3	6	7
$\vec{c}_3$	$\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$	−5	−3	0	3	4
$\vec{c}_4$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	−8	−6	−3	0	1
$\vec{c}_5$	$\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$	−9	−7	−4	−1	0

Table 10
$\vec{x}_i \succ_{v\oplus s} \vec{x}_j$ when $\frac{\succ_s}{\succ_v} = 3$.

$\vec{c}_i$ \ $\vec{c}_j$		$\vec{c}_1$	$\vec{c}_2$	$\vec{c}_3$	$\vec{c}_4$	$\vec{c}_5$
	$\vec{x}_i$ \ $\vec{x}_j$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$	$\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	$\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$
$\vec{c}_1$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_2$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$	–	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_3$	$\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$	–	–	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_4$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	–	–	–	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_5$	$\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$	–	–	–	–	$\sim_{v\oplus s}$

We use a synthetic retail dataset that is similar to the small dataset presented in Examples 1 and 2 to analyze the quantitative combination of crisp clustering. The structure of the synthetic dataset – shown in Fig. 4 – is the same as the one used in Examples 1 and 2. The five clusters in Fig. 4 are in the same region in the combined space as that in Fig. 3. The only difference being that the clusters in synthetic dataset have more objects to further validate the feasibility of our approach.

The quantitative combination was done by varying the pair of weights for (v, s) from $(0, 1)$ to $(1, 0)$ with an increment or decrement of 0.1. For each pair of weights we got five clusters as shown in Fig. 4. The combined clusters were ranked based on the values of their centroid vectors. Let $\vec{c}_i = (c_{vi}, c_{si})$. The value of $\vec{c}_i = \sqrt{c_{vi}^2 + c_{si}^2}$.

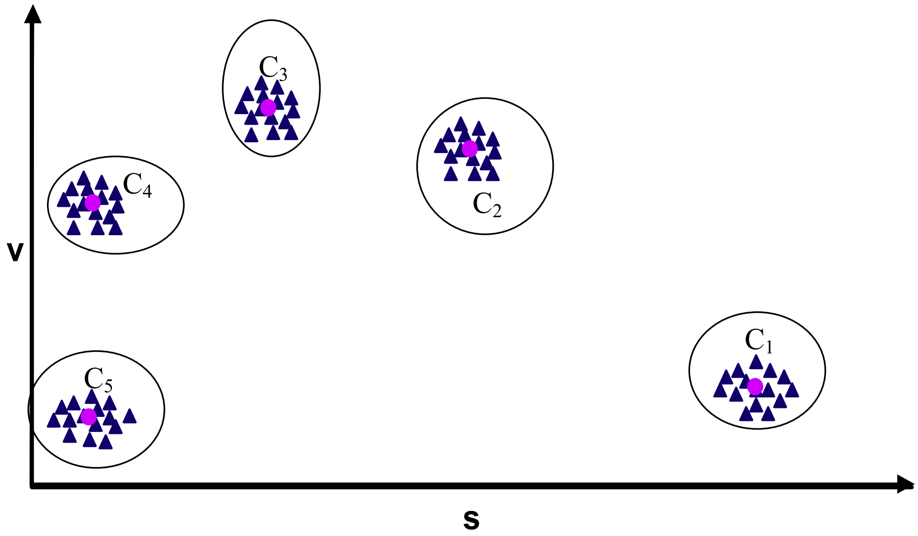

Fig. 4. Synthetic data for combining crisp clustering.

Table 11
Dominance relations between clusters for crisp clustering for synthetic data.

ω_v	ω_s	Dominance relations	$\frac{\succ_v}{\succ_s}$	Dominance relations
0.9	0.1	$\vec{c}_3 \succ \vec{c}_2 \succ \vec{c}_4 \succ \vec{c}_1 \succ \vec{c}_5$	9	$\vec{c}_2 \succ \vec{c}_3 \succ \vec{c}_4 \succ \vec{c}_1 \succ \vec{c}_5$
0.8	0.2		4	
0.75	0.25		3	$\vec{c}_2 \succ \vec{c}_3 \succ \vec{c}_1 \succ \vec{c}_5$ $\vec{c}_2 \succ \vec{c}_3 \succ \vec{c}_4 \succ \vec{c}_5$ $\vec{c}_1 ? \vec{c}_4$
0.7	0.3		$\frac{7}{3}$	$\vec{c}_2 \succ \vec{c}_3 \succ \vec{c}_1 \succ \vec{c}_4 \succ \vec{c}_5$
0.6	0.4		$\frac{3}{2}$	$\vec{c}_2 \succ \vec{c}_1 \succ \vec{c}_3 \succ \vec{c}_4 \succ \vec{c}_5$
0.5	0.5	$\vec{c}_2 \succ \vec{c}_3 \succ \vec{c}_1 \succ \vec{c}_4 \succ \vec{c}_5$	1	$\vec{c}_1 \succ \vec{c}_3 \succ \vec{c}_4 \succ \vec{c}_5$ $\vec{c}_2 \succ \vec{c}_3 \succ \vec{c}_4 \succ \vec{c}_5$ $\vec{c}_1 ? \vec{c}_2$
0.4	0.6	$\vec{c}_1 \succ \vec{c}_2 \succ \vec{c}_3 \succ \vec{c}_4 \succ \vec{c}_5$	$\frac{2}{3}$	$\vec{c}_1 \succ \vec{c}_2 \succ \vec{c}_3 \succ \vec{c}_4 \succ \vec{c}_5$
0.3	0.7		$\frac{3}{7}$	
0.25	0.75		$\frac{1}{3}$	
0.2	0.8		$\frac{1}{4}$	
0.1	0.9		$\frac{1}{9}$	

Since the synthetic dataset in Fig. 4 is similar to the one used in Examples 1 and 2, the resulting qualitative combination yields the same results as shown in Example 2. Table 11 shows the results for selected weights that produce dominance relations similar to the qualitative combinations using three weights.

Table 11 shows a correspondence between qualitative and quantitative combinations. The qualitative combinations allow us to specify no relation between clusters, while we will usually have a strict dominance relation between quantitatively combined clusters. Sometimes such a partial dominance relation may be more suitable. For example, customers in cluster c_1 are less loyal but bigger spenders than those in cluster c_2. Therefore, it is not obvious that c_1 is better or worse than c_2, when both visits and spending are equally weighted. This fact is captured by the qualitative combination, which chooses to specify no relation between the two clusters. The qualitative combination may also be useful in cases where one can only communicate individual clustering schemes due to privacy considerations in sharing attribute values or due to large communication costs. It can also be useful for simply devising an appropriate weighting scheme for a quantitative combination scheme.

6. Qualitative rough clustering

The rough set based clustering makes it possible for the cluster boundaries to overlap through the introduction of a boundary region. Let us consider a rough clustering scheme $rc_v = \{\vec{c}_{v1}, \ldots, \vec{c}_{vk_v}\}$. Each rough cluster c_{vi} will be represented by a rough set $(lower(c_{vi}), upper(c_{vi}))$. Let $bnd(c_{vi}) = upper(c_{vi}) - lower(c_{vi})$ be the boundary region of c_{vi}.

Similar to its crisp counterpart, in rough clustering, we will use the strict preference relation to specify the partial ordering between cluster centroids $\vec{c}_{vi}$. For the two centroids $\vec{c}_{vi}$ and $\vec{c}_{vj}$ there are three possibilities:

- $\vec{c}_{vi} \succ_v \vec{c}_{vj}$.
- $\vec{c}_{vj} \succ_v \vec{c}_{vi}$.
- $\vec{c}_{vi} ?_v \vec{c}_{vj}$.

All the objects in $lower(c_{vi})$ will be similar to exactly one centroid. That is, for each object $\vec{x}_t \in lower(c_{vi})$, there exists one and only one $\vec{c}_{vi}$ such that $\vec{x}_t \sim_v \vec{c}_{vi}$.

Relationship between individual pair of objects in lower bounds of clusters can be defined as follows:

- If $\vec{x}_t, \vec{x}_r \in lower(c_{vi})$, then $\vec{x}_t \sim_v \vec{x}_r$.
- If $\vec{x}_t \in lower(c_{vi})$ and $\vec{x}_r \in lower(c_{vj})$ and $\vec{c}_{vi} \succ_v \vec{c}_{vj}$, then $\vec{x}_t \succ_v \vec{x}_r$.
- If $\vec{x}_t \in lower(c_{vi})$ and $\vec{x}_r \in lower(c_{vj})$ and $\vec{c}_{vi} ?_v \vec{c}_{vj}$, then $\vec{x}_t ?_v \vec{x}_r$.

The objects in the boundary region belong to two or more clusters. This fact can be accommodated in dominance based qualitative framework by the addition of a *roughly similar* relation $\backsim_v$ for a rough clustering scheme rc_v. An object $\vec{x}_b$ in the boundary region $bnd(c_{vi})$ will be related to $\vec{c}_{vi}$ as $\vec{x}_b \backsim_v \vec{c}_{vi}$. For every object $\vec{x}_b$ in the boundary region, there will exist at least two clusters $\vec{c}_{vi}$ and $\vec{c}_{vj}$ such that $\vec{x}_b \backsim_v \vec{c}_{vi}$ and $\vec{x}_b \backsim_v \vec{c}_{vj}$.
Assuming that $\vec{x}_i \neq \vec{x}_j$, the roughly similar relation obeys the following properties:

Symmetry: $\vec{x}_i \backsim_v \vec{x}_j \Rightarrow \vec{x}_j \backsim_v \vec{x}_i$.

Reflexivity: $\vec{x}_i \backsim_v \vec{x}_i$.

Unlike the indifference relation, the roughly similar relation does not obey the transitivity axiom. Instead, we infer the roughly similar relations through the cluster centroids as follows:

- If $\vec{x}_a \backsim_v \vec{c}_{vi}$ and $\vec{x}_b \backsim_v \vec{c}_{vi}$, then $\vec{x}_a \backsim_v \vec{x}_b$.
- If $\vec{x}_a \backsim_v \vec{c}_{vi}$ and $\vec{x}_b \sim_v \vec{c}_{vi}$, then $\vec{x}_a \backsim_v \vec{x}_b$.

Based on the indifference and roughly similar relations, we can define the lower, upper, and boundary regions as follows:

$lower(c_{vi}) = \{\vec{x}_t | \vec{x}_t \sim_v \vec{c}_{vi}\}$

$upper(c_{vi}) = \{\vec{x}_u | \vec{x}_u \backsim_v \vec{c}_{vi} \vee (\vec{x}_u \sim_v \vec{c}_{vi})\}$.

$bnd(c_{vi}) = \{\vec{x}_b | \vec{x}_b \backsim_v \vec{c}_{vi} \wedge \neg(\vec{x}_b \sim_v \vec{c}_{vi})\}$.

The degree of dominance for cluster centroids and objects in their lower bounds can be defined using the same definition as crisp clustering.

If either one or both of the objects $\vec{x}_a$ and $\vec{x}_b$ are in the boundary region, $degree(\vec{x}_a \succ_v \vec{x}_b)$ needs to defined using a rough value [60] $(\underline{d}, \overline{d})$ as follows.

Let $A = \{\vec{c}_{vi} | \vec{x}_a \in upper(c_{vi})\}$ and $B = \{\vec{c}_{vj} | \vec{x}_b \in upper(c_{vj})\}$.

$$\underline{d} = \min\{degree(\vec{c}_{vi} \succ_v \vec{c}_{vj}) | \vec{c}_{vi} \in A, \vec{c}_{vj} \in B\},$$

$$\overline{d} = \max\{degree(\vec{c}_{vi} \succ_v \vec{c}_{vj}) | \vec{c}_{vi} \in A, \vec{c}_{vj} \in B\}.$$

It should be noted that if both the objects $\vec{x}_a$ and $\vec{x}_b$ are in lower bounds of clusters, the above definition yields a rough value such that:

$$\underline{d} = \overline{d} = degree(\vec{x}_a \succ_v \vec{x}_b).$$

If $\vec{x}_a$ and $\vec{x}_b$ are in boundary regions and $\neg(\vec{x}_i \backsim_v \vec{x}_j)$, the rough degree of dominance can be used to define dominance relations between them as follows:

- Let $degree(\vec{x}_a \succ_v \vec{x}_b) = (\underline{d}, \overline{d})$.
- If $\underline{d} = \overline{d} = 0$, then $\vec{x}_a ?_v \vec{x}_b$
- else if $\underline{d} = 0 \wedge \overline{d} > 0$, then $\vec{x}_a \succeq_v \vec{x}_b$
- else if $\underline{d} > 0 \wedge \overline{d} > 0$, then $\vec{x}_a \succ_v \vec{x}_b$.

Example 3. We will illustrate the usefulness of rough clustering schemes by modifying our small customer dataset from Example 1 by adding a few boundary region objects. As shown in Fig. 5, the rough clustering rs_v of customers based on visits still broadly falls into two clusters $rs_v = \{\vec{c}_{v1}, \vec{c}_{v2}\}$. The objects previously in the crisp clusters are now in the lower bound of those clusters as shown in Table 12. However, now we have three objects in the boundary region of these two clusters, namely, $\vec{x}_{19}, \vec{x}_{20}, \vec{x}_{21}$. Their rough degrees of dominance are given in Table 13.

The clustering of the augmented dataset based on spending rs_s also groups the objects into four clusters $rs_s = \{\vec{c}_{s1}, \vec{c}_{s2}, \vec{c}_{s3}, \vec{c}_{s4}\}$. The objects previously in the crisp clusters are now in the lower bound of those clusters as shown in Table 14.

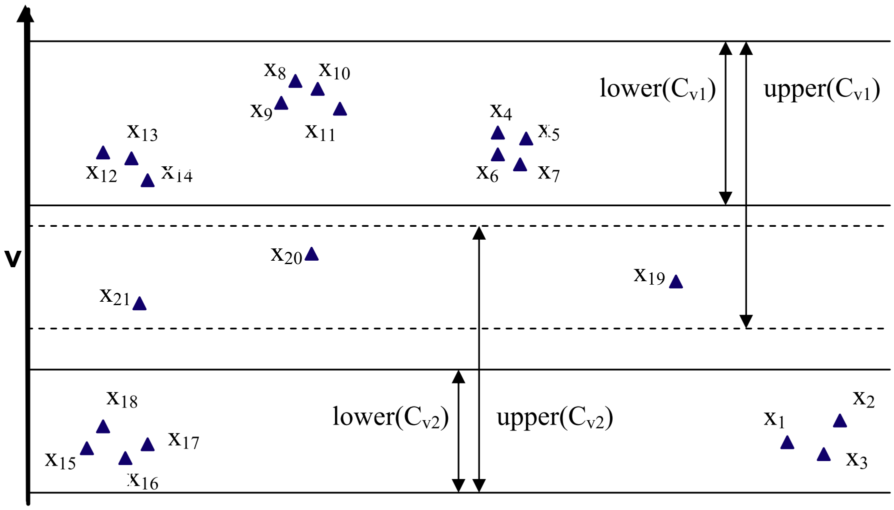

Fig. 5. The rough partition of objects by rs_v.

Table 12
$\vec{x}_i \succ_v \vec{x}_j$.

$\vec{c}_{vi}$ \ $\vec{c}_{vj}$		$lower(c_{v1})$	$lower(c_{v2})$	$bnd(c_{v1}), bnd(c_{v2})$		
	$\vec{x}_i$ \ $\vec{x}_j$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$ $\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$ $\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$ $\vec{x}_{15}, \vec{x}_{16}$ $\vec{x}_{17}, \vec{x}_{18}$	$\vec{x}_{19}$	$\vec{x}_{20}$	$\vec{x}_{21}$
$lower(c_{v1})$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$ $\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$ $\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	$\sim_v$	$\succ_v$	$\wp_v$	$\wp_v$	$\wp_v$
$lower(c_{v2})$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$ $\vec{x}_{15}, \vec{x}_{16}$ $\vec{x}_{17}, \vec{x}_{18}$	–	$\sim_v$	$\wp_v$	$\wp_v$	$\wp_v$
$bnd(c_{v1})$ $bnd(c_{v2})$	$\vec{x}_{19}$ / $\vec{x}_{20}$ / $\vec{x}_{21}$	$\wp_v$	$\wp_v$	$\wp_v$	$\wp_v$	$\wp_v$

Table 13
$degree(\vec{x}_i \succ_v \vec{x}_j)$.

$\vec{c}_{vi}$ \ $\vec{c}_{vj}$		$lower(c_{v1})$	$lower(c_{v2})$	$bnd(c_{v1}), bnd(c_{v2})$		
	$\vec{x}_i$ \ $\vec{x}_j$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$ $\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$ $\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$ $\vec{x}_{15}, \vec{x}_{16}$ $\vec{x}_{17}, \vec{x}_{18}$	$\vec{x}_{19}$	$\vec{x}_{20}$	$\vec{x}_{21}$
$lower(c_{v1})$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$ $\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$ $\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	0	1	(0, 1)	(0, 1)	(0, 1)
$lower(c_{v2})$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$ $\vec{x}_{15}, \vec{x}_{16}$ $\vec{x}_{17}, \vec{x}_{18}$	−1	0	(−1, 0)	(−1, 0)	(−1, 0)
$bnd(c_{v1})$ $bnd(c_{v2})$	$\vec{x}_{19}$ / $\vec{x}_{20}$ / $\vec{x}_{21}$	(−1, 0)	(0, 1)	(−1, 1)	(−1, 1)	(−1, 1)

However, now we have one object $\vec{x}_{19}$ in the boundary region of clusters c_{s1} and c_{s2}. Its rough degree of dominance is given in Table 15.

7. Qualitative combination of rough clustering

The qualitative combination of rough clustering can be implemented similar to the crisp clustering. The process is made a little complicated by the fact that we need to calculate the lower and upper bounds of the combined clusters.

The lower bounds of combined clusters will be the intersections of the lower bounds of original clusters. Similarly, the upper bounds of combined clusters will be the intersections of the upper bounds of original clusters. We will use the combined cluster only if its upper bound is non-empty.

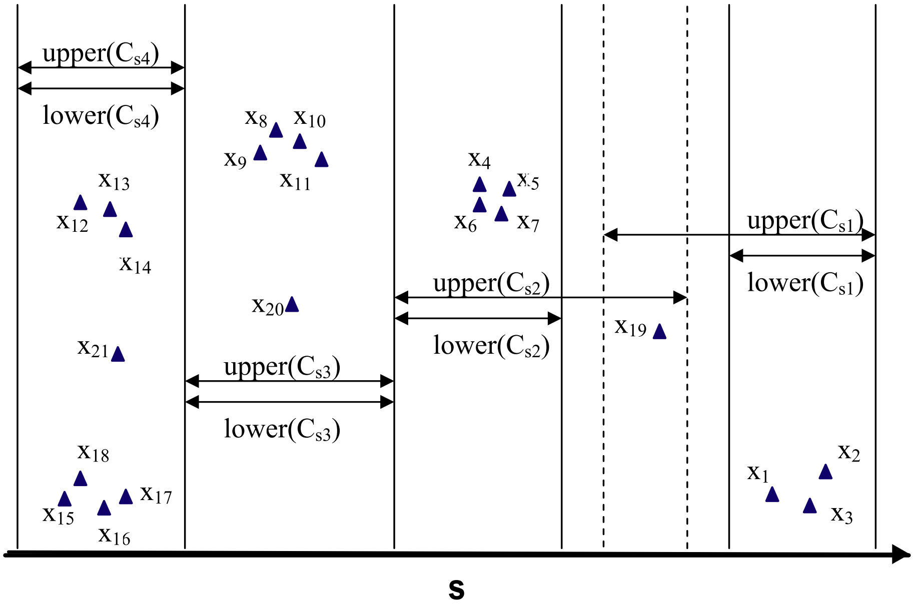

Fig. 6. The rough partition of objects by rs_s.

Table 14
$\vec{x}_i \succ_s \vec{x}_j$.

$\vec{c}_{si}$ \ $\vec{c}_{sj}$	$\vec{x}_i$ \ $\vec{x}_j$	$lower(c_{s1})$	$lower(c_{s2})$	$lower(c_{s3})$	$lower(c_{s4})$	$bnd(c_{s1})$ $bnd(c_{s2})$
		$\vec{x}_1, \vec{x}_2, \vec{x}_3$	$\vec{x}_4, \vec{x}_5$ $\vec{x}_6, \vec{x}_7$	$\vec{x}_8, \vec{x}_9$ $\vec{x}_{10}, \vec{x}_{11}$ $\vec{x}_{20}$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$ $\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}$ $\vec{x}_{18}, \vec{x}_{21}$	$\vec{x}_{19}$
$lower(c_{s1})$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	$\sim_s$	$\succ_s$	$\succ_s$	$\succ_s$	$\wp_s$
$lower(c_{s2})$	$\vec{x}_4, \vec{x}_5$ $\vec{x}_6, \vec{x}_7$	–	$\sim_s$	$\succ_s$	$\succ_s$	$\wp_s$
$lower(c_{s3})$	$\vec{x}_8, \vec{x}_9, \vec{x}_{10}$ $\vec{x}_{11}, \vec{x}_{20}$	–	–	$\sim_s$	$\succ_s$	–
$lower(c_{s4})$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$ $\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}$ $\vec{x}_{18}, \vec{x}_{21}$	–	–	–	$\sim_s$	–
$bnd(c_{s1})$ $bnd(c_{s2})$	$\vec{x}_{19}$	$\wp_s$	$\wp_s$	–	–	$\wp_s$

Table 15
$degree(\vec{x}_i \succ_s \vec{x}_j)$.

$\vec{c}_{si}$ \ $\vec{c}_{sj}$	$\vec{x}_i$ \ $\vec{x}_j$	$lower(c_{s1})$	$lower(c_{s2})$	$lower(c_{s3})$	$lower(c_{s4})$	$bnd(c_{s1})$ $bnd(c_{s2})$
		$\vec{x}_1, \vec{x}_2, \vec{x}_3$	$\vec{x}_4, \vec{x}_5$ $\vec{x}_6, \vec{x}_7$	$\vec{x}_8, \vec{x}_9$ $\vec{x}_{10}, \vec{x}_{11}$ $\vec{x}_{20}$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$ $\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}$ $\vec{x}_{18}, \vec{x}_{21}$	$\vec{x}_{19}$
$lower(c_{s1})$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	0	1	2	3	(0, 1)
$lower(c_{s2})$	$\vec{x}_4, \vec{x}_5$ $\vec{x}_6, \vec{x}_7$	−1	0	1	2	(−1, 0)
$lower(c_{s3})$	$\vec{x}_8, \vec{x}_9, \vec{x}_{10}$ $\vec{x}_{11}, \vec{x}_{20}$	−2	−1	0	1	(−2, −1)
$lower(c_{s4})$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$ $\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}$ $\vec{x}_{18}, \vec{x}_{21}$	−3	−2	−1	0	(−3, −2)
$bnd(c_{s1})$ $bnd(c_{s2})$	$\vec{x}_{19}$	(−1, 0)	(0, 1)	(1, 2)	(2, 3)	(−1, 1)

Let rs_v and rs_s be two rough clustering schemes.

$$rs_v = \{\vec{c}_{v1}, \ldots, \vec{c}_{vk_v}\},$$
$$rs_s = \{\vec{c}_{s1}, \ldots, \vec{c}_{sk_s}\}.$$

k_v and k_s are the number of clusters for rs_v and rs_s, respectively. Combination of clustering schemes rs_v and rs_s will be given as:

$$rs_{v\oplus s} = \left\{ \vec{c}_l \;\middle|\; \begin{array}{c} lower(c_l) = lower(c_{vi}) \cap lower(c_{sj}) \wedge \\ upper(c_l) = upper(c_{vi}) \cap upper(c_{sj}) \neq \emptyset, \\ \forall \vec{c}_{vi} \in rs_v \wedge \forall \vec{c}_{sj} \in rs_s \end{array} \right\}.$$

Let $k_{v\oplus s}$ be the number of clusters for $rs_{v\oplus s}$. Similar to the combination of crisp clusters, it can be easily verified that

$$max(k_v, k_s) \leqslant k_{v\oplus s} \leqslant k_v * k_s.$$

Combined indifference relations $\sim_{v\oplus s}$ will be defined as follows: If $\vec{x}_t, \vec{x}_r \in lower(c_i)$, then $\vec{x}_t \sim_{v\oplus s} \vec{x}_r$, where $\vec{c}_i \in rs_{v\oplus s}$.
Combined roughly similar relations $\backsim\!\!\backsim_{v\oplus s}$ will be defined as follows: If $\vec{x}_t, \vec{x}_r \in upper(c_i)$, then $\vec{x}_t \backsim\!\!\backsim_{v\oplus s} \vec{x}_r$, where $\vec{c}_i \in rs_{v\oplus s}$.
Before calculating dominance relations for combined clustering scheme, we will first define the concept of distance between clusters, which will be useful in distribution of objects in the boundary region. For the original clustering schemes, the distance between two clusters is given as absolute value of their degree of dominance, i.e.

$$distance_v(\vec{c}_{vi}, \vec{c}_{vj}) = abs\left(degree(\vec{c}_{vi} \succ_v \vec{c}_{vj})\right), \text{ and}$$
$$distance_s(\vec{c}_{si}, \vec{c}_{sj}) = abs\left(degree(\vec{c}_{si} \succ_s \vec{c}_{sj})\right).$$

It should be noted that since there is a strict ordering between clusters, the *distance* function obeys all the properties of Euclidean distance in a one-dimensional space.
Let us now calculate the distance, degrees of dominance, and dominance relations between combined clusters.
Let $\vec{c}_l \in rs_{v\oplus s}$ such that

$$lower(c_l) = lower(c_{vi}) \cap lower(c_{st}) \wedge upper(c_l) = upper(c_{vi}) \cap upper(c_{st}) \neq \emptyset,$$

and $\vec{c}_m \in rs_{v\oplus s}$ such that

$$lower(c_m) = lower(c_{vj}) \cap lower(c_{sr}) \wedge upper(c_m) = upper(c_{vj}) \cap upper(c_{sr}) \neq \emptyset.$$

The distance between combined clusters will be a simple addition of the two distances given by:

$$distance_{v\oplus s}(\vec{c}_l, \vec{c}_m) = distance_v(\vec{c}_{vi}, \vec{c}_{vj}) + distance_s(\vec{c}_{st}, \vec{c}_{sr}).$$

In order to calculate the dominance, we use the concept of weight similar to the one used for combination of crisp clusters. Assuming $\frac{\succ_v}{\succ_s} \geqslant 1$, we define the degree of dominance for the combined clustering scheme as:

$$degree(\vec{c}_l \succ_{v\oplus s} \vec{c}_m) = \frac{\succ_v}{\succ_s} * degree(\vec{c}_{vi} \succ_v \vec{c}_{vj}) + degree(\vec{c}_{st} \succ_s \vec{c}_{sr}).$$

Let $\vec{c}_i, \vec{c}_j \in c_{v\oplus s}$. The degree defined above will then be used to define the strict preference relation between $\vec{c}_i$ and $\vec{c}_j$ as:

If $degree(\vec{c}_i \succ_{v\oplus s} \vec{c}_j) > 0$ then $\vec{c}_i \succ_{v\oplus s} \vec{c}_j$
else if $degree(\vec{c}_i \succ_{v\oplus s} \vec{c}_j) < 0$ then $\vec{c}_j \succ_{v\oplus s} \vec{c}_i$
else if $\neg(\vec{c}_i \sim_{v\oplus s} \vec{c}_j)$ and $degree(\vec{c}_i \succ_{v\oplus s} \vec{c}_j) = 0$ there is no relation between $\vec{c}_i$ and $\vec{c}_j$, i.e. $\vec{c}_i ?_{v\oplus s} \vec{c}_j$.

Relationship between individual pair of objects in lower bounds of two different combined clusters can be inferred by the transitivity axioms for strict preference and indifference relations as well as axiom 7 that defines the transitivity between $\sim$ and $\succ$.
It should be noted here that the qualitative rough combination may lead to proliferation of combined clusters with empty lower bounds. We should eliminate such clusters. However, if some of the objects in the boundary region of these clusters belong to none or only one of the remaining clusters, we may have to make it part of other neighboring clusters. This assignment will preserve property (P3) of rough clustering. Neighboring clusters are defined using the distance between the clusters. The specifics of eliminating the clusters are as follows.

- Let $\vec{x}_b$ be an object from an eliminated cluster.
- If $\vec{x}_b$ belongs to upper bound of two or more of the remaining clusters, there is no need to assign it to additional clusters.

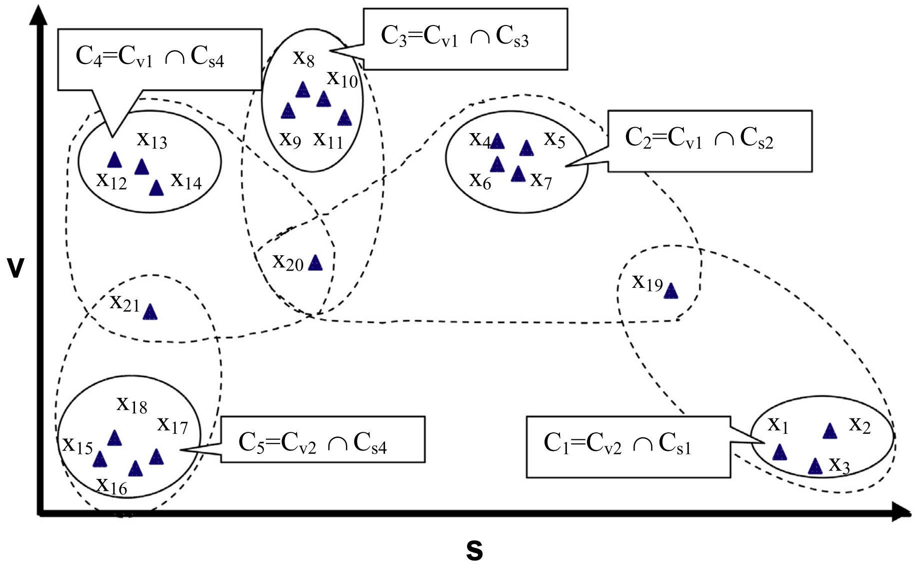

Fig. 7. The distribution of data objects for rough clustering.

- If $\vec{x}_b$ belongs to upper bound of only one of the remaining clusters, $\vec{c}_i$, then we assign $\vec{x}_b$ to the upper bounds of all the clusters that are neighbors of $\vec{c}_i$. The neighbors of $\vec{c}_i$ are calculated as follows.
  - Let

$$d = min\{distance(\vec{c}_j, \vec{c}_i) \mid \vec{c}_j \in rs_{v \oplus s} \wedge lower(c_j) \neq \emptyset\}.$$

  - $Neighbours(\vec{c}_i) = \{\vec{c}_j \mid distance(\vec{c}_j, \vec{c}_i) = d\}$
- If $\vec{x}_b$ belongs to upper bound of none of the remaining clusters, then it should be made part of upper bound of all the remaining clusters.

Once we have eliminated the clusters with empty lower bounds, the final step is to calculate the rough degree of dominance. If one or both of the objects $\vec{x}_a$ and $\vec{x}_b$ are in a boundary region, we need to calculate the rough degrees of dominance between them using the formula described in the previous section. The dominance relationships between $\vec{x}_a$ and $\vec{x}_b$ can then be calculated again using the procedure described in the previous section.

In summary, the steps in the qualitative combination of rough clustering are as follows.

Step 1: Calculate lower and upper bounds of combined clusters using intersections of lower and upper bounds of original clusters.

Step 2: Eliminate the clusters with empty lower bounds.

Step 3: Redistribute objects from boundary regions of eliminated clusters, if necessary.

Step 4: Calculate the degrees and rough degrees of dominance using the weight.

Step 5: Determine dominance relations.

Example 4. We use the same dataset presented in Example 3 to illustrate qualitative combination of rough clustering schemes. We want to group the customers based on a combination of loyalty and spending.

It is obvious that the combined clustering scheme will have five clusters with a few boundary region objects. The lower bounds and boundary region of each cluster obtained from qualitative combination are shown in Table 16. It can be seen that three of the eight clusters formed have empty lower bounds. In order to consolidate these eight clusters into five clusters, we first find the distances between the combined clusters.

Table 17 presents the distances between the combined clustering schemes. We can eliminate clusters $\vec{c}_6$ and $\vec{c}_7$ without any redistribution of objects, since $\vec{x}_{19}$ already belongs to two boundary regions. Elimination of $\vec{c}_8$, however, makes it necessary to redistribute $\vec{x}_{20}$, since it now only belongs to upper bound of $\vec{c}_3$. From Table 17, two closest neighbors of $\vec{c}_3$ are clusters $\vec{c}_2$ and $\vec{c}_4$. Therefore, we add $\vec{x}_{20}$ to the upper bounds of $\vec{c}_2$ and $\vec{c}_4$. Table 18 shows the final five clusters $C = \{\vec{c}_1, \ldots, \vec{c}_5\}$.

We can now calculate the degrees of dominance for different weights. Table 19, Table 21, and Table 23 describe the combined degrees of dominance for $\frac{\succ_v}{\succ_s} = 1$, $\frac{\succ_v}{\succ_s} = 3$ and $\frac{\succ_s}{\succ_v} = 3$, respectively. Table 20, Table 22 and Table 24 show the corresponding dominance relationships.

Table 16
Clusters in rough clustering before consolidation.

$\vec{c}_i$	$\vec{c}_{vi} \cap \vec{c}_{sj}$	$lower(c_i)$	$bnd(c_i)$
$\vec{c}_1$	$\vec{c}_{v2} \cap \vec{c}_{s1}$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	$\vec{x}_{19}$
$\vec{c}_2$	$\vec{c}_{v1} \cap \vec{c}_{s2}$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$	$\vec{x}_{19}$
$\vec{c}_3$	$\vec{c}_{v1} \cap \vec{c}_{s3}$	$\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$	$\vec{x}_{20}$
$\vec{c}_4$	$\vec{c}_{v1} \cap \vec{c}_{s4}$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	$\vec{x}_{21}$
$\vec{c}_5$	$\vec{c}_{v2} \cap \vec{c}_{s4}$	$\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$	$\vec{x}_{21}$
$\vec{c}_6$	$\vec{c}_{v1} \cap \vec{c}_{s1}$	$\emptyset$	$\vec{x}_{19}$
$\vec{c}_7$	$\vec{c}_{v2} \cap \vec{c}_{s2}$	$\emptyset$	$\vec{x}_{19}$
$\vec{c}_8$	$\vec{c}_{v2} \cap \vec{c}_{s3}$	$\emptyset$	$\vec{x}_{20}$

Table 17
$distance_{v\oplus s}(\vec{c}_i, \vec{c}_j)$ before consolidation.

$\vec{c}_i$ \ $\vec{c}_j$	$\vec{c}_1$	$\vec{c}_2$	$\vec{c}_3$	$\vec{c}_4$	$\vec{c}_5$
$\vec{c}_1$	0	2	3	4	3
$\vec{c}_2$	2	0	1	2	3
$\vec{c}_3$	3	1	0	1	2
$\vec{c}_4$	4	2	1	0	1
$\vec{c}_5$	3	3	2	1	0

Table 18
Clusters in rough clustering after consolidation.

$\vec{c}_i$	$\vec{c}_{vi} \cap \vec{c}_{sj}$	$lower(c_i)$	$bnd(c_i)$
$\vec{c}_1$	$\vec{c}_{v2} \cap \vec{c}_{s1}$	$\vec{x}_1, \vec{x}_2, \vec{x}_3$	$\vec{x}_{19}$
$\vec{c}_2$	$\vec{c}_{v1} \cap \vec{c}_{s2}$	$\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7$	$\vec{x}_{19}, \vec{x}_{20}$
$\vec{c}_3$	$\vec{c}_{v1} \cap \vec{c}_{s3}$	$\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}$	$\vec{x}_{20}$
$\vec{c}_4$	$\vec{c}_{v1} \cap \vec{c}_{s4}$	$\vec{x}_{12}, \vec{x}_{13}, \vec{x}_{14}$	$\vec{x}_{20}, \vec{x}_{21}$
$\vec{c}_5$	$\vec{c}_{v2} \cap \vec{c}_{s4}$	$\vec{x}_{15}, \vec{x}_{16}, \vec{x}_{17}, \vec{x}_{18}$	$\vec{x}_{21}$

Table 19
$degree(\vec{c}_i \succ_{v\oplus s} \vec{c}_j)$ when $\frac{\succ_v}{\succ_s} = 1$.

$\vec{c}_i$ \ $\vec{c}_j$	$\vec{c}_1$	$\vec{c}_2$	$\vec{c}_3$	$\vec{c}_4$	$\vec{c}_5$
$\vec{c}_1$	0	0	1	2	3
$\vec{c}_2$	0	0	1	2	3
$\vec{c}_3$	−1	−1	0	1	2
$\vec{c}_4$	−2	−2	−1	0	1
$\vec{c}_5$	−3	−3	−2	−1	0

Table 20
$\vec{c}_i \succ_{v\oplus s} \vec{c}_j$ when $\frac{\succ_v}{\succ_s} = 1$.

$\vec{c}_i$ \ $\vec{c}_j$	$\vec{c}_1$	$\vec{c}_2$	$\vec{c}_3$	$\vec{c}_4$	$\vec{c}_5$
$\vec{c}_1$	$\sim_{v\oplus s}$	$?_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_2$	$?_{v\oplus s}$	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_3$	–	–	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_4$	–	–	–	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_5$	–	–	–	–	$\sim_{v\oplus s}$

Table 21
$degree(\vec{c}_i \succ_{v\oplus s} \vec{c}_j)$ when $\frac{\succ_v}{\succ_s} = 3$.

$\vec{c}_i$ \ $\vec{c}_j$	$\vec{c}_1$	$\vec{c}_2$	$\vec{c}_3$	$\vec{c}_4$	$\vec{c}_5$
$\vec{c}_1$	0	−2	−1	0	3
$\vec{c}_2$	2	0	1	2	5
$\vec{c}_3$	1	−1	0	1	4
$\vec{c}_4$	0	−2	−1	0	3
$\vec{c}_5$	−3	−5	−4	−3	0

Table 22
$\vec{c}_i \succ_{v\oplus s} \vec{c}_j$ when $\frac{\succ_v}{\succ_s} = 3$.

$\vec{c}_i$ \ $\vec{c}_j$	$\vec{c}_1$	$\vec{c}_2$	$\vec{c}_3$	$\vec{c}_4$	$\vec{c}_5$
$\vec{c}_1$	$\sim_{v\oplus s}$	–	–	$?_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_2$	$\succ_{v\oplus s}$	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_3$	$\succ_{v\oplus s}$	–	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_4$	$?_{v\oplus s}$	–	–	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_5$	–	–	–	–	$\sim_{v\oplus s}$

Table 23
$degree(\vec{c}_i \succ_{v\oplus s} \vec{c}_j)$ when $\frac{\succ_s}{\succ_v} = 3$.

$\vec{c}_i$ \ $\vec{c}_j$	$\vec{c}_1$	$\vec{c}_2$	$\vec{c}_3$	$\vec{c}_4$	$\vec{c}_5$
$\vec{c}_1$	0	2	5	8	9
$\vec{c}_2$	−2	0	3	6	7
$\vec{c}_3$	−5	−3	0	3	4
$\vec{c}_4$	−8	−6	−3	0	1
$\vec{c}_5$	−9	−7	−4	−1	0

Table 24
$\vec{c}_i \succ_{v\oplus s} \vec{c}_j$ when $\frac{\succ_s}{\succ_v} = 3$.

$\vec{c}_i$ \ $\vec{c}_j$	$\vec{c}_1$	$\vec{c}_2$	$\vec{c}_3$	$\vec{c}_4$	$\vec{c}_5$
$\vec{c}_1$	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_2$	–	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_3$	–	–	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_4$	–	–	–	$\sim_{v\oplus s}$	$\succ_{v\oplus s}$
$\vec{c}_5$	–	–	–	–	$\sim_{v\oplus s}$

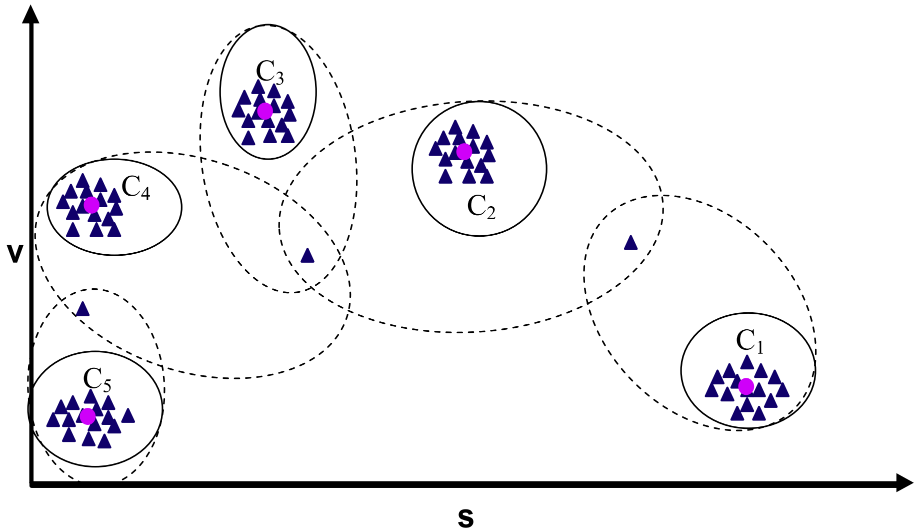

Fig. 8. Synthetic data for combining rough clustering.

8. Qualitative and quantitative combination of rough clustering

The previous section discussed qualitative combination of two rough clustering schemes. In this section, we compare the quantitative alternative used for rough clustering based on multicriteria, and study the resulting set of dominance relations. Similar to crisp clustering, the quantitative combination of rough clustering scheme multiplies the numeric values of the two criteria with different weights depending on their relative importance. The weighted numeric values are then clustered using rough K-means.

We used a synthetic dataset that is similar to the small dataset presented in Examples 3 and 4 to analyze the quantitative combination of rough clustering. The synthetic dataset is shown in Fig. 8. It is clear that there are five clusters in this dataset with a few objects in the boundary region.

For rough clustering, threshold and ω_l are set as 1.4 and 0.7, respectively. The number of clusters is 5. The combined clustering is shown in Fig. 8. Lower bounds are outlined by solid lines and boundary area are outlined by dotted lines.

The quantitative combination was done by varying the pair of weights for visits and spending from $(0, 1)$ to $(1, 0)$ with an increment or decrement of 0.1. Similar to the qualitative combination, for each pair of weights we got five clusters as shown in Fig. 8. For extreme values of the weights, the objects $\vec{x}_{19}, \vec{x}_{20}, \vec{x}_{21}$ sometimes moved to lower bounds of one of the clusters. The combined clusters were ranked based on the value of their centroid vectors.

Table 25 shows a correspondence between qualitative and quantitative rough combinations. As we saw for crisp clustering, the qualitative combinations allow us to specify no relation between clusters, while we will usually have a strict dominance relation between quantitatively combined clusters.

Table 25
Dominance relations between clusters for rough clustering with synthetic data.

ω_v	ω_s	Dominance relations	$\frac{\succ_v}{\succ_s}$	Dominance relations
0.9	0.1	$\vec{c}_3 \succ \vec{c}_2 \succ \vec{c}_4 \succ \vec{c}_1 \succ \vec{c}_5$	9	$\vec{c}_2 \succ \vec{c}_3 \succ \vec{c}_4 \succ \vec{c}_1 \succ \vec{c}_5$
0.8	0.2		4	
0.75	0.25		3	$\vec{c}_2 \succ \vec{c}_3 \succ \vec{c}_1 \succ \vec{c}_5$ $\vec{c}_2 \succ \vec{c}_3 \succ \vec{c}_4 \succ \vec{c}_5$ $\vec{c}_1 ? \vec{c}_4$
0.7	0.3		$\frac{7}{3}$	$\vec{c}_2 \succ \vec{c}_3 \succ \vec{c}_1 \succ \vec{c}_4 \succ \vec{c}_5$
0.6	0.4	$\vec{c}_3 \succ \vec{c}_2 \succ \vec{c}_1 \succ \vec{c}_4 \succ \vec{c}_5$	$\frac{3}{2}$	$\vec{c}_2 \succ \vec{c}_1 \succ \vec{c}_3 \succ \vec{c}_4 \succ \vec{c}_5$
0.5	0.5	$\vec{c}_1 \succ \vec{c}_2 \succ \vec{c}_3 \succ \vec{c}_4 \succ \vec{c}_5$	1	$\vec{c}_1 \succ \vec{c}_3 \succ \vec{c}_4 \succ \vec{c}_5$ $\vec{c}_2 \succ \vec{c}_3 \succ \vec{c}_4 \succ \vec{c}_5$ $\vec{c}_1 ? \vec{c}_2$
0.4	0.6		$\frac{2}{3}$	$\vec{c}_1 \succ \vec{c}_2 \succ \vec{c}_3 \succ \vec{c}_4 \succ \vec{c}_5$
0.3	0.7		$\frac{3}{7}$	
0.25	0.75		$\frac{1}{3}$	
0.2	0.8		$\frac{1}{4}$	
0.1	0.9		$\frac{1}{9}$	

9. Summary and conclusions

This paper described a preference/dominance relation based framework for analyzing crisp and rough clustering. Fishburn proposed the use of preference relations for qualitative decision theory. Use of these relations in clustering analysis will help us introduce semantic structure to the unsupervised clustering process. The proposed qualitative framework is further used to propose crisp and rough combination rules. The framework also helps us better understand the quantitative combination of clustering schemes. The paper studied the correspondence between qualitative and quantitative combinations. The qualitative combinations allow us to specify no relation between clusters, while usually there is a strict preference relation between quantitatively combined clusters. Sometimes such a partial dominance relation may be more suitable. The qualitative combination may be useful in cases where one can only communicate individual clustering schemes due to privacy considerations in sharing attribute values or due to large communication costs. It can also be useful for simply devising an appropriate weighting scheme for quantitative combination scheme. The proposed framework and combination rules are explained using analysis of customers from a synthetic retail dataset based on their spending and visit patterns. While the qualitative combination described in this paper in this paper is based on axiomatic properties of preference relations, it is mainly focused on providing a practical combination rule. There are many theoretical aspects of the combination of multiple clustering schemes that need to be explored as part of our future research.

Acknowledgements

The authors would like to thank China Scholarship Council and NSERC Canada for their financial support.

References

[1] P.C. Fishburn, Utility Theory for Decision Making, Clearinghouse, Springfield, VA, USA, 1970.
[2] Y.Y. Yao, Generalized rough set models, Rough Sets in Knowledge Discovery (1998) 286–318.
[3] Y.Y. Yao, Interval sets and interval sets algebras, in: Proc. of the 2009 IEEE International Conference on Cognitive Informatics, 2009, pp. 307–314.
[4] Y.Y. Yao, An outline of a theory of three-way decisions, in: J. Yao, Y. Yang, R. Slowinski, S. Greco, H. Li, S. Mitra, L. Polkowski (Eds.), Proc. of the 2012 Rough Sets and Current Trends in Computing, 2012, pp. 1–17.
[5] Y.Y. Yao, Decision-theoretic rough set models, in: Proc. of 2nd International Conference on Rough Set and Knowledge Technology, 2007, pp. 1–12.
[6] Y.Y. Yao, S. Wong, P. Lingras, A decision-theoretic rough set model, in: Proc. of ISMIS 1990, 1990.
[7] Y.Y. Yao, S. Wong, A decision theoretic framework for approximating concepts, International Journal of Man-Machine Studies 37 (6) (1992) 793–809.
[8] Y. Sai, Y.Y. Yao, N. Zhong, Data analysis and mining in ordered information tables, in: Proc. of the 2001 IEEE International Conference on Data Mining, 2001, pp. 497–504.
[9] Y.Y. Yao, P. Lingras, R. Wang, D. Miao, Interval set cluster analysis, in: Proc. of 2009 Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, RSFDGrC'09, 2009, pp. 398–405.
[10] D. Liu, T. Li, H. Li, A multiple-category classification approach with decision-theoretic rough sets, Fundamenta Informaticae 115 (2012) 173–188.
[11] X. Yang, J.T. Yao, Modelling multi-agent three-way decisions with decision-theoretic rough sets, Fundamenta Informaticae 115 (2012) 157–171.
[12] Y.Y. Yao, Y. Zhao, Attribute reduction in decision-theoretic rough set models, Information Sciences 178 (17) (2008) 3356–3373.
[13] C.-C. Chan, G.-H. Tzeng, Computing approximations of dominance-based rough sets by bit-vector encodings, in: Proc. of the Sixth International Conference on Rough Sets and Current Trends in Computing, 2008, pp. 131–141.
[14] S. Greco, B. Matarazzo, R. Slowinski, Rough approximation of a preference relation by dominance relations, European Journal of Operation Research 117 (1) (1999) 63–83.
[15] S. Greco, B. Matarazzo, R. Slowinski, Rough sets theory for multicriteria decision analysis, European Journal of Operation Research 130 (3) (2001) 449–467.
[16] S. Greco, B. Matarazzo, R. Slowinski, Rough sets methodology for sorting problems in presence of multiple attributes and criteria, European Journal of Operation Research 138 (2) (2002) 247–259.
[17] W. Pedrycz, Fuzzy sets and neurocomputation: knowledge-based networks (invited paper), in: Proc. of SPIE Applications of Fuzzy Logic Technology II, 1995, pp. 2–20.
[18] J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, Cybernetics 3 (3) (1973) 32–57. http://dx.doi.org/10.1080/01969727308546046.
[19] J.C. Dunn, Well separated clusters and optimal fuzzy partitions, Cybernetics 4 (1974) 95–104.

[20] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Kluwer Academic Publishers, Norwell, MA, USA, 1981.
[21] W. Pedrycz, J. Waletzky, Fuzzy clustering with partial supervision, IEEE Transactions on Systems, Man, and Cybernetics, Part B 27 (5) (1997) 787–795.
[22] T. Denoeux, M.H. Masson, Evclus: evidential clustering of proximity data, IEEE Transactions on Systems Man and Cybernetics 34 (1) (2004) 95–109.
[23] N.R. Pal, K. Pal, J.M. Keller, J. Bezdek, A possibilistic fuzzy c-means clustering algorithm, IEEE Transaction on Fuzzy Systems 13 (4) (2005) 517–530.
[24] Y. Xie, V.V. Raghavan, P. Dhatric, X.Q. Zhao, A new fuzzy clustering algorithm for optimally finding granular prototypes, International Journal of Approximate Reasoning 40 (2005) 109–124.
[25] M.H. Masson, T. Denoeux, ECM: an evidential version of the fuzzy c-means algorithm, Pattern Recognition 41 (2008) 1384–1397.
[26] M.H. Masson, T. Denoeux, Ensemble clustering in the belief functions framework, International Journal of Approximate Reasoning 52 (1) (2011) 92–109.
[27] G. Peters, F. Crespo, P. Lingras, R. Weber, Soft clustering – fuzzy and rough approaches and their extensions and derivatives, International Journal of Approximate Reasoning 54 (2) (2013) 307–322.
[28] M. Banerjee, S. Mitra, S.K. Pal, Rough fuzzy MLP: knowledge encoding and classification, IEEE Transactions on Neural Networks 9 (6) (1998) 1203–1216.
[29] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers., Boston, 1992.
[30] S. Hirano, T. Okuzaki, Y. Hata, S. Tsumoto, K. Tsumoto, A clustering method based on rough sets and its application to biochemical data analysis, in: Proc. of Atlantic Symp. Computational Biology and Genome Inf. Syst. and Technol., 2001, pp. 16–20.
[31] T.B. Ho, N.B. Nguyen, Nonhierarchical document clustering by a tolerance rough set model, International Journal of Intelligent Systems 17 (2) (2002) 199–212.
[32] S. Mitra, An evolutionary rough partitive clustering, Pattern Recognition Letters 25 (12) (2004) 1439–1449.
[33] S. Mitra, H. Banka, W. Pedrycz, Rough-fuzzy collaborative clustering, IEEE Transactions on Systems, Man, and Cybernetics. Part B: Cybernetics 36 (4) (2006) 795–805.
[34] H.S. Nguyen, T.B. Ho, Rough document clustering and the internet, Handbook on Granular Computing, John Wiley & Sons, Hoboken, NJ, 2007, pp. 987–1005.
[35] G. Peters, Some refinements of rough k-means, Pattern Recognition 39 (8) (2006) 1481–1491.
[36] J.F. Peters, A. Skowron, Z. Suraj, W. Rzasa, M. Borkowski, Clustering: a rough set approach to constructing information granules, in: Proc. of the 6th International Conference on Soft Computing and Distributed Processing, 2002, pp. 57–61.
[37] P. Lingras, Unsupervised rough set classification using gas, Journal of Intelligent Information Systems 16 (3) (2001) 215–228.
[38] P. Lingras, C. West, Interval set clustering of web users with rough k-means, Journal of Intelligent Information System 23 (1) (2004) 5–16.
[39] P. Lingras, M. Hogo, M. Snorek, Interval set clustering of web users using modified Kohonen self-organizing maps based on the properties of rough sets, Web Intelligence and Agent Systems 2 (3) (2004) 217–230.
[40] P. Lingras, M. Chen, D. Miao, Precision of rough set clustering, in: C.-C. Chan, J. Grzymala-Busse, W. Ziarko (Eds.), Rough Sets and Current Trends in Computing, Lecture Notes in Computer Science, vol. 5306, Springer, Berlin/Heidelberg, 2008, pp. 369–378. http://dx.doi.org/10.1007/978-3-540-88425-5_38.
[41] P. Lingras, M. Chen, D.Q. Miao, Rough multi-category decision theoretic framework, in: Proc of 2008 Rough Sets and Knowledge Technology, 2008, pp. 676–683.
[42] P. Lingras, M. Chen, D.Q. Miao, Rough cluster quality index based on decision theory, IEEE Transactions on Knowledge and Data Engineering 21 (7) (2009) 1014–1026.
[43] H. Yu, S. Chu, Autonomous knowledge-oriented clustering using decision-theoretic rough set theory, Fundamenta Informaticae 115 (2012) 141–156.
[44] H. Yu, Y. Wang, Three-way decisions method for overlapping clustering, in: Proc. of the 2012 Rough Sets and Current Trends in Computing, 2012, pp. 277–286.
[45] J.A. Hartigan, M.A. Wong, Algorithm as136: a k-means clustering algorithm, Applied Statistics 28 (1979) 100–108.
[46] J. MacQueen, Some methods fir classification and analysis of multivariate observations, in: Proc. of Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967, pp. 281–297.
[47] L. Polkowski, A. Skowron, Rough mereological approach to knowledge-based distributed AI, in: Proc. of the Third World Congress on Expert Systems, Cognizant Communication Corporation, 1996, pp. 774–781.
[48] X. Jia, Z. Tang, W. Liao, L. Shang, On an optimization representation of decision-theoretic rough set model, International Journal of Approximate Reasoning. http://dx.doi.org/10.1016/j.ijar.2013.02.010.
[49] M.-W. Shao, Y. Leung, W.-Z. Wu, Rule acquisition and complexity reduction in formal decision contexts, International Journal of Approximate Reasoning, 2013, in press.
[50] T.-J. Li, X.-P. Yang, An axiomatic characterization of probabilistic rough sets, International Journal of Approximate Reasoning. http://dx.doi.org/10.1016/j.ijar.2013.02.012.
[51] D. Liu, T. Li, D. Liang, Incorporating logistic regression to decision-theoretic rough sets for classifications, International Journal of Approximate Reasoning. http://dx.doi.org/10.1016/j.ijar.2013.02.013.
[52] Y. Qian, H. Zhang, Y. Sang, J. Liang, Multigranulation decision-theoretic rough sets, International Journal of Approximate Reasoning. http://dx.doi.org/10.1016/j.ijar.2013.03.004.
[53] N. Azam, J.T. Yao, Analyzing uncertainties of probabilistic rough set regions with game-theoretic rough sets, International Journal of Approximate Reasoning. http://dx.doi.org/10.1016/j.ijar.2013.03.015.
[54] F. Min, Q. Hu, W. Zhu, Feature selection with test cost constraint, International Journal of Approximate Reasoning. http://dx.doi.org/10.1016/j.ijar.2013.04.003.
[55] J.W. Grzymala-Busse, P.G. Clark, M. Kuehnhausen, Generalized probabilistic approximations of incomplete data, International Journal of Approximate Reasoning. http://dx.doi.org/10.1016/j.ijar.2013.04.007.
[56] B. Zhou, Multi-class decision-theoretic rough sets, International Journal of Approximate Reasoning. http://dx.doi.org/10.1016/j.ijar.2013.04.006.
[57] F. Li, M. Ye, X. Chen, An extension to rough c-means clustering based on decision-theoretic rough set model, International Journal of Approximate Reasoning.
[58] H. Yu, Z. Liu, G. Wang, Automatically determining the number of clusters using decision-theoretic rough set, in: Proc. of the 2011 Rough Sets and Knowledge Technology, 2011, pp. 504–513.
[59] H. Yu, Z. Liu, G. Wang, An automatic method to determine the number of clusters using decision-theoretic rough set, International Journal of Approximate Reasoning. http://dx.doi.org/10.1016/j.ijar.2013.03.018.
[60] P. Lingras, Applications of rough patterns, in: L. Polkowski, A. Skowron (Eds.), Rough Sets in Data Mining and Knowledge Discovery, Soft Computing, Physica Verlag (Springer), Berlin, 1998, pp. 369–384 (Chapter 3).