Constrained Evidential Clustering

V. Antoine[1] B. Quost[2] M.-H. Masson[2] T. Denœux[2]

[1]Limos laboratory, UMR CNRS 6158, Clermont-Ferrand, France
[2]Heudiasyc laboratory, UMR CNRS 6599, Compiègne, France

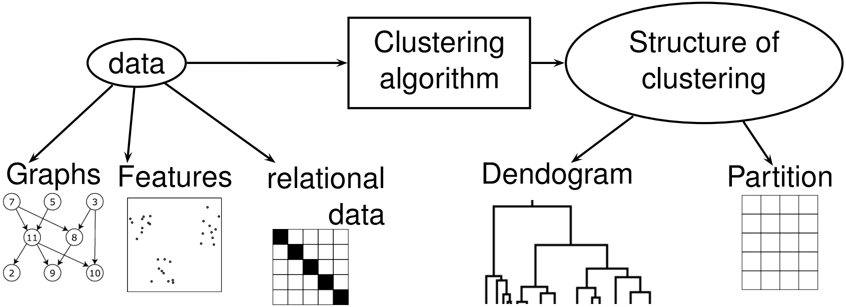
December 2013

Introduction

Clustering algorithm

Determine groups of N objects

- $\mathbf{x}_i \in \{\mathbf{x}_1 \ldots \mathbf{x}_N\}$ the set of objects with p attributes
- $\omega_k \in \Omega = \{\omega_1 \ldots \omega_c\}$ the set of clusters

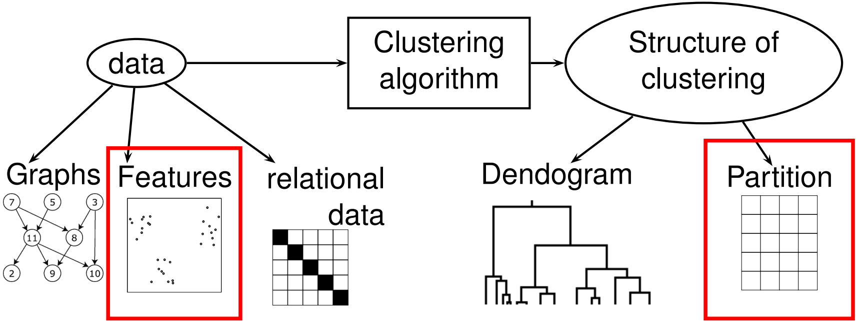


Introduction

Clustering algorithm

Determine groups of N objects

- $\mathbf{x}_i \in \{\mathbf{x}_1 \ldots \mathbf{x}_N\}$ the set of objects with p attributes
- $\omega_k \in \Omega = \{\omega_1 \ldots \omega_c\}$ the set of clusters



Introduction

Clustering methods

Group data objects into clusters based on a similarity notion

Problematic

No background knowledge

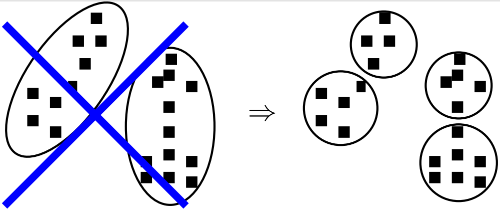
- How to define a similarity notion ?
- How to detect the expected classification ?

Introduction

Constrained clustering

Incorporating constraints into a clustering method

- Model level
  - balanced clusters
  - negative information : one model rejected
- Cluster level
- Instance level

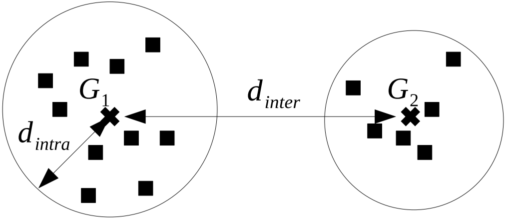


Introduction

Constrained clustering

Incorporating constraints into a clustering method

- Model level
  - balanced clusters
  - negative information : one model rejected
- Cluster level
- Instance level



Introduction

Constrained clustering

Incorporating constraints into a clustering method

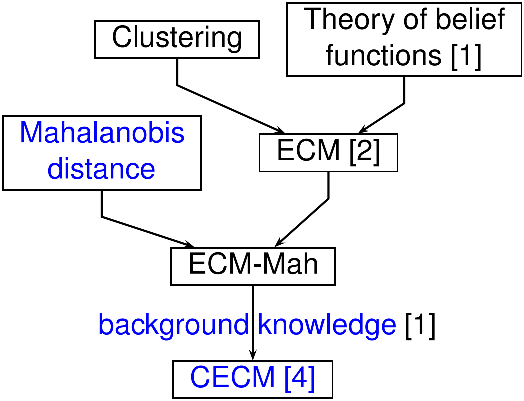
- Model level
  - balanced clusters
  - negative information : one model rejected
- Cluster level
- Instance level

Must-Link

Cannot-Link

- $\mathcal{M}$ Must-Link set of constraints
- $\mathcal{C}$ Cannot-Link set of constraints

Motivations



[1] P. Smets, *The transferable belief model for quantified belief representation*, 1998

[2] M.-H. Masson & al, *ECM : An evidential version of the fuzzy c-means algorithm*, 2008

[3] K. Wagstaff & al, *Constrained k-means clustering with background knowledge*, 2001

[4] V. Antoine & al, *CECM : Constrained Evidential C-Means algorithm*, 2012

Outline

- Background
  - Theory of belief functions
  - FCM and ECM
- Our contributions
  - Using an adaptive metric
  - Integrating constraints
  - Active learning
- Experiments
- Conclusion and Perspectives

Outline

- Background
  - Theory of belief functions
  - FCM and ECM
- Our contributions
  - Using an adaptive metric
  - Integrating constraints
  - Active learning
- Experiments
- Conclusion and Perspectives

Mass Function

Let Y be a variable taking values in a finite set Ω.

Mass function $m : 2^{\Omega} \to [0, 1]$

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

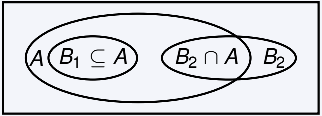
- $m(A)$: degree of belief specific to $Y \in A$
- If $m(A) > 0$ then A is a focal set

Derivative notions

Plausibility function

Potential degree of belief that *could be* given to A :

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad \forall A \subseteq \Omega$$



Making decision : the pignistic transformation

- Belief functions space $\longrightarrow$ probability space

$$BetP(\omega) = \frac{1}{1 - m(\emptyset)} \sum_{\{A \subseteq \Omega | \omega \in A\}} \frac{m(A)}{|A|}$$

Credal partition

Clustering framework

- $\Rightarrow$ Ω : set of clusters $\{\omega_1, \ldots, \omega_c\}$
- $\Rightarrow$ Y : actual class of the object $\mathbf{x}_i$
- $\Rightarrow$ $\mathbf{m}_i$: partial knowledge on the class of $\mathbf{x}_i$
- $\Rightarrow$ $\mathbf{M} = (\mathbf{m}_i)$: credal partition

Exemple

A	m_1	m_2	m_3	m_4
$\emptyset$	0	0	0	1
$\{\omega_1\}$	1	0.3	0	0
$\{\omega_2\}$	0	0.7	0	0
$\{\omega_1, \omega_2\}$	0	0	1	0

Evidential algorithms

- model with features : **ECM**
- relational model : EVCLUS, RECM

Fuzzy *c*-means (FCM)

Geometrical model

- Each object x_i has a degree of membership in each cluster k : u_{ik}
- Each cluster ω_k is represented by a center v_k
- Distance
  - Euclidean $d_{ik}^2 = \|\mathbf{x}_i - \mathbf{v}_k\|^2$
  - Mahalanobis, Gustafson and Kessel method : $d_{ik}^2 = (\mathbf{x}_i - \mathbf{v}_k)^\top \Sigma_k (\mathbf{x}_i - \mathbf{v}_k)$

Aternate optimization

$$opt(u_{ik}) \rightleftarrows opt(\mathbf{v}_k)$$

Objective function

$$J_{FCM} = \sum_{i=1}^{N} \sum_{k=1}^{C} u_{ik}^{\beta} d_{ik}^2$$

Subject to

$$\sum_{k=1}^{C} u_{ik} = 1 \text{ and } u_{ik} \geq 0 \quad \forall i, k$$

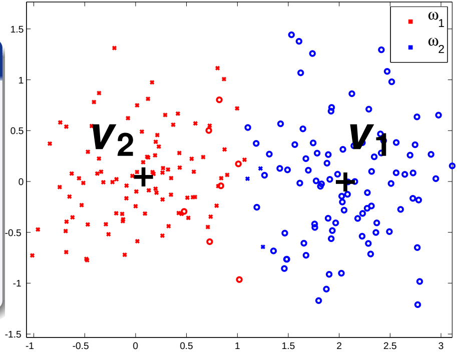
ECM

Principle

- Generalization of fuzzy c-means
- goal : enhance the concept of partition by using a credal partition

Geometrical model

- Each cluster ω_k is represented by a center v_k
- Centroid $\overline{v}_j$: barycenter of centers associated to classes composing $A_j \subseteq \Omega$
- Distance d_{ij}^2 between x_i and $\overline{v}_j$



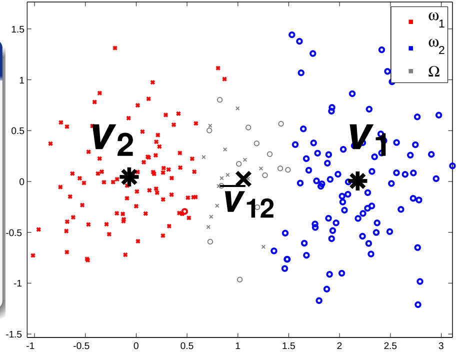
ECM

Principle

- Generalization of fuzzy *c*-means
- goal : enhance the concept of partition by using a credal partition

Geometrical model

- Each cluster ω_k is represented by a center v_k
- Centroid $\overline{v}_j$: barycenter of centers associated to classes composing $A_j \subseteq \Omega$
- Distance d_{ij}^2 between x_i and $\overline{v}_j$



ECM

Objective function

$$J_{ECM} = \sum_{i=1}^{N} \sum_{A_j \subseteq \Omega,\ A_j \neq \emptyset} |A_j|^{\alpha} m_i(A_j)^{\beta} d_{ij}^2 + \sum_{i=1}^{N} \delta^2 m_i(\emptyset)^{\beta}$$

$$\text{subject to : } \begin{cases} \sum_{A_j \subseteq \Omega,\ A_j \neq \emptyset} m_i(A_j) + m_i(\emptyset) = 1 \\ m_i(A_j) \geq 0 \quad \forall i, j \end{cases}$$

Optimisation

Minimize J_{ECM} w.r.t m_{ij}, $\mathbf{v}_k$

$\Rightarrow$ Use of the Lagrangian multipliers

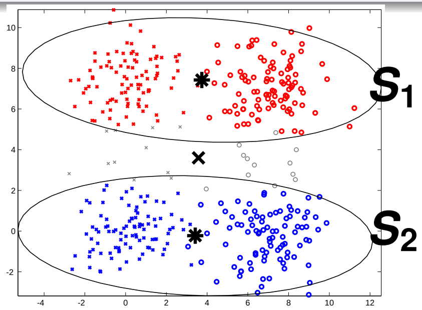
Outline

- Background
  - Theory of belief functions
  - FCM and ECM
- Our contributions
  - Using an adaptive metric
  - Integrating constraints
  - Active learning
- Experiments
- Conclusion and Perspectives

Using an adaptive metric

Mahalanobis distance for each class ω_k

- Each cluster ω_k is represented by a center $\mathbf{v_k}$
- Each cluster ω_k has a covariance matrix $\mathbf{S_k}$



Definition

$$d_{ij}^2 = (\mathbf{x}_i - \overline{\mathbf{v}}_j)^t \overline{\mathbf{S}}_j (\mathbf{x}_i - \overline{\mathbf{v}}_j)$$

such that

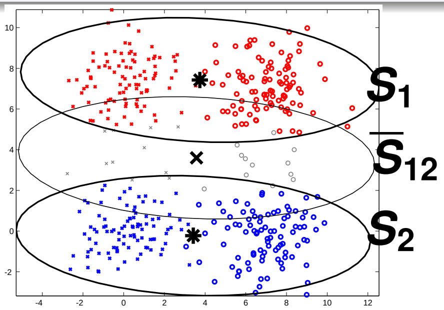
$$\overline{\mathbf{S}}_j = \frac{1}{|A_j|} \sum_{\omega_k \in A_j} \mathbf{S}_k,$$

$$\forall A_j \subseteq \Omega, A_j \neq \emptyset$$

Using an adaptive metric

Mahalanobis distance for each class ω_k

- Each cluster ω_k is represented by a center $\mathbf{v_k}$
- Each cluster ω_k has a covariance matrix $\mathbf{S_k}$



Definition

$$d_{ij}^2 = (\mathbf{x}_i - \overline{\mathbf{v}}_j)^t \overline{\mathbf{S}}_j (\mathbf{x}_i - \overline{\mathbf{v}}_j)$$

such that

$$\overline{\mathbf{S}}_j = \frac{1}{|A_j|} \sum_{\omega_k \in A_j} \mathbf{S}_k,$$

$$\forall A_j \subseteq \Omega, A_j \neq \emptyset$$

Using an adaptive metric

New Objective function

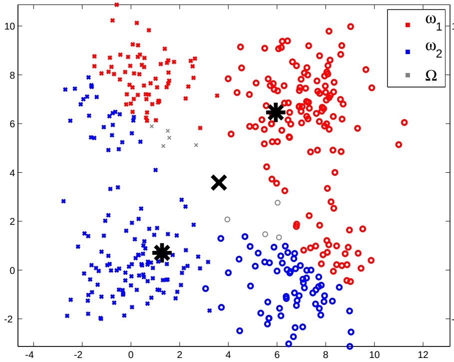
Minimize J_{ECM} w.r.t m_{ij}, $\mathbf{v}_k$, $\mathbf{S}_k$ s.t. $|\mathbf{S}_k| = 1 \quad \forall k = 1, C$

Optimisation

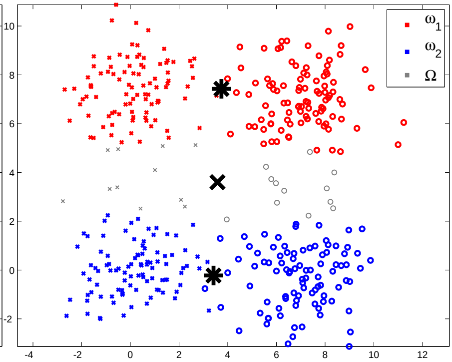
Kuhn–Tucker conditions give :

- $m_i(A_j)$ identical to ECM with a Euclidean distance
- $\mathbf{v}_k$: system of linear equations
- $\mathbf{S}_k$: similar to Gustafson et Kessel

Using an adaptive metric



ECM+Euclidean distance



ECM+Mahalanobis distance

Adding constraints in ECM

Formalization

- Joint class membership for $\mathbf{x}_i$, $\mathbf{x}_j$

$$m_{i\times j}(A \times B) = m_i(A)m_j(B) \quad \forall A, B \subseteq \Omega, A \neq \emptyset, B \neq \emptyset$$

- In Ω^2, two events
  - $\theta \Rightarrow$ "$\mathbf{x}_i$ and $\mathbf{x}_j$ belong to the same class"
  - $\overline{\theta} \Rightarrow$ "$\mathbf{x}_i$ and $\mathbf{x}_j$ do not belong to the same class"

$\Rightarrow$ Plausibility to belong to the same class

$$pl_{i\times j}(\theta) = \sum_{A\cap B\neq\emptyset} m_i(A)\, m_j(B)$$

$\Rightarrow$ Plausibility to belong to a different class

$$pl_{i\times j}(\overline{\theta}) = 1 - m_{i\times j}(\emptyset) - \sum_{k=1\ldots c} m_i(\{\omega_k\})m_j(\{\omega_k\})$$

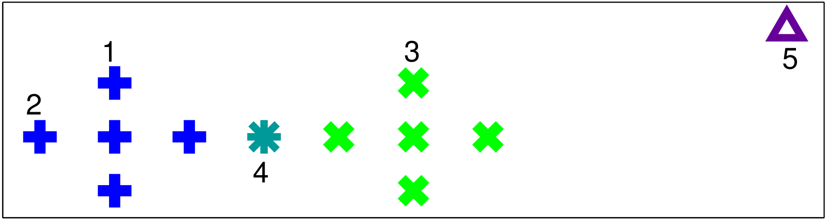
Adding constraints in ECM

Example

A	m_1	m_2	m_3	m_4	m_5
$\emptyset$	0	0	0	0	1
ω_1	1	1	0	0	0
ω_2	0	0	1	0	0
Ω	0	0	0	1	0

$\Rightarrow$

	$pl_{1\times 2}$	$pl_{1\times 3}$	$pl_{1\times 4}$	$pl_{1\times 5}$
θ	1	0	1	0
$\overline{\theta}$	0	1	1	0



Adding constraints in ECM

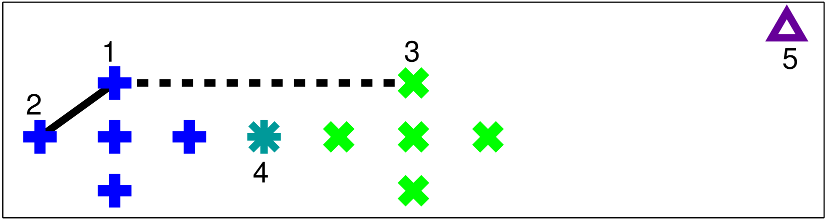
Example

A	m_1	m_2	m_3	m_4	m_5
$\emptyset$	0	0	0	0	1
ω_1	1	1	0	0	0
ω_2	0	0	1	0	0
Ω	0	0	0	1	0

	$pl_{1\times 2}$	$pl_{1\times 3}$	$pl_{1\times 4}$	$pl_{1\times 5}$
θ	1	0	1	0
$\overline{\theta}$	0	1	1	0

$\Updownarrow$

$(o_1, o_2) \in \mathcal{M}$ $(o_1, o_3) \in \mathcal{C}$



Adding constraints in ECM

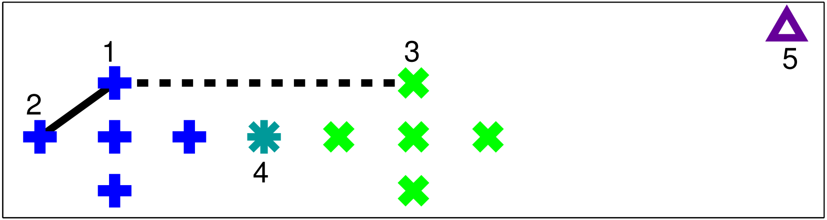
Example

A	m_1	m_2	m_3	m_4	m_5
$\emptyset$	0	0	0	0	1
ω_1	1	1	0	0	0
ω_2	0	0	1	0	0
Ω	0	0	0	1	0

	$pl_{1\times 2}$	$pl_{1\times 3}$	$pl_{1\times 4}$	~~$pl_{1\times 5}$~~
θ	1	0	1	~~0~~
$\overline{\theta}$	0	1	1	~~0~~

$\Updownarrow$

$(o_1, o_2) \in \mathcal{M}$ $\quad (o_1, o_3) \in \mathcal{C}$



Evidential clustering with constraints : CECM

Basic idea

If $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{M} \Rightarrow pl_{i\times j}(\overline{\theta})$ low and if $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{C} \Rightarrow pl_{i\times j}(\theta)$ low

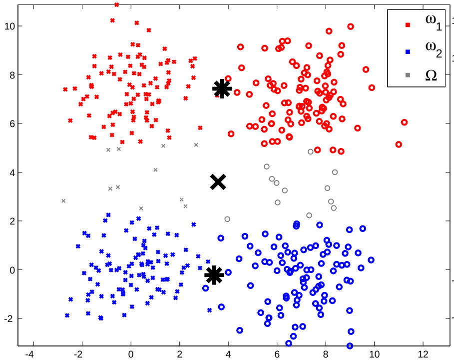
Objective function

$$J_{CECM} = (1-\varphi)\Big(\sum_{i=1}^{N} \sum_{A_j\subseteq\Omega,\ A_j\neq\emptyset} |A_j|^\alpha m_i(A_j)^\beta d_{ij}^2 + \sum_{i=1}^{N} \delta^2 m_i(\emptyset)^\beta\Big) + \varphi\Big(\sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{M}} pl_{i\times j}(\overline{\theta}) + \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{C}} pl_{i\times j}(\theta)\Big)$$

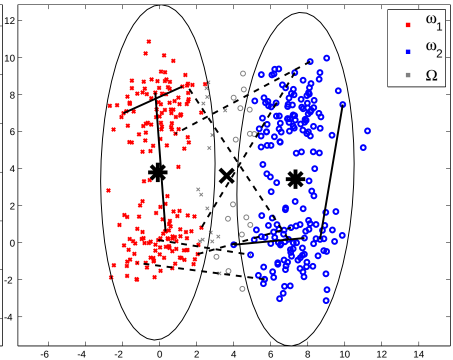
subject to $\sum m_i(A_j) + m_i(\emptyset) = 1$ and $m_i(A_j) \geq 0 \quad \forall i, j$

$\Rightarrow$ J_{CECM} quadratic when $\beta = 2$, constraints are linear

Evidential clustering with constraints : CECM



ECM+Mahalanobis distance

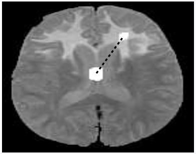


CECM+Mahalanobis distance

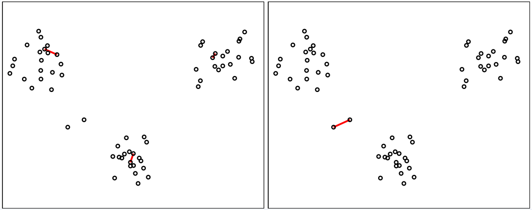
Problematic

How to retrieve a constraints set ?

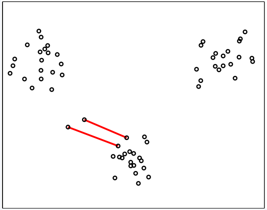
- Background knowledge
  - $\Rightarrow$ wide number of constraints, not necessarely interesting ones
- Expert
  - $\Rightarrow$ time consuming



Uninformative constraints



Redundancy



* image provided by Prof. Adamsbaum (AP-HP hospital, Paris) and Prof. Bloch (ENST, Paris)

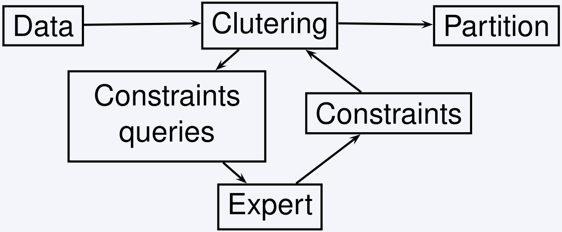
Active learning

Automatic and smart selection of constraints

Goal :

- acquire a few number of constraints at low cost
- greatly improve the clustering result

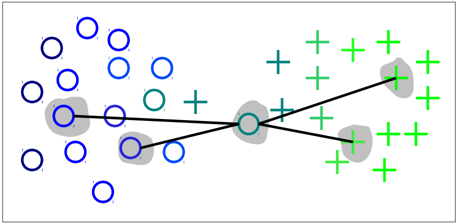
Global scheme



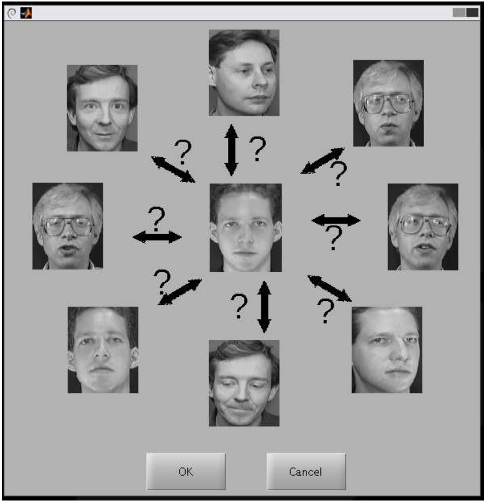
Active learning

Proposed method

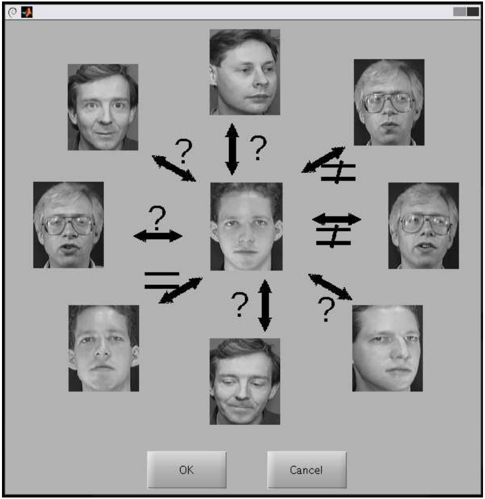
1. Select an object classified with uncertainty
2. Select one or several objects classified with certainty
3. Ask the link to an expert



Active learning : possible application



Active learning : possible application



Outline

- Background
  - Theory of belief functions
  - FCM and ECM
- Our contributions
  - Using an adaptive metric
  - Integrating constraints
  - Active learning
- Experiments
- Conclusion and Perspectives

Experiments

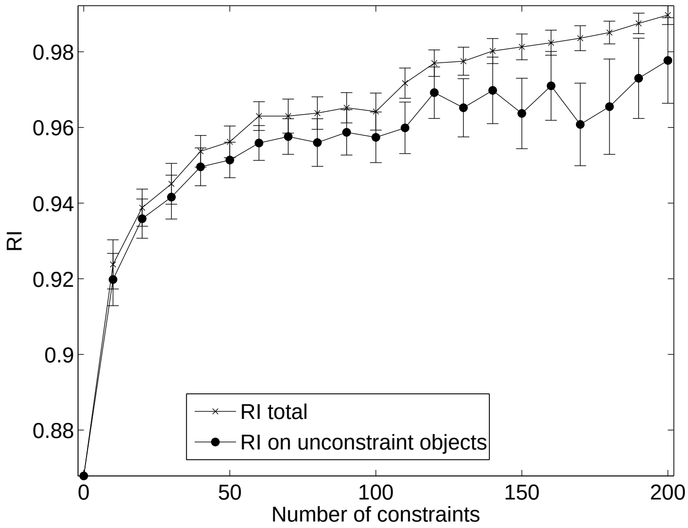
Datasets

	# attributes	# objects	# classes
Iris	4	150	3
LettersIJL	16	227	3

Evaluation method

- Random constraints selection : based on the true known classes
- Decision : Maximum of pignistic probability
- Criterion : Rand Index

Behavior of CECM on Iris dataset

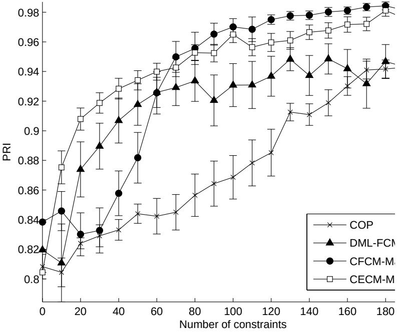


Comparison with other algorithms

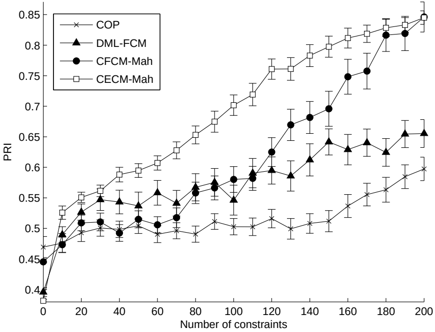
	Clustering base	Constraints respected	Distance modification
COP [1]	k-Means	Yes	
DML-FCM [2]	k-Means	/	X
CFCM [3]	FCM	/	X
CECM	ECM	/	X

- [1] K. Wagstaff & al, *Constrained k-means clustering with background knowledge*, KDID, 2001
- [2] E. Xing & al, *Distance Metric Learning with application to clustering with side-information*, 2002
- [3] N. Grira & al, *Active semi-supervised fuzzy clustering*, Pattern Recognition, 2008

Comparison with other algorithms



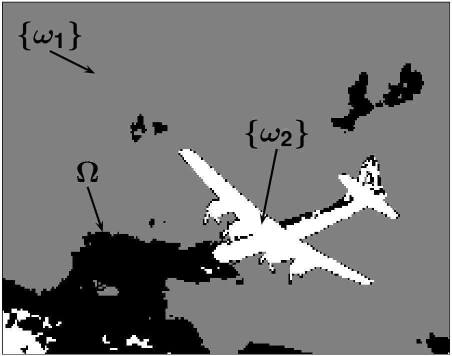
Iris



LettersIJL

Plane image

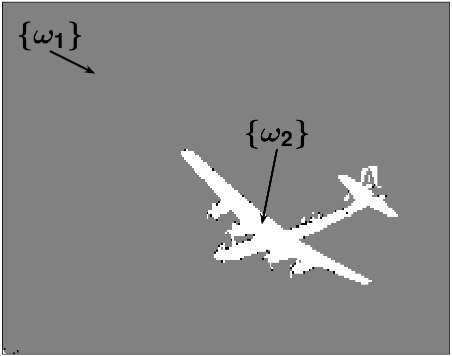
Original image



ECM+Euclidean distance

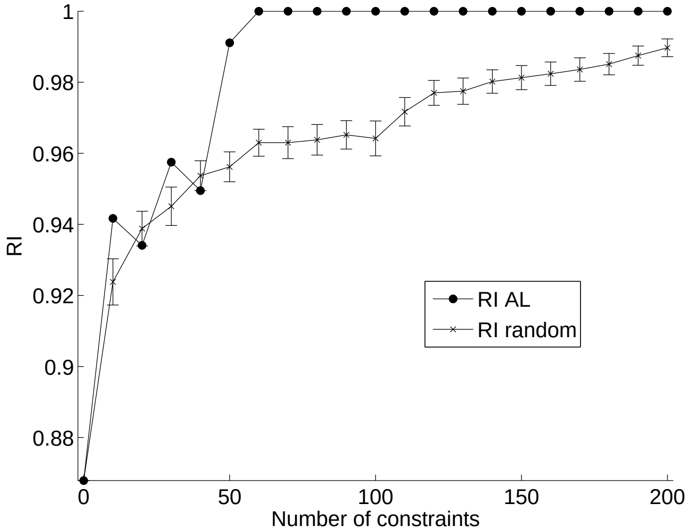
Plane image

constraints selection



CECM+Mahalanobis distance

Active learning on CECM for Iris dataset



Outline

Conclusion

Background knowledge in clustering algorithm

clustering
belief function theory } → ECM

adaptive metric & background knowledge } → CECM

Benefits

- constraints lead to the desired solution
- Improved performance

Problems

- Computational complexity
- sensitivity to constraints selection

Active learning

New method based on belief function theory

Benefits

- improve clustering result
- reduce the execution time

Problems

- bad selection of constraints

Perspectives

Evidential clustering

- lower the computational complexity of ECM
  - $\rightarrow$ decrease the number of existing subsets
- improve the use of a Mahalanobis distance in ECM
  - $\rightarrow$ new computation of the centers

Constrained clustering

- new constrained clustering methods
- studying active learning methods
  - $\rightarrow$ guideline
  - $\rightarrow$ new measures of utility for constraints
  - $\rightarrow$ new algorithms

Thank you for your attention