Evidential Clustering: Review and some New Developments

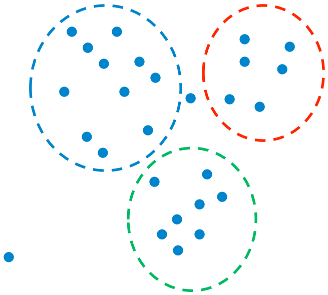
Thierry Denœux

Université de Technologie de Compiègne
Heudiasyc (UMR CNRS 7253)

https://www.hds.utc.fr/~tdenoeux

18th Int. Symp. on Advanced Intelligent Systems (ISIS 2017)
Daegu, South Korea
October 12, 2017

Cluster Analysis



- n objects described by
  - Attribute vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ (attribute data) or
  - Dissimilarities (proximity data)
- Goals:
  1. Discover groups in the data
  2. Assess the uncertainty in group membership

Hard and soft clustering

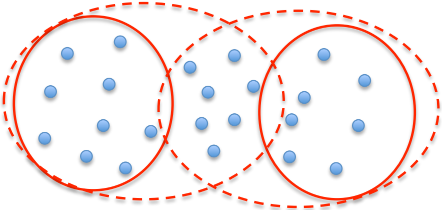
Hard clustering: no representation of uncertainty. Each object is assigned to one and only one group. Group membership is represented by binary variables u_{ik} such that $u_{ik} = 1$ if object i belongs to group k and $u_{ik} = 0$ otherwise.

Fuzzy clustering: each object has a degree of membership $u_{ik} \in [0, 1]$ to each group, with $\sum_{k=1}^{c} u_{ik} = 1$. The u_{ik}'s can be interpreted as probabilities.

Possibilistic clustering: the u_{ik} are free to take any value in $[0, 1]^c$. Each number u_{ik} is interpreted as a degree of possibility that object i belongs to group k.

Hard and soft clustering (continued)

Rough clustering: each cluster ω_k is characterized by a lower approximation $\underline{\omega}_k$ and an upper approximation $\overline{\omega}_k$, with $\underline{\omega}_k \subseteq \overline{\omega}_k$; the membership of object i to cluster k is described by a pair $(\underline{u}_{ik}, \overline{u}_{ik}) \in \{0,1\}^2$, with $\underline{u}_{ik} \leq \overline{u}_{ik}$, $\sum_{k=1}^{c} \underline{u}_{ik} \leq 1$ and $\sum_{k=1}^{c} \overline{u}_{ik} \geq 1$.



Clustering and belief functions

clustering structure	uncertainty framework
fuzzy partition	probability theory
possibilistic partition	possibility theory
rough partition	(rough) sets
?	belief functions

- As belief functions extend probabilities, possibilities and sets, could the theory of belief functions provide a more general and flexible framework for cluster analysis?
- Objectives:
  - Unify the various approaches to clustering
  - Achieve a richer and more accurate representation of uncertainty
  - New clustering algorithms and new tools to compare and combine clustering structures.

Outline

1. Theory of belief functions: a brief introduction
2. Evidential clustering
   - Credal partition
   - Summarization of a credal partition
   - Relational representation of a credal partition
3. Algorithms
   - EVCLUS
   - Handling a large number of clusters
   - Constrained evidential clustering
4. Comparing and combining the results of soft clustering algorithms
   - Credal Rand index
   - Combining clustering structures

Outline

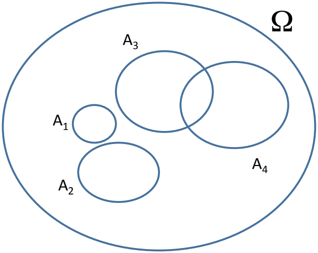
1. Theory of belief functions: a brief introduction
2. Evidential clustering
   - Credal partition
   - Summarization of a credal partition
   - Relational representation of a credal partition
3. Algorithms
   - EVCLUS
   - Handling a large number of clusters
   - Constrained evidential clustering
4. Comparing and combining the results of soft clustering algorithms
   - Credal Rand index
   - Combining clustering structures

Mass function

Definition

- Let Q be a question of interest, and Ω be a finite set of possible answers, one and only one of which is true.
- Evidence about Q can be represented by a mass function m from 2^Ω to $[0,1]$ such that

$$\sum_{A \subseteq \Omega} m(A) = 1$$



- The subsets A of Ω such that $m(A) \neq 0$ are called the focal sets of m.
- If $m(\emptyset) = 0$, m is said to be normalized (usually assumed).

Example

- Consider a road scene analysis application. Let Q concern the type of object in some region of the image, and $\Omega = \{G, R, T, O, S\}$, corresponding to the possibilities **G**rass, **R**oad, **T**ree/Bush, **O**bstacle, **S**ky
- Assume that a lidar sensor (laser telemeter) returns the information that the object is either a Tree or an Obstacle, but we there is a probability $p = 0.1$ that the information is not reliable (because, e.g., the sensor is out of order).
- This uncertain evidence can be represented by the following mass function m on Ω,

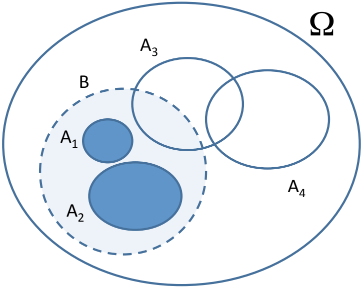
$$m(\{T, O\}) = 0.9, \quad m(\Omega) = 0.1$$

Mass function

Interpretation

- The meaning of the evidence is uncertain: it points to several subsets of Ω, with different degrees of support.
- Each mass $m(A)$ is a measure of the support given exactly to A, and to no more specific subset.
- In particular, $m(\Omega)$ is a measure of lack of information (ignorance).
- The vacuous mass function defined by $m_?(\Omega) = 1$ represents total lack of information (complete ignorance).

Belief and plausibility functions



Given a normalized mass function m, we can define

- Tshe degree of belief in B as

$$Bel(B) = \sum_{A \subseteq B} m(A)$$

- The degree of plausibility of B as

$$Pl(B) = 1 - Bel(\overline{B}) = \sum_{A \cap B \neq \emptyset} m(A).$$

Interpretation: $Bel(B)$ measures the total support given to B, while $Pl(A)$ measures the lack of support given to $\overline{B}$.

Two-dimensional representation

- The uncertainty on a proposition B is represented by two numbers: $Bel(B)$ and $Pl(B)$, with $Bel(B) \leq Pl(B)$.
- The intervals $[Bel(B), Pl(B)]$ have maximum length when m is the vacuous mass function. Then,

$$[Bel(B), Pl(B)] = [0, 1]$$

  for all subset B of Ω, except $\emptyset$ and Ω.
- The intervals $[Bel(B), Pl(B)]$ are reduced to points when the focal sets of m are singletons (m is then said to be Bayesian); then,

$$Bel(B) = Pl(B)$$

  for all B, and Bel is a probability measure.

Logical/Consonant mass function

- If m has only one focal set, it is said to be logical.
- If the focal sets of m are nested ($A_1 \subset A_2 \subset \ldots \subset A_n$), m is said to be consonant.
- Pl is then a possibility measure, i.e.,

$$Pl(A \cup B) = \max(Pl(A), Pl(B))$$

  for all $A, B \subseteq \Omega$.
- We have

$$Pl(A) = \max_{\omega \in A} Pl(\{\omega\}) \text{ for all } A \subseteq \Omega.$$

Dempster's rule

- Let m_1 and m_2 be two mass functions induced by independent pieces of evidence.
- Their orthogonal sum is the mass function $m_1 \oplus m_2$ defined by

$$(m_1 \oplus m_2)(A) = \frac{1}{1-\kappa} \sum_{B \cap C = A} m_1(B) m_2(C)$$

for all $A \neq \emptyset$ and $(m_1 \oplus m_2)(\emptyset) = 0$, where

$$\kappa = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$$

is the degree of conflict.

- $\oplus$ is commutative, associative, and $m_?$ is its single neutral element.

Belief-probability transformation

- It may be useful to transform a mass function m into a probability distribution for approximation or decision-making.
- Plausibility-probability transformation

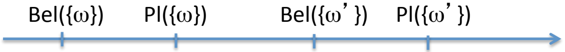
$$p_m(\omega) = \frac{Pl(\{\omega\})}{\sum_{\omega \in \Omega} Pl(\{\omega\})}$$

- Property:

$$p_{m_1 \oplus m_2} = p_{m_1} \oplus p_{m_2}$$

Decision

- Given a normalized mass function m, how to select an element or a subset of Ω?
- Several solutions: for instance, choose ω with the largest degree of belief $Bel(\{\omega\})$ or the largest plausibility $Pl(\{\omega\})$.
- Interval dominance:



  - ω is dominated by ω' iff $Pl(\{\omega\}) < Bel(\{\omega'\})$
  - We may select the set A^* of possibilities ω that are dominated by no other possibility ω'.

Main ideas

- The theory of belief function combines sets with probabilities, by assigning basic probabilities to focal sets.
- A mass function can be seen as a generalized set or as a generalized probability distribution.
- Possibility theory is also recovered as a special case, when the focal sets are nested.
- A mass function can be transformed into
  - a probability distribution,
  - an element of Ω, or
  - a subset of Ω.

Outline

1. Theory of belief functions: a brief introduction
2. Evidential clustering
   - Credal partition
   - Summarization of a credal partition
   - Relational representation of a credal partition
3. Algorithms
   - EVCLUS
   - Handling a large number of clusters
   - Constrained evidential clustering
4. Comparing and combining the results of soft clustering algorithms
   - Credal Rand index
   - Combining clustering structures

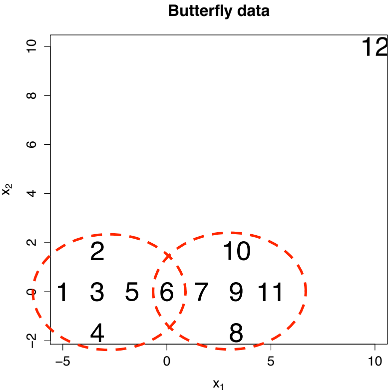
Outline

1. Theory of belief functions: a brief introduction
2. Evidential clustering
   - Credal partition
   - Summarization of a credal partition
   - Relational representation of a credal partition
3. Algorithms
   - EVCLUS
   - Handling a large number of clusters
   - Constrained evidential clustering
4. Comparing and combining the results of soft clustering algorithms
   - Credal Rand index
   - Combining clustering structures

Evidential clustering

- Let $O = \{o_1, \ldots, o_n\}$ be a set of n objects and $\Omega = \{\omega_1, \ldots, \omega_c\}$ be a set of c groups (clusters).
- Each object o_i belongs to at most one group.
- Evidence about the group membership of object o_i is represented by a mass function m_i on Ω:
  - for any nonempty set of clusters $A \subseteq \Omega$, $m_i(A)$ is the degree of support given to the proposition "o_i belongs to one of the clusters in A"
  - $m_i(\emptyset)$ measures the support given to the proposition "o_i does not belong to any of the c groups"
- The n-tuple $M = (m_1, \ldots, m_n)$ is called a credal partition.

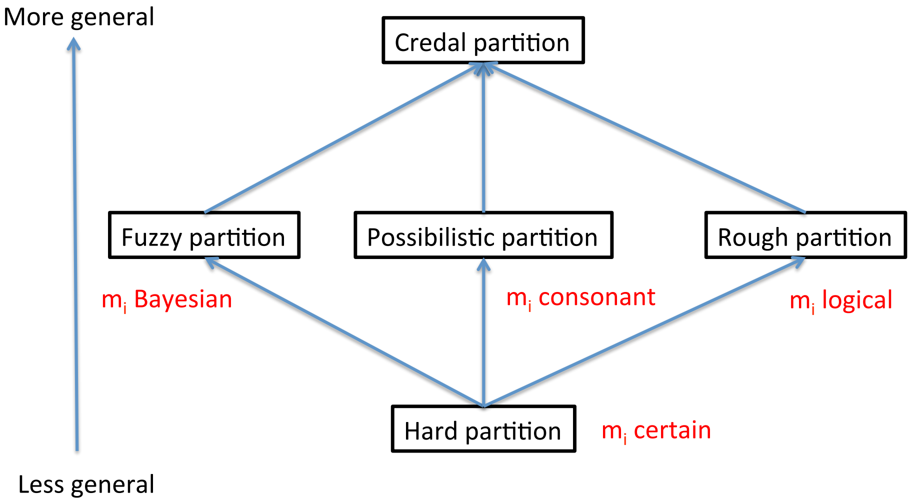
Example



Credal partition

	$\emptyset$	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_1, \omega_2\}$
m_3	0	1	0	0
m_5	0	0.5	0	0.5
m_6	0	0	0	1
m_{12}	0.9	0	0.1	0

Relationship with other clustering structures

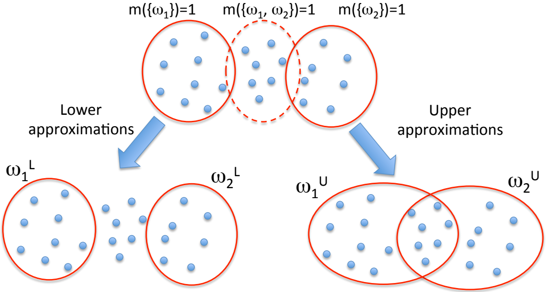


Rough clustering as a special case

- Assume that each m_i is logical, i.e., $m_i(A_i) = 1$ for some $A_i \subseteq \Omega$, $A_i \neq \emptyset$.
- We can then define the lower and upper approximations of cluster ω_k as

$$\underline{\omega}_k = \{o_i \in O | A_i = \{\omega_k\}\}, \quad \overline{\omega}_k = \{o_i \in O | \omega_k \in A_i\}.$$

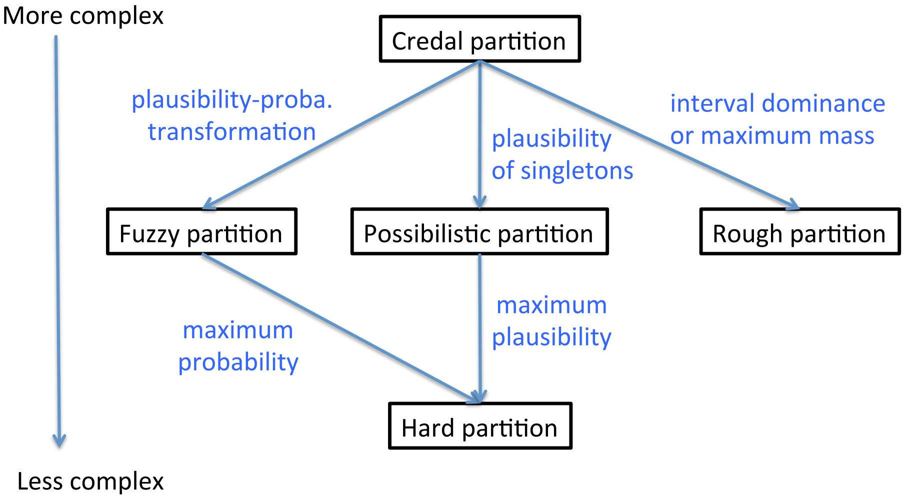
- The membership values to the lower and upper approximations of cluster ω_k are $\underline{u}_{ik} = Bel_i(\{\omega_k\})$ and $\overline{u}_{ik} = Pl_i(\{\omega_k\})$.



Outline

1. Theory of belief functions: a brief introduction
2. Evidential clustering
   - Credal partition
   - Summarization of a credal partition
   - Relational representation of a credal partition
3. Algorithms
   - EVCLUS
   - Handling a large number of clusters
   - Constrained evidential clustering
4. Comparing and combining the results of soft clustering algorithms
   - Credal Rand index
   - Combining clustering structures

Summarization of a credal partition



From evidential to rough clustering

- For each i, let $A_i \subseteq \Omega$ be the set of non dominated clusters

$$A_i = \{\omega \in \Omega | \forall \omega' \in \Omega, Bel_i^*(\{\omega'\}) \leq Pl_i^*(\{\omega\})\},$$

  where Bel_i^* and Pl_i^* are the normalized belief and plausibility functions.
- Lower approximation:

$$\underline{u}_{ik} = \begin{cases} 1 & \text{if } A_i = \{\omega_k\} \\ 0 & \text{otherwise.} \end{cases}$$

- Upper approximation:

$$\overline{u}_{ik} = \begin{cases} 1 & \text{if } \omega_k \in A_i \\ 0 & \text{otherwise.} \end{cases}$$

- The outliers can be identified separately as the objects for which $m_i(\emptyset) \geq m_i(A)$ for all $A \neq \emptyset$.

Outline

1. Theory of belief functions: a brief introduction
2. Evidential clustering
   - Credal partition
   - Summarization of a credal partition
   - Relational representation of a credal partition
3. Algorithms
   - EVCLUS
   - Handling a large number of clusters
   - Constrained evidential clustering
4. Comparing and combining the results of soft clustering algorithms
   - Credal Rand index
   - Combining clustering structures

Relational representation of a hard partition

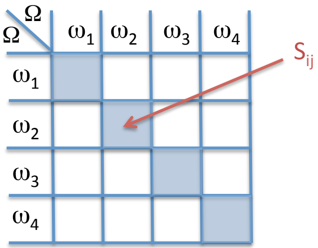
- A hard partition can be represented equivalently by
  - the $n \times c$ membership matrix $U = (u_{ik})$ or
  - an $n \times n$ relation matrix $R = (r_{ij})$ representing the equivalence relation

$$r_{ij} = \begin{cases} 1 & \text{if } o_i \text{ and } o_j \text{ belong to the same group} \\ 0 & \text{otherwise.} \end{cases}$$

- The relational representation R is invariant under renumbering of the clusters, and is thus more suitable to compare or combine several partitions.
- What is the counterpart of matrix R in the case of a credal partition?

Pairwise representation

- Let $M = (m_1, \ldots, m_n)$ be a credal partition.
- For a pair of objects $\{o_i, o_j\}$, let Q_{ij} be the question "Do o_i and o_j belong to the same group?" defined on the frame $\Theta_{ij} = \{S_{ij}, \neg S_{ij}\}$.
- Θ_{ij} is a coarsening of Ω^2.



Given m_i and m_j on Ω, a mass function m_{ij} on Θ_{ij} can be computed as follows:

1. Extend m_i and m_j to Ω^2
2. Combine the extensions of m_i and m_j by the unnormalized Dempster's rule
3. Compute the restriction of the combined mass function to Θ_{ij}

Pairwise mass function

- Mass function:

$$m_{ij}(\emptyset) = m_i(\emptyset) + m_j(\emptyset) - m_i(\emptyset)m_j(\emptyset)$$
$$m_{ij}(\{S_{ij}\}) = \sum_{k=1}^{c} m_i(\{\omega_k\})m_j(\{\omega_k\})$$
$$m_{ij}(\{\neg S_{ij}\}) = \kappa_{ij} - m_{ij}(\emptyset)$$
$$m_{ij}(\Theta_{ij}) = 1 - \kappa_{ij} - \sum_k m_i(\{\omega_k\})m_j(\{\omega_k\})$$

where κ_{ij} is the degree of conflict between m_i and m_j.

- In particular,

$$Pl_{ij}(\{S_{ij}\}) = 1 - \kappa_{ij}$$
$$Pl_{ij}(\{\neg S_{ij}\}) = 1 - m_{ij}(\emptyset) - \sum_k m_i(\{\omega_k\})m_j(\{\omega_k\})$$

Return to CEVCLUS

Relational representation of a credal partition

- Let $M = (m_1, \ldots, m_n)$ be a credal partition.
- The tuple $R = (m_{ij})_{1 \leq i < j \leq n}$ is called the relational representation of credal partition M.

$$M = (m_1, m_2, m_3, m_4, m_5) \longrightarrow R = \begin{pmatrix} & 1 & 2 & 3 & 4 & 5 \\ 1 & \cdot & m_{12} & m_{13} & m_{14} & m_{15} \\ 2 & \cdot & \cdot & m_{23} & m_{24} & m_{25} \\ 3 & \cdot & \cdot & \cdot & m_{34} & m_{35} \\ 4 & \cdot & \cdot & \cdot & \cdot & m_{45} \\ 5 & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

Outline

1. Theory of belief functions: a brief introduction

2. Evidential clustering
   - Credal partition
   - Summarization of a credal partition
   - Relational representation of a credal partition

3. Algorithms
   - EVCLUS
   - Handling a large number of clusters
   - Constrained evidential clustering

4. Comparing and combining the results of soft clustering algorithms
   - Credal Rand index
   - Combining clustering structures

Evidential clustering algorithms

1. Evidential *c*-means (ECM): (Masson and Denoeux, 2008):
   - Attribute data
   - HCM, FCM family
2. EK-NNclus (Denoeux et al, 2015)
   - Attribute or proximity data
   - Searches for the most plausible partition of a dataset
3. EVCLUS (Denoeux and Masson, 2004; Denoeux et al., 2016):
   - Attribute or proximity (possibly non metric) data
   - Multidimensional scaling approach

Outline

1. Theory of belief functions: a brief introduction
2. Evidential clustering
   - Credal partition
   - Summarization of a credal partition
   - Relational representation of a credal partition
3. Algorithms
   - EVCLUS
   - Handling a large number of clusters
   - Constrained evidential clustering
4. Comparing and combining the results of soft clustering algorithms
   - Credal Rand index
   - Combining clustering structures

Learning a Credal Partition from proximity data

- Problem: given the dissimilarity matrix $D = (d_{ij})$, how to build a "reasonable" credal partition ?
- We need a model that relates cluster membership to dissimilarities.
- Basic idea: "The more similar two objects, the more plausible it is that they belong to the same group".
- How to formalize this idea?

Formalization

- Let m_i and m_j be mass functions regarding the group membership of objects o_i and o_j.
- We have seen that the plausibility that objects o_i and o_j belong to the same group is

$$Pl_{ij}(\{S_{ij}\}) = \sum_{A \cap B \neq \emptyset} m_i(A) m_j(B) = 1 - \kappa_{ij}$$

  where κ_{ij} = degree of conflict between m_i and m_j.
- Problem: find a credal partition $M = (m_1, \ldots, m_n)$ such that larger degrees of conflict κ_{ij} correspond to larger dissimilarities d_{ij}.

Cost function

- Approach: minimize the discrepancy between the dissimilarities d_{ij} and the degrees of conflict κ_{ij}.
- Cost (stress) function:

$$J(M) = \sum_{i<j} (\kappa_{ij} - \varphi(d_{ij}))^2$$

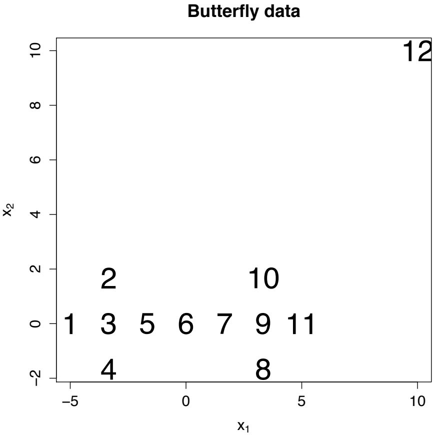
where φ is an increasing function from $[0, +\infty)$ to $[0, 1]$, for instance

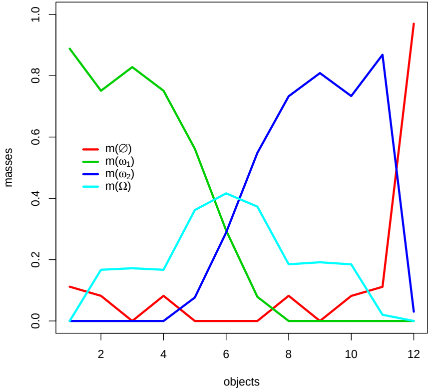
$$\varphi(d) = 1 - \exp(-\gamma d^2).$$

- $J(M)$ can be minimized efficiently using an Iterative Row-wise Quadratic Programming (IRQP) algorithm.

Butterfly example

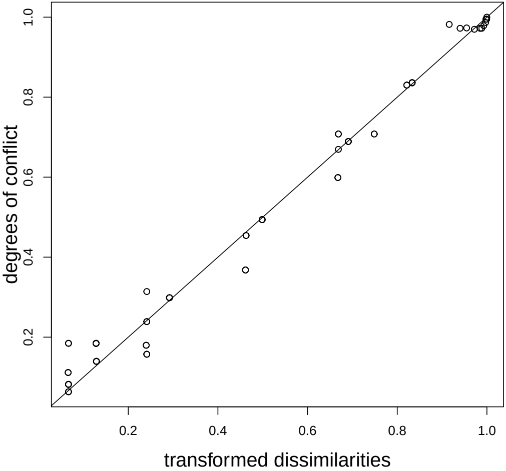
Credal partition



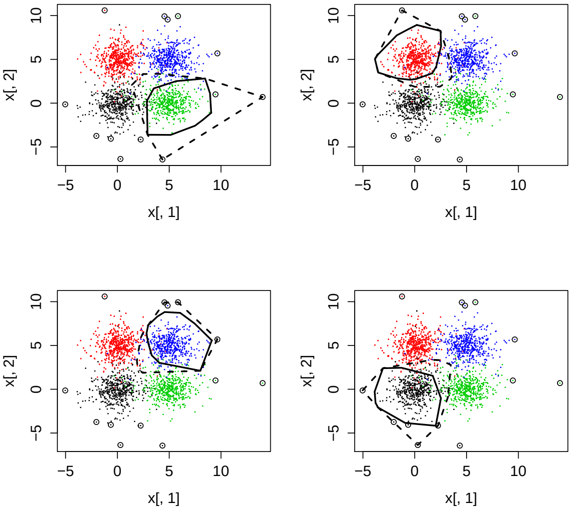


Butterfly example

Shepard diagram



Example with a four-class dataset (2000 objects)



Modification for large datasets

- EVCLUS requires to store the whole dissimilarity matrix: it is inapplicable to large proximity data.
- However, there is usually some redundancy in a dissimilarity matrix.
- In particular, if two objects o_1 and o_2 are very similar, then any object o_3 that is dissimilar from o_1 is usually also dissimilar from o_2.
- Because of such redundancies, it might be possible to compute the differences between degrees of conflict and dissimilarities, for only a subset of randomly sampled dissimilarities.

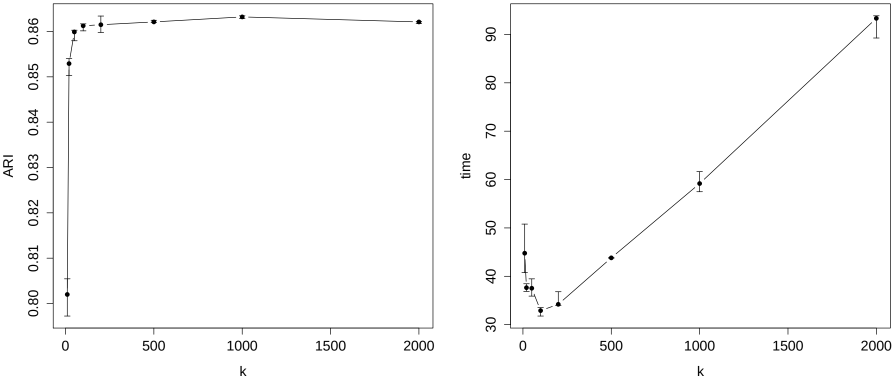
New stress function

- Let $j_1(i), \ldots, j_k(i)$ be k integers sampled at random from the set $\{1, \ldots, i-1, i+1, \ldots, n\}$, for $i = 1, \ldots, n$.
- Let J_k the following stress criterion,

$$J_k(M) = \sum_{i=1}^{n} \sum_{r=1}^{k} \left(\kappa_{i,j_r(i)} - \varphi(d_{i,j_r(i)})\right)^2.$$

- The calculation of $J_k(M)$ requires only $O(nk)$ operations.
- If k can be kept constant as n increases, then time and space complexities are reduced from quadratic to linear.
- This modification makes EVCLUS applicable to large datasets ($\sim 10^4 - 10^5$ objects and hundreds of classes).

Example with simulated data ($n = 10,000$)



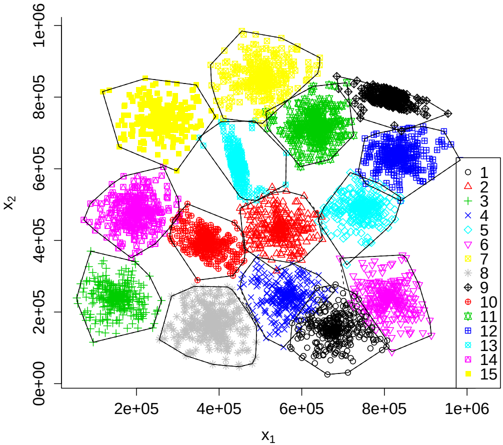
Outline

1. Theory of belief functions: a brief introduction
2. Evidential clustering
   - Credal partition
   - Summarization of a credal partition
   - Relational representation of a credal partition
3. Algorithms
   - EVCLUS
   - Handling a large number of clusters
   - Constrained evidential clustering
4. Comparing and combining the results of soft clustering algorithms
   - Credal Rand index
   - Combining clustering structures

Need to limit the number of focal sets

- If no restriction is imposed on the focal sets, the number of parameters to be estimated in evidential clustering grows exponentially with the number c of clusters, which makes it intractable unless c is small.
- If we allow masses to be assigned to all pairs of clusters, the number of focal sets becomes proportional to c^2, which is manageable for moderate values of c (say, until 10), but still impractical for larger n.
- Idea: assign masses only to pairs of contiguous clusters.
- If each cluster has at most q neighbors, then the number of focal sets is proportional to c.

Example



The S_2 dataset ($n = 5000$) and the 15 clusters found by k-EVCLUS with $k = 100$

Method

Step1: Run a clustering algorithm (e.g., ECM or EVCLUS) with focal sets of cardinalities 0, 1 and (optionally) c. A credal partition M_0 is obtained.

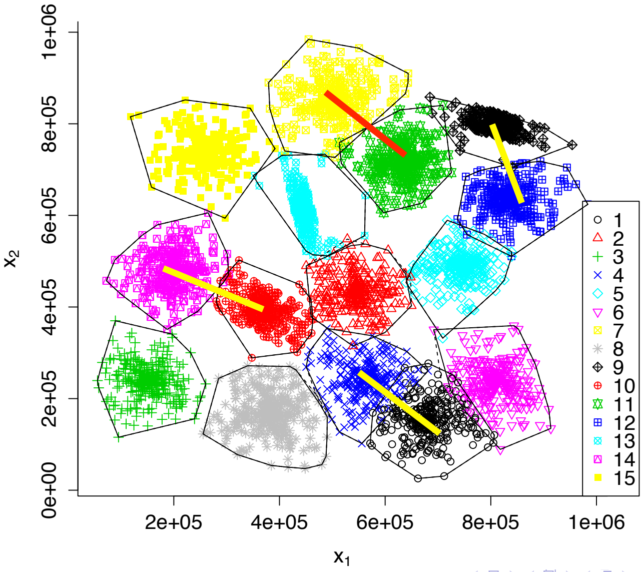
Step 2: Compute the similarity between each pair of clusters (ω_j, ω_ℓ) as

$$S(j, \ell) = \sum_{i=1}^{n} pl_{ij} pl_{i\ell},$$

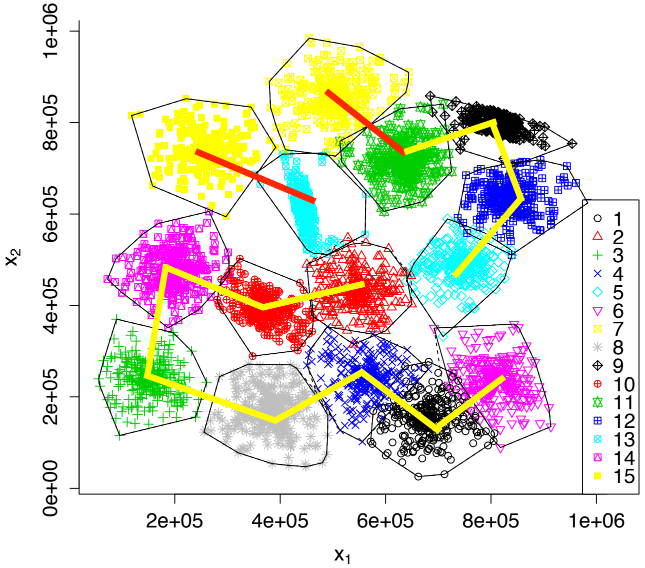
where pl_{ij} and $pl_{i\ell}$ are the normalized plausibilities that object i belongs, respectively, to clusters j and ℓ. Determine the set P_K of pairs $\{\omega_j, \omega_\ell\}$ that are mutual q nearest neighbors.

Step 3: Run the clustering algorithm again, starting from the previous credal partition M_0, and adding as focal sets the pairs in P_K.

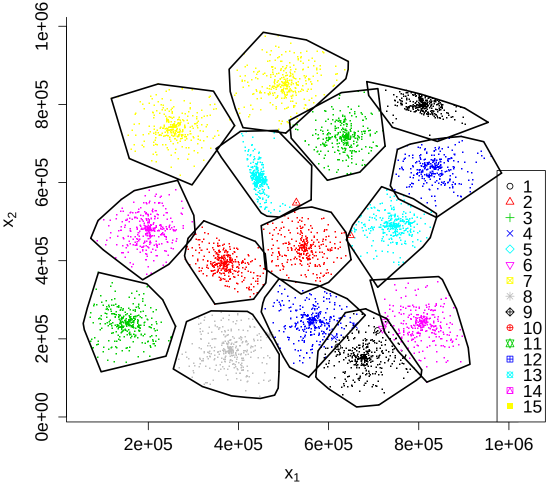
Pairs of mutual neighbors with $q = 1$



Pairs of mutual neighbors with $q = 2$

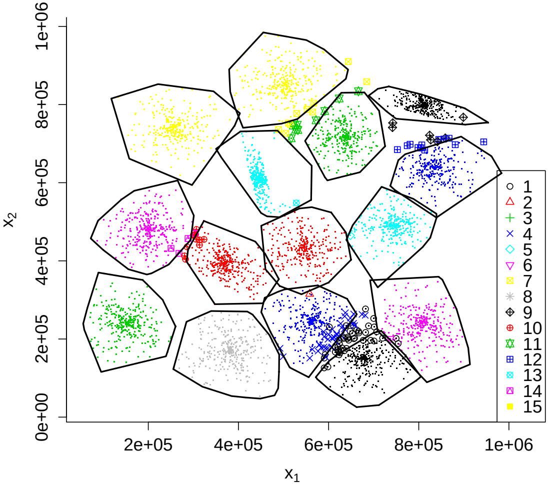


Initial credal partition $\mathcal{M}_0$



Lower approximations and ambiguous objects for the initial credal partition $\mathcal{M}_0$

Final credal partition ($q = 1$)



Lower approximations and ambiguous objects for the final credal partition

Outline

1. Theory of belief functions: a brief introduction
2. Evidential clustering
   - Credal partition
   - Summarization of a credal partition
   - Relational representation of a credal partition
3. Algorithms
   - EVCLUS
   - Handling a large number of clusters
   - Constrained evidential clustering
4. Comparing and combining the results of soft clustering algorithms
   - Credal Rand index
   - Combining clustering structures

Constrained EVCLUS

- In some cases, we may have some prior knowledge about the group membership of some objects.
- Such knowledge may take the form of instance-level constraints of two kinds:
  1. Must-link (ML) constraints, which specify that two objects certainly belong to the same cluster
  2. Cannot-link (CL) constraints, which specify that two objects certainly belong to different clusters
- How to take into account such constraints?

Modified cost-function

- To take into account ML and CL constraints, we can modify the cost function of k-EVCLUS as

$$J_{kC}(\mathcal{M}) = \eta \sum_{i=1}^{n} \sum_{r=1}^{k} \left(\kappa_{i,j_r(i)} - \delta_{i,j_r(i)}\right)^2 + \frac{\xi}{2(|\mathrm{ML}| + |\mathrm{CL}|)}(J_{\mathrm{ML}} + J_{\mathrm{CL}}),$$

with

$$J_{\mathrm{ML}} = \sum_{(i,j)\in \mathrm{ML}} Pl_{ij}(\{\neg S_{ij}\}) + 1 - Pl_{ij}(\{S_{ij}\}),$$

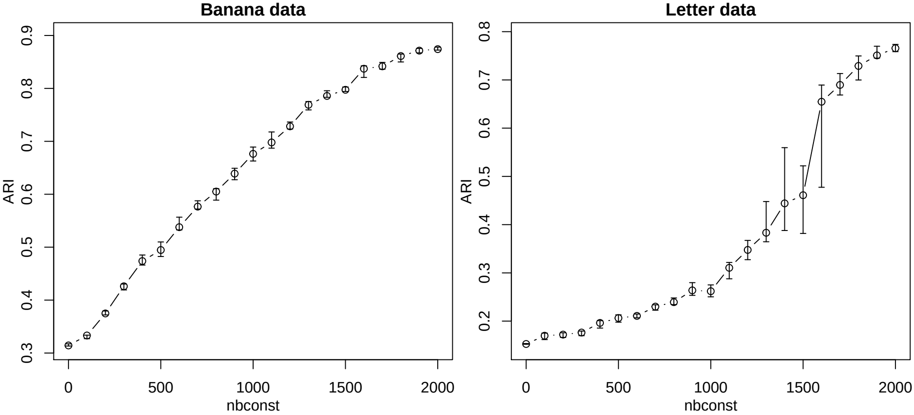
$$J_{\mathrm{CL}} = \sum_{(i,j)\in \mathrm{CL}} Pl_{ij}(\{S_{ij}\}) + 1 - Pl_{ij}(\{\neg S_{ij}\}),$$

where

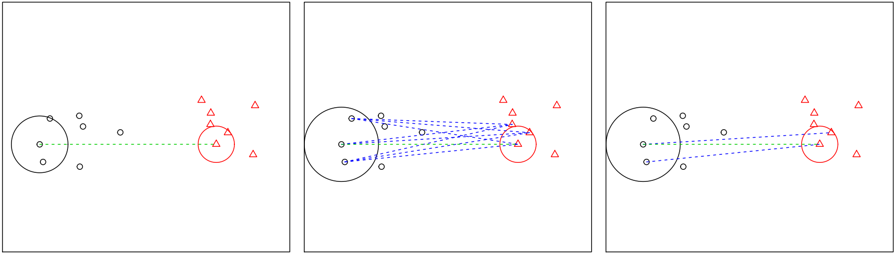
- ML and CL are, respectively, the sets of ML and CL constraints.
- $Pl_{ij}(\{\neg S_{ij}\})$ and $Pl_{ij}(\{\neg S_{ij}\})$ are computed from pairwise mass function m_{ij}

▸ Go back to pairwise mass functions

Results

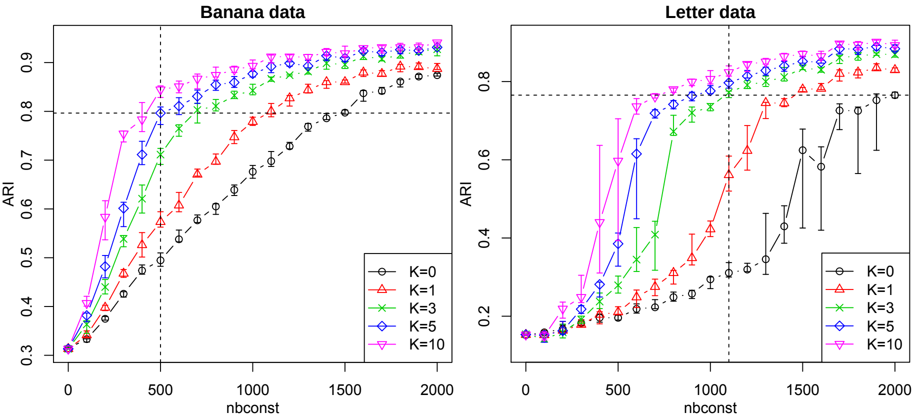

Banana data
ARI
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0
500
1000
1500
2000
nbconst
Letter data
ARI
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0
500
1000
1500
2000
nbconst

Constraint expansion



(a) A CL constraint $(o_i, o_j) \in$ CL, with the K-neighborhoods $\mathcal{N}_K(o_i)$ and $\mathcal{N}_K(o_j)$ of o_i and o_j, respectively ($K = 2$). (b) The set $\mathcal{P}_K(o_i, o_j)$ of pairs of a neighbor of o_i and a neighbor of o_j. (c) The $K = 2$ new CL constraints.

Results


Banana data
Letter data
ARI
nbconst
0
500
1000
1500
2000
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0.2
K=0
K=1
K=3
K=5
K=10

Outline

1 Theory of belief functions: a brief introduction

2 Evidential clustering
- Credal partition
- Summarization of a credal partition
- Relational representation of a credal partition

3 Algorithms
- EVCLUS
- Handling a large number of clusters
- Constrained evidential clustering

4 Comparing and combining the results of soft clustering algorithms
- Credal Rand index
- Combining clustering structures

Exploiting the generality of evidential clustering

- We have seen that the concept of credal partition subsumes the main hard and soft clustering structures.
- Consequently, methods designed to evaluate or combine credal partitions can be used to evaluate or combine the results of any hard or soft clustering algorithms.
- Two such methods will be described:
    1. A generalization of the Rand index to compute the distance between two credal partitions;
    2. A method to combine credal partitions.

Outline

1. Theory of belief functions: a brief introduction
2. Evidential clustering
   - Credal partition
   - Summarization of a credal partition
   - Relational representation of a credal partition
3. Algorithms
   - EVCLUS
   - Handling a large number of clusters
   - Constrained evidential clustering
4. Comparing and combining the results of soft clustering algorithms
   - Credal Rand index
   - Combining clustering structures

Rand index

- The Rand index is a widely used measure of similarity between two hard partitions.
- It is defined as

$$\mathrm{RI} = \frac{a+b}{n(n-1)/2}$$

  with
  - $a =$ number of pairs of objects that are grouped together in both partitions
  - $b =$ number of pairs of objects that are assigned to different clusters in both partitions.
- How to generalize the Rand Index to credal partitions?

Belief distance

- Let $R = (m_{ij})$ and $R' = (m'_{ij})$ be the relational representations of two credal partitions.
- The assess the distance between R and R', we can average the distances between the m_{ij}'s and m'_{ij}'s.
- Belief distance between mass functions:

$$\delta_B(m_{ij}, m'_{ij}) = \frac{1}{2} \sum_{A \subseteq \Theta} \mid Bel_{ij}(A) - Bel'_{ij}(A) \mid,$$

  where Bel_{ij} and Bel'_{ij} are, respectively, the belief functions associated to m_{ij} and m'_{ij}.
- Property: $\delta_B(m_{ij}, m'_{ij}) \in [0, 1]$.

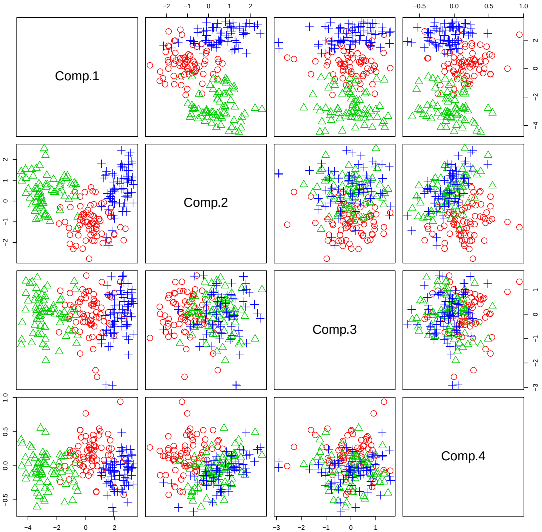
Similarity index

- We define the Credal Rand Index between two credal partitions as

$$\mathrm{CRI}(R, R') = 1 - \frac{\sum_{i<j} \delta_B(m_{ij}, m'_{ij})}{n(n-1)/2}.$$

- Properties:
  - $0 \leq \mathrm{CRI}(R, R') \leq 1$
  - $\mathrm{CRI} = \mathrm{RI}$ when the two partitions are hard
  - Symmetry: $\mathrm{CRI}(R, R') = \mathrm{CRI}(R', R)$
  - If $R = R'$, then $\mathrm{CRI}(R, R') = 1$
  - $1 - \mathrm{CRI}$ is a metric in the space $\mathcal{R}$ of relational representations
- The CRI can be used to compare the results of any two hard or soft clustering algorithms.

Example: Seeds data

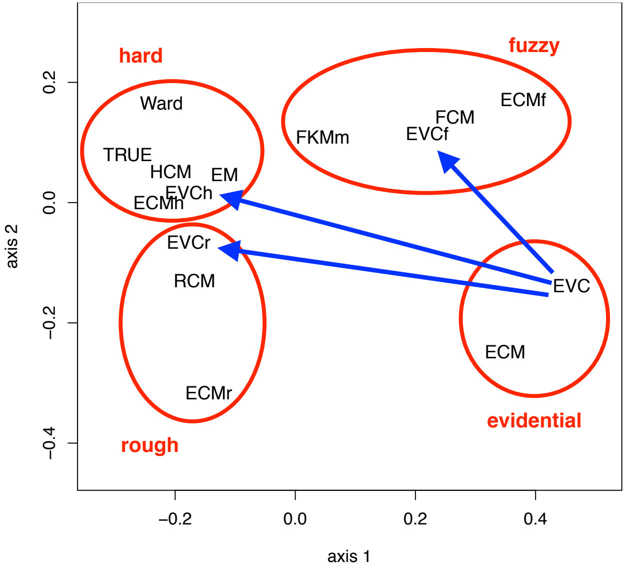


- Seeds from three different varieties of wheat: Kama, Rosa and Canadian, 70 elements each
- 7 features

Clustering algorithms

algorithm	R function	R package
ECM	`ecm`	`evclust`
EVCLUS	`kevclus`	`evclust`
HCM	`kmeans`	`stats`
Hierarchical clust. (Ward distance)	`hclust`	`stats`
FCM	`FKM`	`fclust`
Fuzzy k-medoids	`FKM.med`	`fclust`
π-Rough k-means	`RoughKMeans_PI`	`SoftClustering`
EM	`Mclust`	`mclust`

Result: MDS configuration



Outline

1. Theory of belief functions: a brief introduction

2. Evidential clustering
   - Credal partition
   - Summarization of a credal partition
   - Relational representation of a credal partition

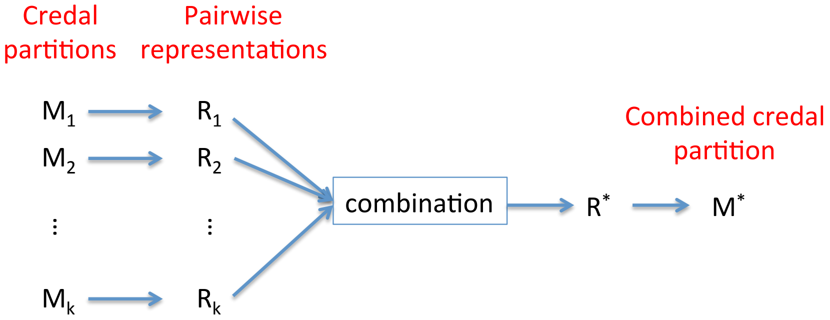
3. Algorithms
   - EVCLUS
   - Handling a large number of clusters
   - Constrained evidential clustering

4. Comparing and combining the results of soft clustering algorithms
   - Credal Rand index
   - Combining clustering structures

Motivations for combining clustering structures

- Let $M_1, \ldots, M_N$ be an ensemble of N credal partitions generated by hard or soft (fuzzy, rough, etc.) clustering structures.
- It may be useful to combine these credal partitions:
  - to increase the chance of finding a good approximation to the true partition, or
  - to highlight invariant patterns across the clustering structures.
- Combination is easily carried out using pairwise representations.

Combination method



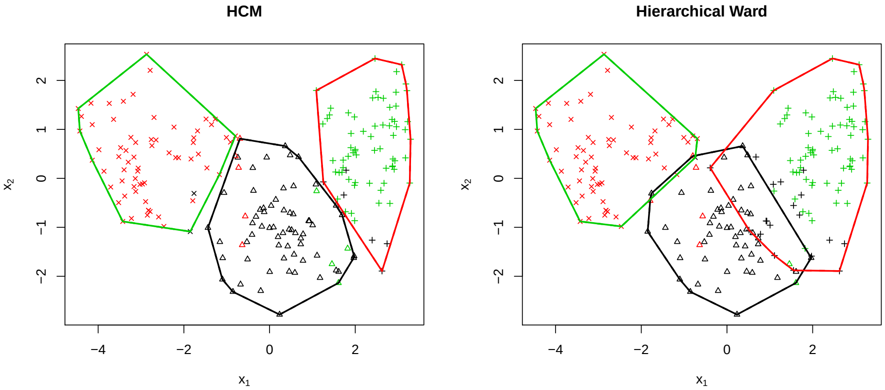
The combined credal partition can be defined as

$$M^* = \arg\max_M \mathrm{CRI}(\mathcal{R}(M), R^*),$$

where $\mathcal{R}(M)$ denotes the relational representation of M.

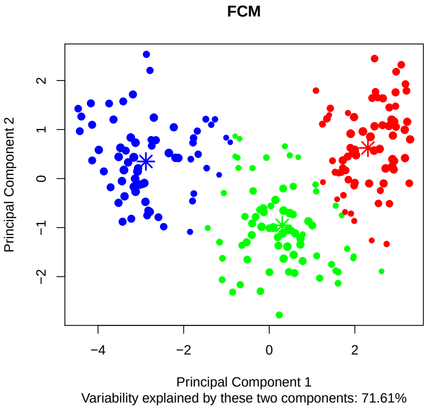
Example: seeds data

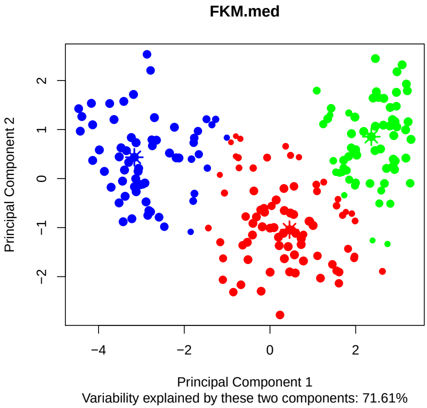
Hard clustering results



Example: seeds data

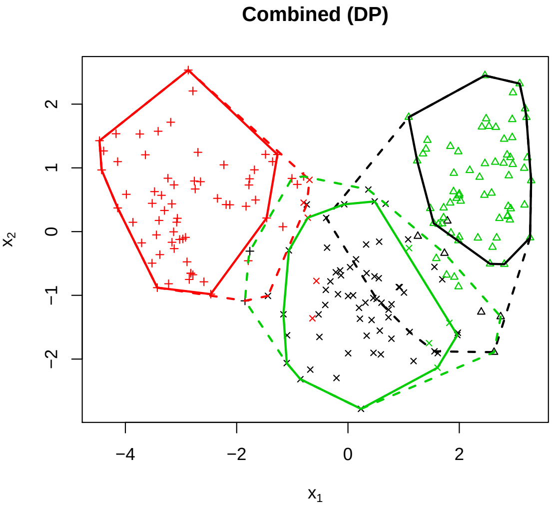
Fuzzy clustering results





Example: seeds data

Combined credal partition (Dubois-Prade rule)



Summary

- The Dempster-Shafer theory of belief functions provides a rich and flexible framework to represent uncertainty in clustering.
- The concept of credal partition encompasses the main existing soft clustering concepts (fuzzy, possibilistic, rough partitions).
- Efficient algorithms exist, allowing one to generate credal partitions from attribute or proximity datasets.
- These algorithms can be applied to large datasets and large numbers of clusters (by carefully selecting the focal sets).
- Concepts from the theory of belief functions make it possible to compare and combine clustering structures generated by various soft clustering algorithms.

Future research directions

- Combining clustering structures in various settings
  - distributed clustering,
  - combination of different attributes, different algorithms,
  - etc.
- Handling huge datasets (several millions of objects)
- Criteria for selecting the number of clusters
- Semi-supervised clustering
- Clustering imprecise or uncertain data
- Applications to image processing, social network analysis, process monitoring, etc.
- Etc...

The evclust package

evclust: Evidential Clustering

Various clustering algorithms that produce a credal partition, i.e., a set of Dempster-Shafer mass functions representing the membership of objects to clusters. The mass functions quantify the cluster-membership uncertainty of the objects. The algorithms are: Evidential c-Means (ECM), Relational Evidential c-Means (RECM), Constrained Evidential c-Means (CECM), EVCLUS and EK-NNclus.

Version:	1.0.3
Depends:	R (≥ 3.1.0)
Imports:	FNN, R.utils, limSolve, Matrix
Suggests:	knitr, rmarkdown
Published:	2016-09-04
Author:	Thierry Denoeux
Maintainer:	Thierry Denoeux <tdenoeux at utc.fr>
License:	GPL-3
NeedsCompilation:	no
In views:	Cluster
CRAN checks:	evclust results

https://cran.r-project.org/web/packages

References I

cf. https://www.hds.utc.fr/~tdenoeux

M.-H. Masson and T. Denœux.
ECM: An evidential version of the fuzzy c-means algorithm.
Pattern Recognition, 41(4):1384-1397, 2008.

T. Denœux and M.-H. Masson.
EVCLUS: Evidential Clustering of Proximity Data.
IEEE Transactions on SMC B, 34(1):95-109, 2004.

V. Antoine, B. Quost, M.-H. Masson and T. Denoeux.
CEVCLUS: Evidential clustering with instance-level constraints for relational data.
Soft Computing, 18(7):1321-1335, 2014.

T. Denœux, S. Sriboonchitta and O. Kanjanatarakul
Evidential clustering of large dissimilarity data.
Knowledge-Based Systems, 106:179–195, 2016.

References II

cf. `https://www.hds.utc.fr/~tdenoeux`

- T. Denoeux, O. Kanjanatarakul and S. Sriboonchitta.
  EK-NNclus: a clustering procedure based on the evidential K-nearest neighbor rule.
  Knowledge-Based Systems, Vol. 88, pages 57-69, 2015.

- T. Denoeux, S. Li and S. Sriboonchitta.
  Evaluating and Comparing Soft Partitions: an Approach Based on Dempster-Shafer Theory.
  IEEE Transactions on Fuzzy Systems, In Press, 2017.