Soft Clustering

Davide Ciucci

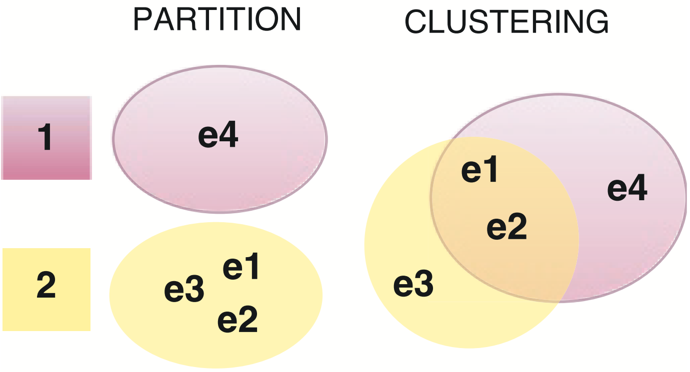
Department of Informatics, Systems and Communication
University of Milan-Bicocca

Uncertainty in Computer Science

Outline

Outline

Example of Soft Clustering



Evidential Clustering

- Set of k clusters: $\{C_1, C_2, \ldots, C_k\}$ and n objects $\{o_1, o_2, \ldots, o_n\}$

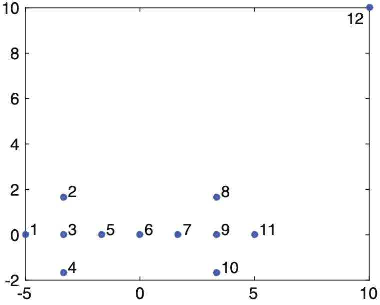
Evidential Clustering

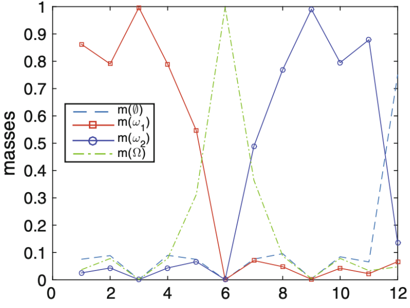
- Set of k clusters: $\{C_1, C_2, \ldots, C_k\}$ and n objects $\{o_1, o_2, \ldots, o_n\}$
- Given an object o_i the membership to a cluster is represented by a mass function m_i

Evidential Clustering

- Set of k clusters: $\{C_1, C_2, \ldots, C_k\}$ and n objects $\{o_1, o_2, \ldots, o_n\}$
- Given an object o_i the membership to a cluster is represented by a mass function m_i
- $m_i(A)$ represents the "degree of support attached to the proposition "the true cluster of object o_i is in A ", and to no more specific proposition"
- $m_1, \ldots, m_n$ is a credal partition

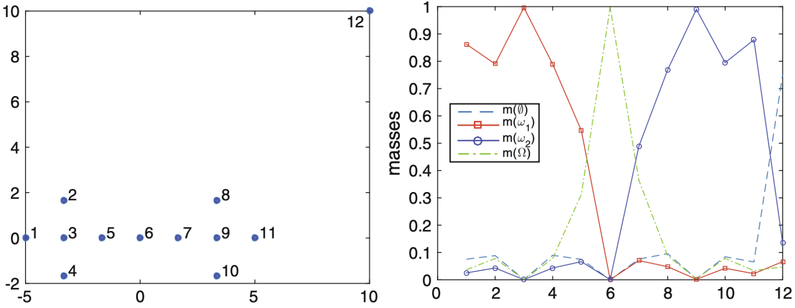
Credal partition - example





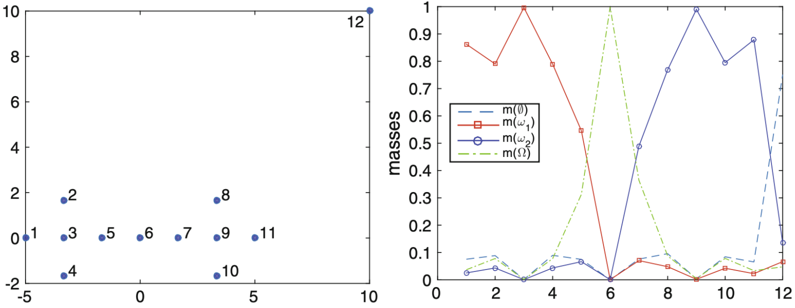
- Butterfly dataset + point 12 (outlier), $k = 2$

Credal partition - example



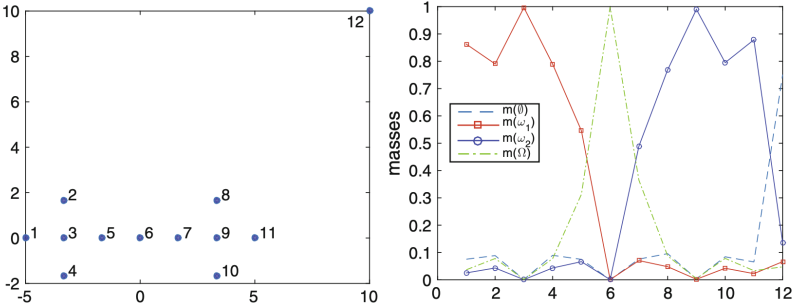
- Butterfly dataset + point 12 (outlier), $k = 2$
- $m_3(C_1) = 1$: object o_3 surely belongs to cluster C_1

Credal partition - example



- Butterfly dataset + point 12 (outlier), $k = 2$
- $m_3(C_1) = 1$: object o_3 surely belongs to cluster C_1
- $m_6(\{C_1, C_2\}) = 1$ max uncertainty: objects o_6 belongs to C_1 or C_2

Credal partition - example



- Butterfly dataset + point 12 (outlier), $k = 2$
- $m_3(C_1) = 1$: object o_3 surely belongs to cluster C_1
- $m_6(\{C_1, C_2\}) = 1$ max uncertainty: objects o_6 belongs to C_1 or C_2
- $m_{12}(\emptyset) = 0.7$ (nb: no normalization): object o_{12} does not belong to any cluster

Thierry Denoeux, Orakanya Kanjanatarakul: Beyond Fuzzy, Possibilistic and Rough: An Investigation of Belief Functions in Clustering. SMPS 2016: 157-164

Credal partition algorithms

- Three different algorithms: Evclus, Evidential k-means, EK-nn
- Evidential k-means, EK-nn: very efficient with discrete data

Credal partition algorithms

- Three different algorithms: Evclus, Evidential k-means, EK-nn
- Evidential k-means, EK-nn: very efficient with discrete data
- Ek-nn: can determine k, the number of clusters, automatically

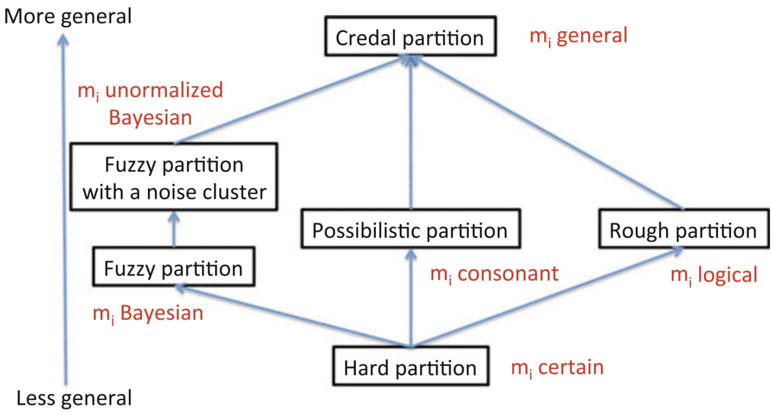
Credal partition algorithms

- Three different algorithms: Evclus, Evidential k-means, EK-nn
- Evidential k-means, EK-nn: very efficient with discrete data
- Ek-nn: can determine k, the number of clusters, automatically
- Evclus, Evidential k-means "produce more informative outputs (with masses assigned to any subsets of clusters)"

Credal partition algorithms

- Three different algorithms: Evclus, Evidential k-means, EK-nn
- Evidential k-means, EK-nn: very efficient with discrete data
- Ek-nn: can determine k, the number of clusters, automatically
- Evclus, Evidential k-means "produce more informative outputs (with masses assigned to any subsets of clusters)"
- Evclus: "very effective for dealing with non metric dissimilarity data, and it is suitable to handle very large datasets"

Credal Partition



Fuzzy Clustering

- Each cluster is described by a fuzzy set f_{C_i}
- Constraint: $\sum_i f_{C_i}(o_j) = 1$ (fuzzy?!?)

Fuzzy Clustering

- Each cluster is described by a fuzzy set f_{C_i}
- Constraint: $\sum_i f_{C_i}(o_j) = 1$ (fuzzy?!?)
- Fuzzy k-means: the membership degree of an object to a specific cluster is calculated based on the relative distances of the object to the other clusters

Fuzzy Clustering

- Each cluster is described by a fuzzy set f_{C_i}
- Constraint: $\sum_i f_{C_i}(o_j) = 1$ (fuzzy?!?)
- Fuzzy k-means: the membership degree of an object to a specific cluster is calculated based on the relative distances of the object to the other clusters
- Many other algorithms
  - Noise clustering: objects far away from all clusters obtain a high membership degree to a noise cluster

Fuzzy Clustering

- Each cluster is described by a fuzzy set f_{C_i}
- Constraint: $\sum_i f_{C_i}(o_j) = 1$ (fuzzy?!?)
- Fuzzy k-means: the membership degree of an object to a specific cluster is calculated based on the relative distances of the object to the other clusters
- Many other algorithms
  - Noise clustering: objects far away from all clusters obtain a high membership degree to a noise cluster
- Tends to avoid local minima
- Can generate wrong global minima, if not handled (i.e., noise clustering) forces outliers to belong to one cluster

KLAWONN, Frank; KRUSE, Rudolf; WINKLER, Roland. Fuzzy clustering: More than just fuzzification. Fuzzy sets and systems, 2015, 281: 272-279

Possibilistic clustering

- Relaxed the constraint that membership degrees must sum up to 1: $\exists i,\ f_{C_i}(o_j) > 0$

Possibilistic clustering

- Relaxed the constraint that membership degrees must sum up to 1: $\exists i,\ f_{C_i}(o_j) > 0$
- f_{C_i} are possibility distributions
- $f_{C_i}(o_j)$ represents the typicality of object o_j relative to cluster C_i

Possibilistic clustering

- Relaxed the constraint that membership degrees must sum up to 1: $\exists i,\ f_{C_i}(o_j) > 0$
- f_{C_i} are possibility distributions
- $f_{C_i}(o_j)$ represents the typicality of object o_j relative to cluster C_i
- an object can have only small membership values $f_{C_1}(o) = f_{C_2}(o) = 0.2$

Possibilistic clustering

- Relaxed the constraint that membership degrees must sum up to 1: $\exists i,\ f_{C_i}(o_j) > 0$
- f_{C_i} are possibility distributions
- $f_{C_i}(o_j)$ represents the typicality of object o_j relative to cluster C_i
- an object can have only small membership values $f_{C_1}(o) = f_{C_2}(o) = 0.2$
- Sensible to initialization

Possibilistic clustering

- Relaxed the constraint that membership degrees must sum up to 1: $\exists i,\ f_{C_i}(o_j) > 0$
- f_{C_i} are possibility distributions
- $f_{C_i}(o_j)$ represents the typicality of object o_j relative to cluster C_i
- an object can have only small membership values $f_{C_1}(o) = f_{C_2}(o) = 0.2$
- Sensible to initialization
- Other models:
  - Hybrid models fuzzy-possibility

Possibilistic clustering

- Relaxed the constraint that membership degrees must sum up to 1: $\exists i,\ f_{C_i}(o_j) > 0$
- f_{C_i} are possibility distributions
- $f_{C_i}(o_j)$ represents the typicality of object o_j relative to cluster C_i
- an object can have only small membership values $f_{C_1}(o) = f_{C_2}(o) = 0.2$
- Sensible to initialization
- Other models:
  - Hybrid models fuzzy-possibility
  - To reduce the complexity: shadowed set c-means (Shadowed set = three-valued set)

Rough Clustering

- Inspired by rough sets: Each cluster C_i is made by a lower region and an uncertain region, named boundary

Rough Clustering

- Inspired by rough sets: Each cluster C_i is made by a lower region and an uncertain region, named boundary
- An instance cannot belong to only one boundary

Rough Clustering

- Inspired by rough sets: Each cluster C_i is made by a lower region and an uncertain region, named boundary
- An instance cannot belong to only one boundary
- Rough k-means: An element x is assigned to the boundary of two (or more) clusters if it is approximately to the same distance to the centroid's clusters

Rough Clustering and Orthopartitions

- One Cluster $\rightarrow$ an orthopair
  Collection of Clusters $\rightarrow$ an orthopartition

Rough Clustering and Orthopartitions

- One Cluster $\rightarrow$ an orthopair
  Collection of Clusters $\rightarrow$ an orthopartition

  A rough clustering generates an orthopartition

Rough Clustering and Orthopartitions

- One Cluster $\rightarrow$ an orthopair
  Collection of Clusters $\rightarrow$ an orthopartition

  A rough clustering generates an orthopartition

- Indeed, a rough clustering should satisfy (Lingras, Peters):
  - $\forall x \in U$, there *exists* at most one lower approximation containing $x \Rightarrow$ Orthopartition axiom 1

Rough Clustering and Orthopartitions

- One Cluster $\rightarrow$ an orthopair
  Collection of Clusters $\rightarrow$ an orthopartition

  A rough clustering generates an orthopartition

- Indeed, a rough clustering should satisfy (Lingras, Peters):
  - $\forall x \in U$, there *exists* at most one lower approximation containing $x \Rightarrow$ Orthopartition axiom 1
  - $\forall x \in U$, if x does not belong to any lower approximation, then, it belongs to at least two upper approximations $=$ Orthopartition axiom 3

Rough Clustering and Orthopartitions

- One Cluster $\rightarrow$ an orthopair
  Collection of Clusters $\rightarrow$ an orthopartition

  A rough clustering generates an orthopartition

- Indeed, a rough clustering should satisfy (Lingras, Peters):
  - $\forall x \in U$, there *exists* at most one lower approximation containing $x \Rightarrow$ Orthopartition axiom 1
  - $\forall x \in U$, if x does not belong to any lower approximation, then, it belongs to at least two upper approximations $=$ Orthopartition axiom 3
  - the above conditions $+$ (lower $\subseteq$ upper) $\Rightarrow$ Orthopartition axiom 2

Soft Clustering Techniques

Three-way clustering

- Lower approximations cannot be empty

Soft Clustering Techniques

Three-way clustering

- Lower approximations cannot be empty
- An instance can be in only one boundary

Soft Clustering Techniques

Three-way clustering

- Lower approximations cannot be empty
- An instance can be in only one boundary

Example

$C_1 = (\{x_3, x_4\}, \{x_3, x_4, x_5\})$
$C_2 = (\{x_6, x_9\}, \{x_6, x_9, x_1, x_8\})$
$C_3 = (\{x_2, x_7\}, \{x_2, x_7, x_8\}))\}$

x_5 is only in the boundary of C_1

Interval/Three-way Clustering and Orthopartitions

Problem: interval-set/three-way clustering allows an object to be in the boundary of only one cluster (not permitted in orthopartitions)

Interval/Three-way Clustering and Orthopartitions

Problem: interval-set/three-way clustering allows an object to be in the boundary of only one cluster (not permitted in orthopartitions)

Solution: collect all these problematic objects (which represent outliers) in a set *Prob* and define the new cluster $(\emptyset, Prob)$

Interval/Three-way Clustering and Orthopartitions

Problem: interval-set/three-way clustering allows an object to be in the boundary of only one cluster (not permitted in orthopartitions)

Solution: collect all these problematic objects (which represent outliers) in a set *Prob* and define the new cluster $(\emptyset, \mathit{Prob})$

Interval set clustering result + $(\emptyset, \mathit{Prob})$ is an orthopartition

Outline

Soft Clustering approaches

Evaluation

- Rough Clustering
- General case

Outline

Hard Clustering evaluation

- External evaluation: the cluster is compared to a gold standard

Hard Clustering evaluation

- External evaluation: the cluster is compared to a gold standard
- Several indices exists
  - Mutual information, purity (here)
  - Rand, Jaccard, Fowlkes-Mellows (no need of orthopartitions)

Normalized Mutual Information

$$NMI(\Omega, C) = \frac{m(\Omega, C)}{min\{h(\Omega), h(C)\}}$$

- m is a mutual information and h its corresponding entropy
- C is the soft-clustering result
- Ω is the gold standard (it can be a partition or an orthopartition)

Purity

- Partition

$$\text{purity}(\Omega, C) = \frac{1}{|U|} \sum_{\omega \in \Omega} \max_{c \in C} \{|\omega \cap c|\}$$

Purity

- Partition

$$\text{purity}(\Omega, C) = \frac{1}{|U|} \sum_{\omega \in \Omega} \max_{c \in C} \{|\omega \cap c|\}$$

- Orthopartition

$$\text{soft-purity}(\Omega, C) = \frac{1}{|U|} \sum_{O_i} \max_{C_j} P(O_i, C_j)$$

Idea $P(O_i, C_j)$ measures the degree of similarity between one of the clusters $O_i \in \Omega$ and one of the classes $C_j \in C$ (weighting the elements in the boundaries differently)

Purity

- Partition

$$\text{purity}(\Omega, C) = \frac{1}{|U|} \sum_{\omega \in \Omega} \max_{c \in C} \{ |\omega \cap c| \}$$

- Orthopartition

$$\text{soft-purity}(\Omega, C) = \frac{1}{|U|} \sum_{O_i} \max_{C_j} P(O_i, C_j)$$

Idea $P(O_i, C_j)$ measures the degree of similarity between one of the clusters $O_i \in \Omega$ and one of the classes $C_j \in C$ (weighting the elements in the boundaries differently)

$$P(O_i, C_j) = |P_i \cap P_j| + \sum_{x \in Bnd_i \cap P_j} \frac{1}{|\{O_k \in \mathcal{O} | x \in Bnd_k\}|} + \sum_{x \in Bnd_j \cap P_i} \frac{1}{|\{C_k \in C | x \in Bnd_k}$$

$$\sum_{x \in Bnd_j \cap Bnd_i} \left[\frac{1}{|\{C_k \in C | x \in Bnd_k\}|} * \frac{1}{|\{O_k \in \mathcal{O} | x \in Bnd_k\}|} \right]$$

Experiments

There exists several rough/three-way clustering algorithm. We compared

- 8 partition based
- 5 density based

Preliminary Results (Mutual Information)

MI	k–means	rough k–means
Iris	0.86	**0.90**
Wine	**0.94**	0.77
Zoo	0.89	**0.91**
Yeast	**0.80**	0.72
Transfusion	**0.36**	0.23
Abalone	**0.90**	**0.90**
Arrhythmia	0.69	**0.70**
Anuran	**0.84**	**0.84**
Dota2	0.25	**0.46**
Adult	**0.40**	0.35

- Manual inspection of the clusters says that to high MI corresponds a similarity with the gold standard
- with the exception of the dataset Abalone. However...

Preliminary Results (Purity)

SP	k–means	rough k–means
Iris	0.84	**0.89**
Wine	**0.95**	0.69
Zoo	**0.88**	0.84
Yeast	**0.52**	0.42
Transfusion	0.76	0.76
Abalone	0.28	0.27
Arrhythmia	0.58	0.58
Anuran	**0.95**	0.80
Dota2	0.53	0.53
Adult	0.76	0.76

- Mutual information and purity should be taken in combination to have a meaningful performance measure

More extensive study

Mean value table of each algorithm on each metric

Model	R-Rand	Lower-Rand	Coverage	SAMI	SLMI
KM*	0.766	0.766	1.0	0.880	0.880
GM*	0.751	0.751	1.0	0.872	0.872
RKM	0.776	0.785	0.803	0.869	0.883
PRKM	0.765	0.769	**0.980**	**0.871**	0.873
RGKM	0.767	0.787	0.763	0.853	0.880
RDCM	**0.791**	**0.846**	0.788	0.837	**0.892**
TWCM	0.766	0.776	0.953	**0.871**	0.875
TWKM	0.741	0.756	0.949	0.844	0.863
TWCK	0.739	0.810	0.686	0.820	0.846
TWCS	0.686	0.758	0.517	0.760	0.831

* Hard clustering, rough clustering, three-way clustering

More extensive study

The differences are not always statistically significant

Model	KM	GM	RKM	PRKM	RGKM	RDCM	TWCM	TWKM	TWCK
KM	1.0	0.419	0.136	0.720	**0.034**	0.435	0.760	0.154	**0.009**
GM	0.419	1.0	0.491	0.652	0.183	0.979	0.614	0.533	0.066
RKM	0.136	0.491	1.0	0.256	0.516	0.474	0.234	0.947	0.245
PRKM	0.720	0.652	0.256	1.0	0.076	0.671	0.958	0.284	**0.023**
RGKM	**0.034**	0.183	0.516	0.076	1.0	0.174	0.068	0.474	0.605
RDCM	0.435	0.979	0.474	0.671	0.174	1.0	0.633	0.516	0.062
TWCM	0.760	0.614	0.234	0.958	0.068	0.633	1.0	0.261	**0.020**
TWKM	0.154	0.533	0.947	0.284	0.474	0.516	0.261	1.0	0.219
TWCK	**0.009**	0.066	0.245	**0.023**	0.605	0.062	**0.020**	0.219	1.0
TWCS	**5.2e-05**	**0.001**	**0.007**	**1.9e-04**	**0.037**	**0.001**	**1.6e-04**	**0.006**	0.114

Table: Pairwise Quade Test: SAMI

Outline

Soft Clustering evaluation

A part from rough set clustering, there exist some extension of standard measure, in particular different version of Rand index

Soft Clustering evaluation

A part from rough set clustering, there exist some extension of standard measure, in particular different version of Rand index

- often they fail to satisfy basic metric properties, for instance it can happen that $rand(F, F) < 1$ or that its value is negative
- hence they cannot be used to compare two clusterings

Soft Clustering evaluation

A part from rough set clustering, there exist some extension of standard measure, in particular different version of Rand index

- often they fail to satisfy basic metric properties, for instance it can happen that $rand(F, F) < 1$ or that its value is negative
- hence they cannot be used to compare two clusterings
- do not distinguish different form of uncertainty: ambiguity, fuzziness

Soft clustering as distributions over hard clusterings

- evidential clustering $\rightarrow$ rough clustering $\rightarrow$ hard clustering

Soft clustering as distributions over hard clusterings

- evidential clustering $\rightarrow$ rough clustering $\rightarrow$ hard clustering
- R rough clustering, $R(x)$ the clusters x belongs to (i.e., the upper approximation)

Soft clustering as distributions over hard clusterings

- evidential clustering → rough clustering → hard clustering
- R rough clustering, $R(x)$ the clusters x belongs to (i.e., the upper approximation)
- an evidential clustering is represented as a mass function over hard clusterings:

$$m_M(R) = \prod_x m_x(R(x))$$

Transport-based measures

A comparison measure for soft clustering obtained by

- computing the cost of making the two distributions equivalent by moving masses from one rough clustering to another

Transport-based measures

A comparison measure for soft clustering obtained by

- computing the cost of making the two distributions equivalent by moving masses from one rough clustering to another
- the cost of such movements is determined by a base distance over hard clusterings.
- any distance is ok $\rightarrow$ it is framework to generalize different hard clustering evaluations

Transport-based measures

An interval $[d_0, d_1]$ such that

Transport-based measures

An interval $[d_0, d_1]$ such that

- d_0 represent the compatibility between two clustering, i.e., whether there exists a hard clustering compatible with both soft clusterings
- d_1 quantifies their similarity

Transport-based measures

An interval $[d_0, d_1]$ such that

- d_0 represent the compatibility between two clustering, i.e., whether there exists a hard clustering compatible with both soft clusterings
- d_1 quantifies their similarity
- Pros: they satisfies good metric properties
- Cons: NP-hard to compute $\rightarrow$ approximate solutions with bounded error

Summary of the proposed comparison measures

Measure	Metric Properties	Computational Complexity
Transport-based Measure	$1 - d_0^E$ consistency d_1^E metric	d_0^E NP-HARD d_1^E NP-HARD
Sampling-based Approximations	-	$O(n^2 s + s^3 \log s)$
Approximation for Rand index	Rand_0 consistency Rand_1 similarity	$O(n^2)$
Approximation for partition distance	$1 - \delta_0^E$ consistency δ_1^E metric	$O(n2^k + k^3)$

A. Campagner, D. Ciucci, T. Denœux, A General Framework for Evaluating and Comparing Soft Clusterings, submitted to Information Sciences