

Contributions à l'Apprentissage Collaboratif non Supervisé

MOHAMAD GHASSANY

Directeur: Younès Bennani

LIPN, UMR CNRS 7030
École Doctorale Galilée
Université Paris 13, Sorbonne Paris Cité

Soutenance de thèse

7 novembre 2013



Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

1 Contexte

2 Clustering Collaboratif à base de Prototypes

3 Collaboration sous cadre Déterministe

4 Collaboration sous cadre Probabiliste

5 Collaboration sous cadre Bayésien

6 Bilan & Perspectives

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

1 Contexte

2 Clustering Collaboratif à base de Prototypes

3 Collaboration sous cadre Déterministe

4 Collaboration sous cadre Probabiliste

5 Collaboration sous cadre Bayésien

6 Bilan & Perspectives

Contexte

Clustering
Collaboratif
à base de
Prototypes

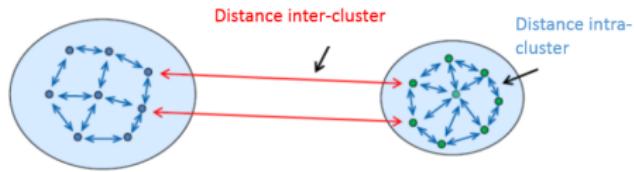
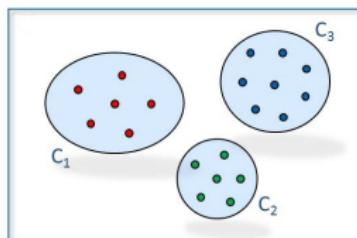
Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

- soit $\mathbf{X} = \{x_i \in \mathbb{R}^D\}_{i=1,\dots,N}$ un ensemble d'observations décrit par D attributs.
- Objectif du clustering : **structuration des données en classes homogènes**. On cherche à regrouper les points en **clusters** ou classes tels que les **données d'un cluster** soient les plus **similaires** possibles.
- Clustering : c'est de **l'apprentissage non supervisé**. Les données ne sont pas étiquetées et le nombre de classes n'est pas connu.



Deux types de Clustering

Contexte

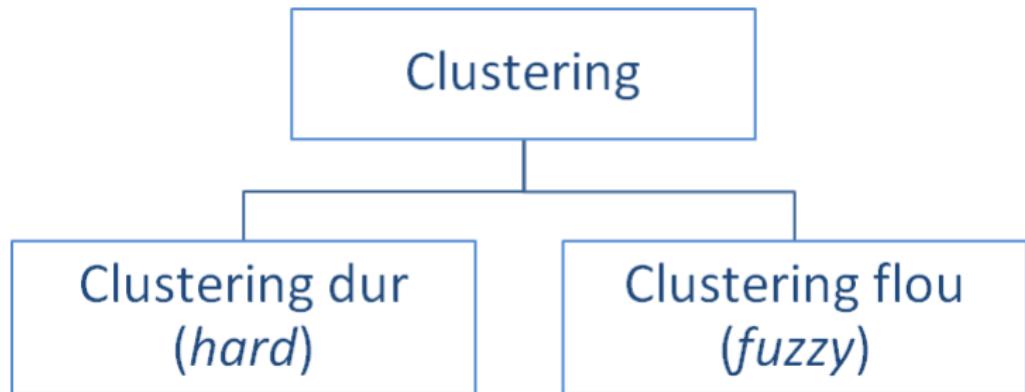
Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives



chaque élément appartient à
un seul cluster

Approches Déterministes

chaque élément appartient à chacun
des clusters selon un degré
d'appartenance

Approches Probabilistes

Contexte

Clustering
Collaboratif
à base de
Prototypes

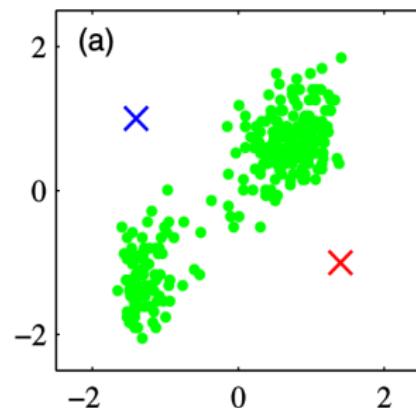
Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Déterministe : Kmeans



choisir k prototypes aléatoires

Contexte

Clustering
Collaboratif
à base de
Prototypes

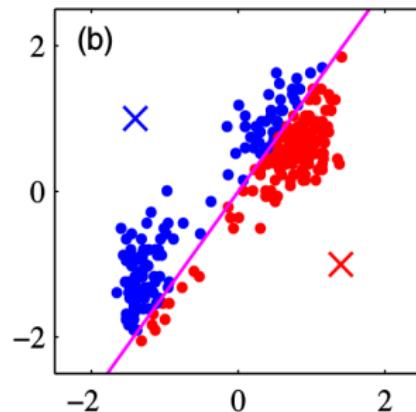
Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Déterministe : Kmeans



attribuer chaque observation au
prototype le plus proche

Contexte

Clustering
Collaboratif
à base de
Prototypes

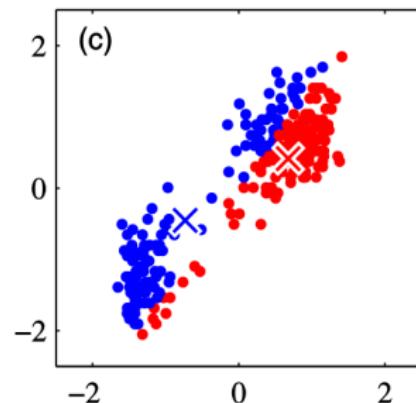
Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Déterministe : Kmeans



mettre à jour les prototypes

Contexte

Clustering
Collaboratif
à base de
Prototypes

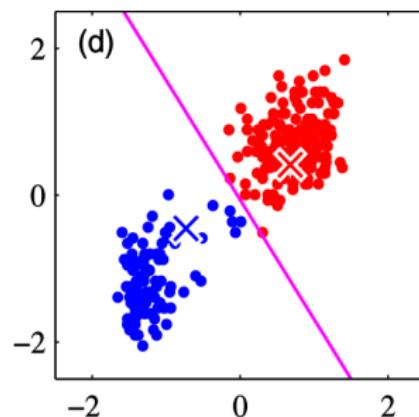
Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Déterministe : Kmeans



Contexte

Clustering
Collaboratif
à base de
Prototypes

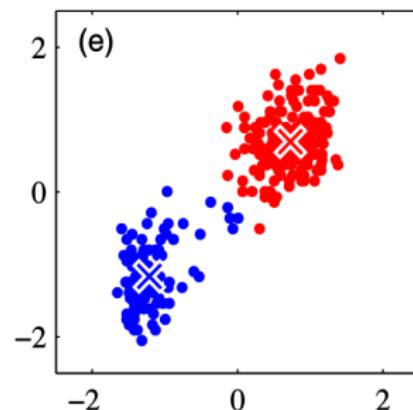
Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Déterministe : Kmeans



Contexte

Clustering
Collaboratif
à base de
Prototypes

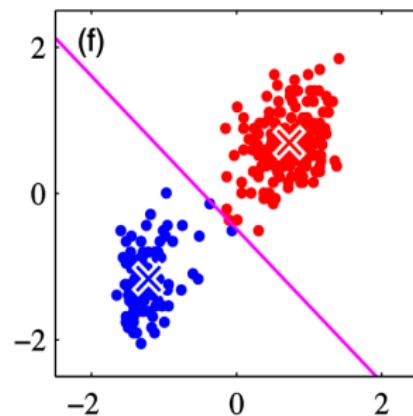
Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Déterministe : Kmeans



Contexte

Clustering
Collaboratif
à base de
Prototypes

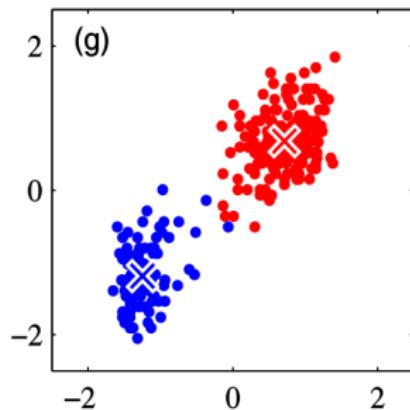
Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

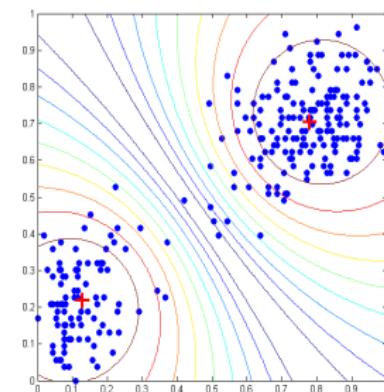
Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Déterministe : Kmeans



Probabiliste : FCM



Contexte

Clustering
Collaboratif
à base de
Prototypes

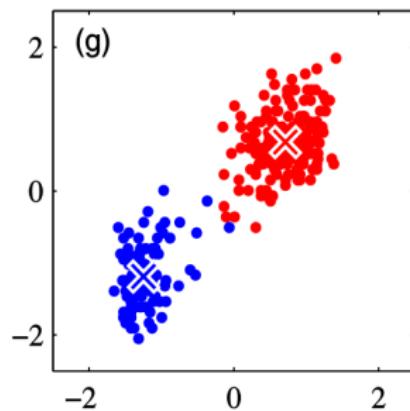
Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

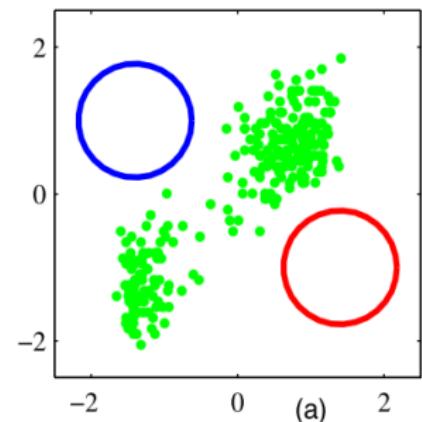
Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Déterministe : Kmeans



Probabiliste : EM



Contexte

Clustering
Collaboratif
à base de
Prototypes

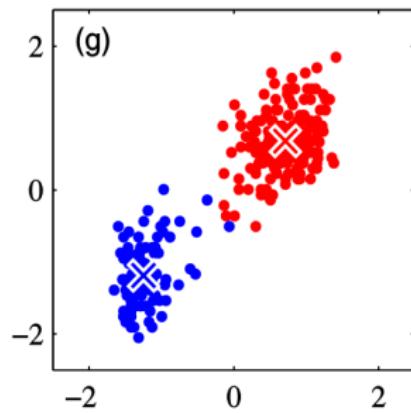
Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

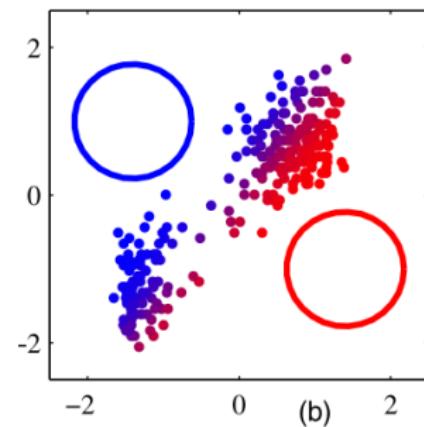
Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Déterministe : Kmeans



Probabiliste : EM



Contexte

Clustering
Collaboratif
à base de
Prototypes

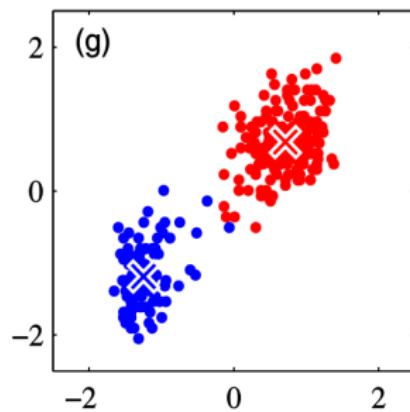
Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

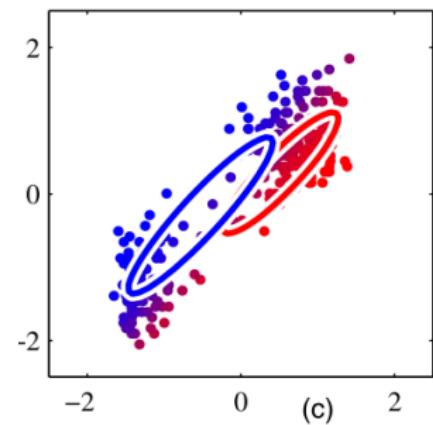
Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Déterministe : Kmeans



Probabiliste : EM



Contexte

Clustering
Collaboratif
à base de
Prototypes

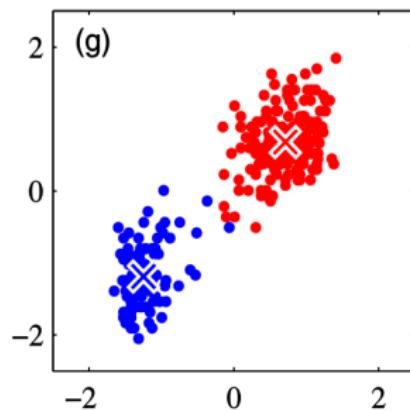
Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

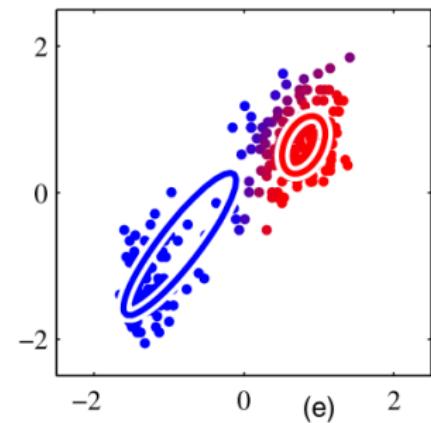
Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Déterministe : Kmeans



Probabiliste : EM



Contexte

Clustering
Collaboratif
à base de
Prototypes

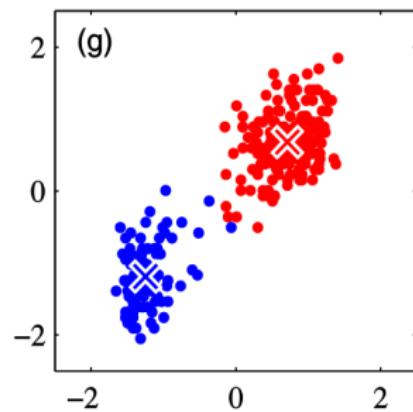
Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

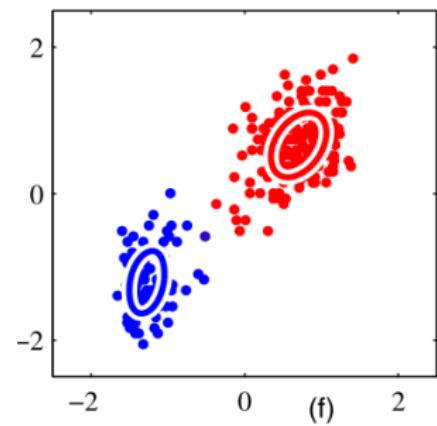
Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Déterministe : Kmeans



Probabiliste : EM



Contexte

*Clustering**Collaboratif**à base de**Prototypes**Collaboration**sous cadre**Déterministe**Collaboration**sous cadre**Probabiliste**Collaboration**sous cadre**Bayésien**Bilan &**Perspectives*

Et si nous traitons des données distribuées sur P sites ?

 $D[ii]$  $D[ij]$  $D[kk]$

...

 $D[P]$

Contexte

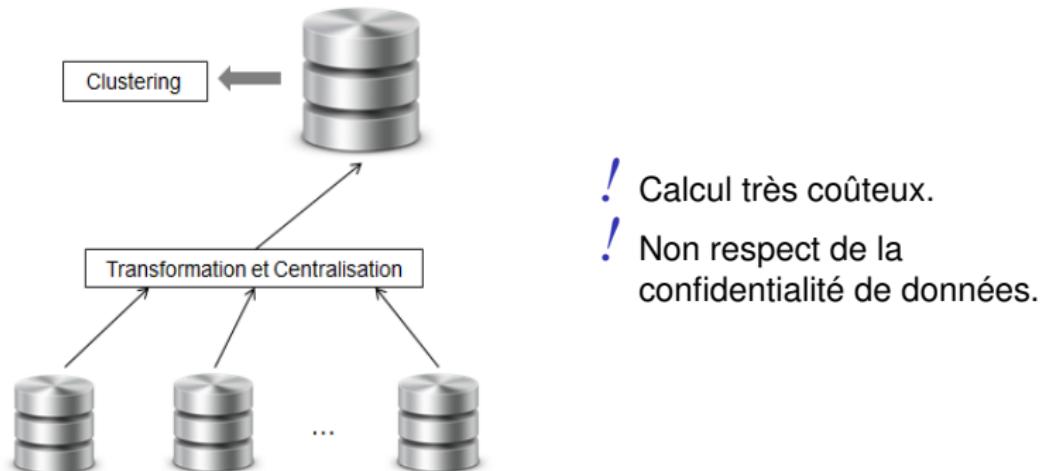
Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives



Techniques standards

Contexte

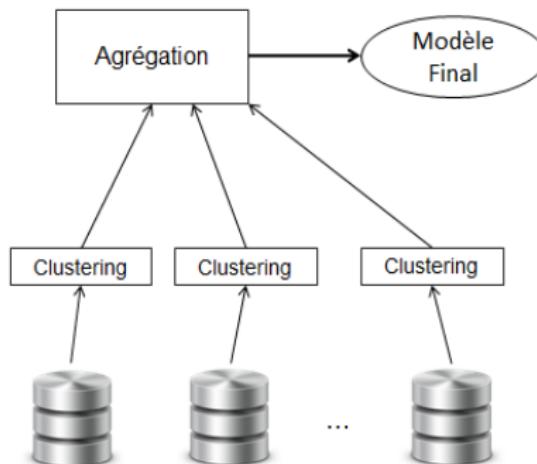
Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives



- ! Ne prend pas en compte la qualité du clustering local.
- ! Abscence de potentielle amélioration locale.

Contexte

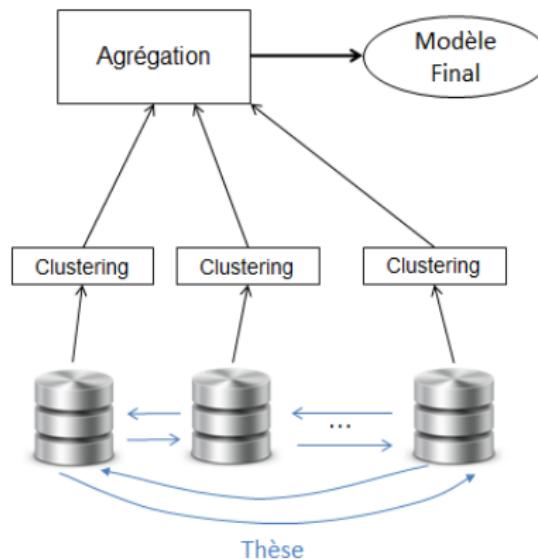
Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives



Techniques standards

Contexte

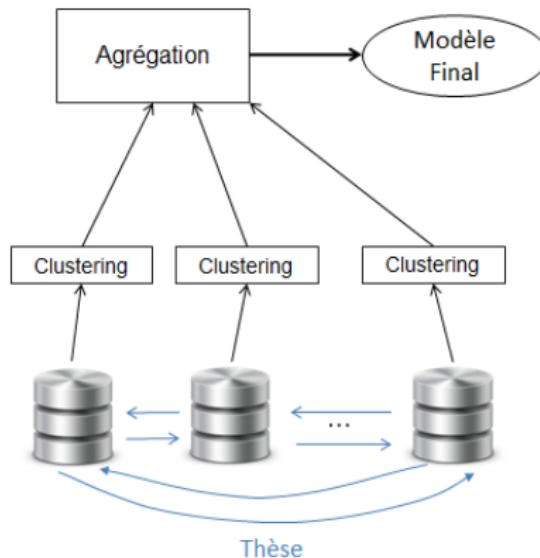
Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives



But : Enrichir les classifications locales à partir des classifications distantes.

Hypothèse fondamentale : les données sont i.i.d. & issues de la même problématique.

Plan de la présentation

Contexte

Clustering
Collaboratif
à base de
Prototypes

Approche
verticale

Approche
horizontale

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

1 Contexte

2 Clustering Collaboratif à base de Prototypes

3 Collaboration sous cadre Déterministe

4 Collaboration sous cadre Probabiliste

5 Collaboration sous cadre Bayésien

6 Bilan & Perspectives

Clustering Collaboratif

Contexte

Clustering
Collaboratif
à base de
Prototypes

Approche
verticale

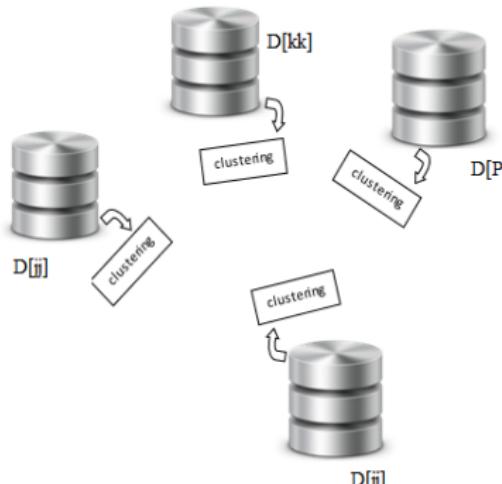
Approche
horizontale

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives



- Phase locale : Appliquer les algorithmes de clustering localement sur les différents sites.

Clustering Collaboratif

Contexte

Clustering
Collaboratif
à base de
Prototypes

Approche
verticale

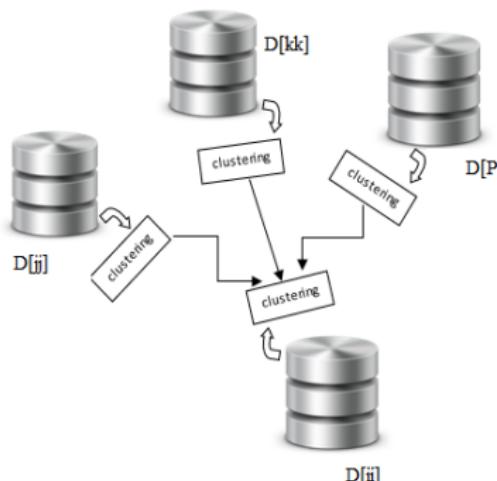
Approche
horizontale

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives



- Phase **locale** : Appliquer les algorithmes de clustering localement sur les différents sites.
- Phase **globale** : Collaborer les sites en partageant les résultats obtenus lors de la phase locale.

Clustering Collaboratif

Contexte

Clustering
Collaboratif
à base de
Prototypes

Approche
verticale

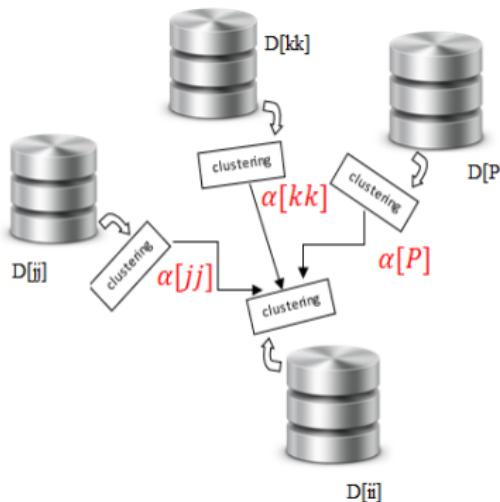
Approche
horizontale

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives



- Phase locale : Appliquer les algorithmes de clustering localement sur les différents sites.
- Phase globale : Collaborer les sites en partageant les résultats obtenus lors de la phase locale.
- La collaboration est contrôlée par un coefficient, i.e. confiance.

Coefficients de Collaboration

Contexte

Clustering
Collaboratif
à base de
Prototypes

Approche
verticale

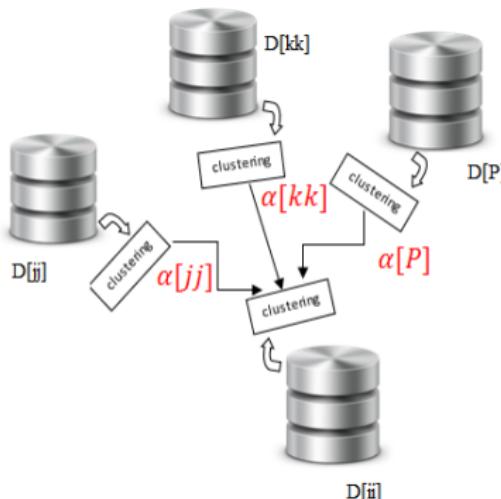
Approche
horizontale

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives



- Coefficients ≥ 0 .
- Coefficient élevé \implies Collaboration forte.
- Ne sont pas symétriques.
- Ce coefficient sera **calculé automatiquement** [Ghassany11].

Contexte

*Clustering
Collaboratif
à base de
Prototypes*

*Approche
verticale*

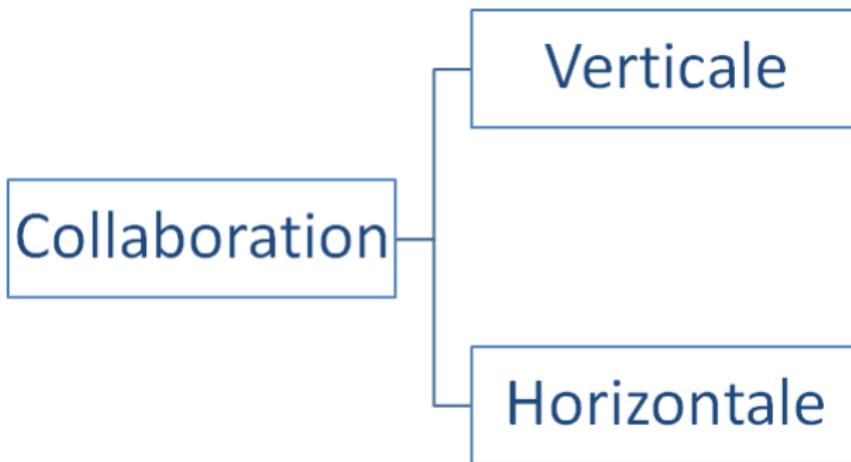
*Approche
horizontale*

*Collaboration
sous cadre
Déterministe*

*Collaboration
sous cadre
Probabiliste*

*Collaboration
sous cadre
Bayésien*

*Bilan &
Perspectives*



Approche verticale

Contexte

Clustering
Collaboratif
à base de
Prototypes

Approche
verticale

Approche
horizontale

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

ID	Att 1	Att 2	Att 3	Att 4
Id1				Base de données [1]
Id2				
Id3				
Id4				

ID	Att 1	Att 2	Att 3	Att 4
Id5				Base de données [2]
Id6				
Id7				

ID	Att 1	Att 2	Att 3	Att 4
Id8				Base de données [P]
Id9				

Mêmes
attributs

Exemple : Plusieurs supermarchés analysent le comportement de leurs clients, selon les mêmes critères. Les clients ne sont pas les mêmes.

Approche horizontale

Contexte
Clustering
Collaboratif
à base de
Prototypes

Approche
verticale

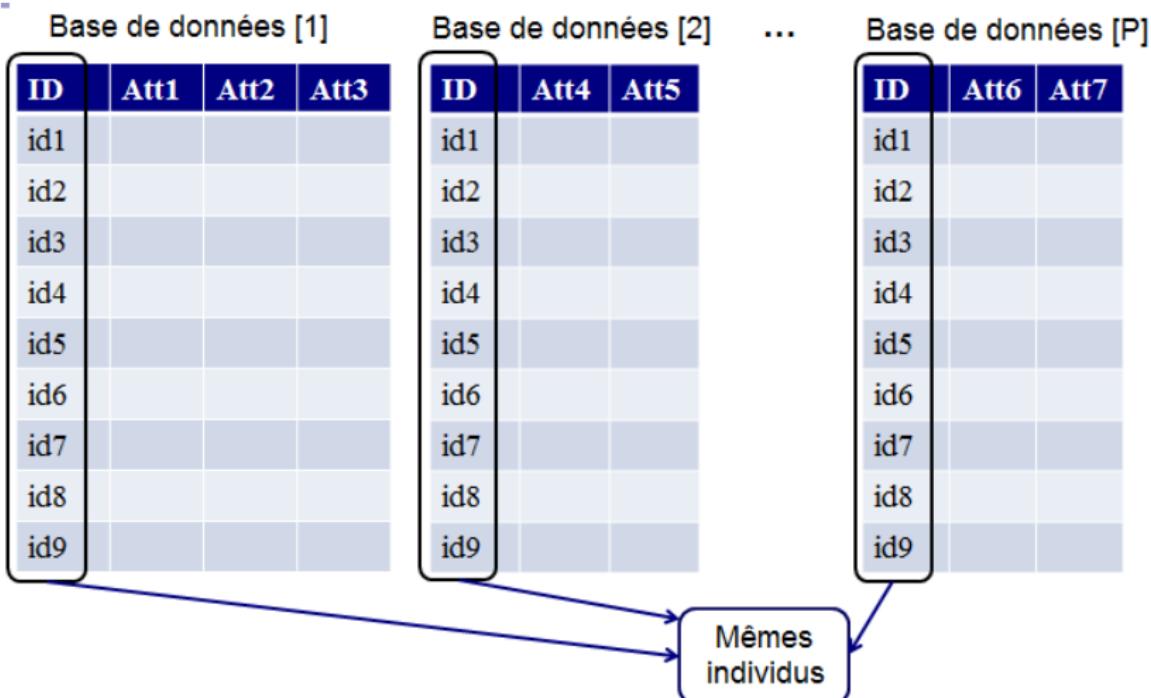
Approche
horizontale

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives



Exemple : Les mêmes personnes sont clients de différents organismes.

Plan de la présentation

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

SOM

Collaboration à
base de SOM

Estimation de
la Confiance

Expérimentation

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

1 Contexte

2 Clustering Collaboratif à base de Prototypes

3 Collaboration sous cadre Déterministe

4 Collaboration sous cadre Probabiliste

5 Collaboration sous cadre Bayésien

6 Bilan & Perspectives

*Contexte**Clustering**Collaboratif
à base de
Prototypes**Collaboration
sous cadre
Déterministe**SOM**Collaboration à
base de SOM**Estimation de
la Confiance**Expérimentation**Collaboration
sous cadre
Probabiliste**Collaboration
sous cadre
Bayésien**Bilan &
Perspectives*

Objectif

Clustering & Visualisation.

Espaces

Principe

Les données vectorielles sont projetées sur une carte de neurones de dimension réduite en conservant la topologie de l'espace d'observation.

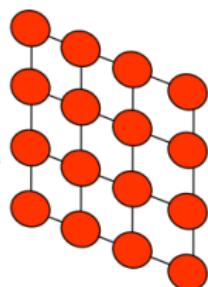
- définition a priori d'une notion de voisinage entre classes.
- des observations voisines dans l'espace de données appartiennent après classement à la même classe ou à des classes voisines.

couche d'entrée



connexions totales
(poids W)

couche de sortie = la carte



Contexte

*Clustering
Collaboratif
à base de
Prototypes*

*Collaboration
sous cadre
Déterministe*

SOM

*Collaboration à
base de SOM*

*Estimation de
la Confiance*

Expérimentation

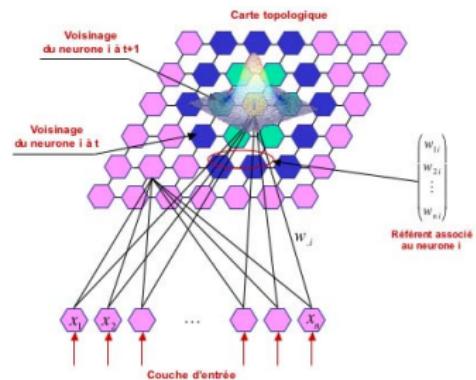
*Collaboration
sous cadre
Probabiliste*

*Collaboration
sous cadre
Bayésien*

*Bilan &
Perspectives*

- Fonction de coût

$$R(\chi, \mathcal{W}) = \sum_{i=1}^N \sum_{j=1}^C \mathcal{K}_{(j, \chi(x_i))} \|x_i - \mathbf{w}_j\|^2$$



Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

SOM

Collaboration à
base de SOMEstimation de
la Confiance

Expérimentation

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

• Fonction de coût

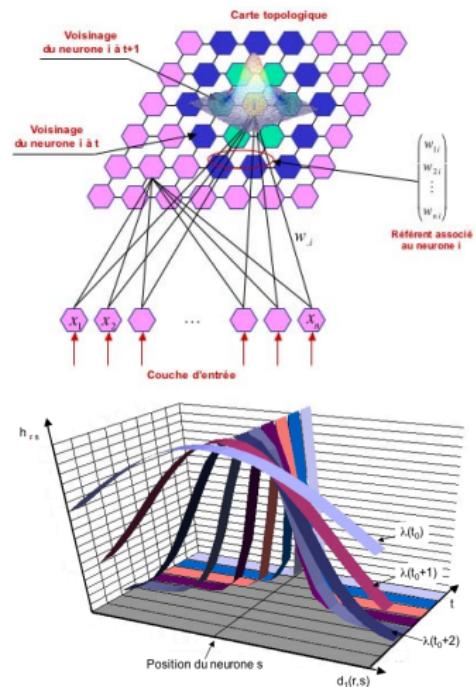
$$R(\chi, \mathcal{W}) = \sum_{i=1}^N \sum_{j=1}^C \mathcal{K}_{(j, \chi(x_i))} \|x_i - \mathbf{w}_j\|^2$$

• Fonction de voisinage

$$\mathcal{K}_{(j, \chi(x_i))} = \frac{1}{\lambda(t)} \exp \left(-\frac{d_1^2(i, j)}{\lambda^2(t)} \right)$$

• Fonction d'affectation

$$\chi(x_i) = \arg \min_j (\|x_i - \mathbf{w}_j\|^2)$$



Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

SOM
Collaboration à
base de SOM

Estimation de
la Confiance

Expérimentation

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Après avoir calculé les prototypes localement

Fonction de coût locale

$$R_V^{[ii]}(\chi, w) = \alpha_{[ii]}^{[jj]} \sum_{i=1}^{N[ii]} \sum_{j=1}^{|w|} \mathcal{K}_{\sigma(j, \chi(x_i))}^{[ii]} \|x_i^{[ii]} - w_j^{[ii]}\|^2$$

*Contexte**Clustering Collaboratif à base de Prototypes**Collaboration sous cadre Déterministe**SOM
Collaboration à base de SOM**Estimation de la Confiance**Expérimentation**Collaboration sous cadre Probabiliste**Collaboration sous cadre Bayésien**Bilan & Perspectives*

Après avoir calculé les prototypes localement

Fonction de coût locale

$$R_V^{[ii]} (\chi, w) = \alpha_{[ii]}^{[jj]} \sum_{i=1}^{N^{[ii]}} \sum_{j=1}^{|w|} \mathcal{K}_{\sigma(j, \chi(x_i))}^{[ii]} \|x_i^{[ii]} - w_j^{[ii]}\|^2$$

$$+ (\alpha_{[ii]}^{[jj]})^2 \sum_{i=1}^{N^{[ii]}} \sum_{j=1}^{|w|} \left(\mathcal{K}_{\sigma(j, \chi(x_i))}^{[ii]} - \mathcal{K}_{\sigma(j, \chi(x_i))}^{[jj]} \right)^2 \|w_j^{[ii]} - w_j^{[jj]}\|^2$$

Terme de Collaboration *Verticale*

*Contexte**Clustering**Collaboratif**à base de**Prototypes**Collaboration*
sous cadre
*Déterministe**SOM*
Collaboration à
*base de SOM**Estimation de*
*la Confiance**Expérimentation**Collaboration*
sous cadre
*Probabiliste**Collaboration*
sous cadre
*Bayésien**Bilan &*
Perspectives

Après avoir calculé les prototypes localement

Fonction de coût locale

$$R_V^{[ii]} (\chi, w) = \alpha_{[ii]}^{[jj]} \sum_{i=1}^{N^{[ii]}} \sum_{j=1}^{|w|} \mathcal{K}_{\sigma(j, \chi(x_i))}^{[ii]} \|x_i^{[ii]} - w_j^{[ii]}\|^2$$

$$+ (\alpha_{[ii]}^{[jj]})^2 \sum_{i=1}^{N^{[ii]}} \sum_{j=1}^{|w|} \left(\mathcal{K}_{\sigma(j, \chi(x_i))}^{[ii]} - \mathcal{K}_{\sigma(j, \chi(x_i))}^{[jj]} \right)^2 \|w_j^{[ii]} - w_j^{[jj]}\|^2$$

Terme de Collaboration *Verticale*

La collaboration se fait au niveau des prototypes

Fonction de coût locale

$$R_V^{[ii]}(\chi, w) = \alpha_{[ii]}^{[jj]} \sum_{i=1}^N \sum_{j=1}^{|w|} \mathcal{K}_{\sigma(j, \chi(x_i))}^{[ii]} \|x_i^{[ii]} - w_j^{[ii]}\|^2$$

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

SOM

Collaboration à
base de SOM

Estimation de
la Confiance

Expérimentation

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

*Contexte**Clustering Collaboratif à base de Prototypes**Collaboration sous cadre Déterministe**SOM**Collaboration à base de SOM**Estimation de la Confiance**Expérimentation**Collaboration sous cadre Probabiliste**Collaboration sous cadre Bayésien**Bilan & Perspectives*

Fonction de coût locale

$$R_V^{[ii]}(\chi, w) = \alpha_{[ii]}^{[jj]} \sum_{i=1}^N \sum_{j=1}^{|w|} \mathcal{K}_{\sigma(j, \chi(x_i))}^{[ii]} \|x_i^{[ii]} - w_j^{[ii]}\|^2$$

$$+ (\alpha_{[ii]}^{[jj]})^2 \sum_{i=1}^N \sum_{j=1}^{|w|} \left(\mathcal{K}_{\sigma(j, \chi(x_i))}^{[ii]} - \mathcal{K}_{\sigma(j, \chi(x_i))}^{[jj]} \right)^2 \|x_i^{[ii]} - w_j^{[ii]}\|^2$$

Terme de Collaboration Horizontale

Contexte

Clustering
Collaboratif
à base de
PrototypesCollaboration
sous cadre
Déterministe

SOM

Collaboration à
base de SOMEstimation de
la Confiance

Expérimentation

Collaboration
sous cadre
ProbabilisteCollaboration
sous cadre
BayésienBilan &
Perspectives

Fonction de coût locale

$$R_V^{[ii]}(\chi, w) = \alpha_{[ii]}^{[jj]} \sum_{i=1}^N \sum_{j=1}^{|w|} \mathcal{K}_{\sigma(j, \chi(x_i))}^{[ii]} \|x_i^{[ii]} - w_j^{[ii]}\|^2$$

$$+ (\alpha_{[ii]}^{[jj]})^2 \sum_{i=1}^N \sum_{j=1}^{|w|} \left(\mathcal{K}_{\sigma(j, \chi(x_i))}^{[ii]} - \mathcal{K}_{\sigma(j, \chi(x_i))}^{[jj]} \right)^2 \|x_i^{[ii]} - w_j^{[ii]}\|^2$$

Terme de Collaboration Horizontale

La collaboration se fait au niveau de **fond de voisinage**

Contexte

Clustering

Collaboratif
à base de
PrototypesCollaboration
sous cadre
Déterministe

SOM

Collaboration à
base de SOMEstimation de
la Confiance

Expérimentation

Collaboration
sous cadre
ProbabilisteCollaboration
sous cadre
BayésienBilan &
Perspectives

Fonction de coût locale

$$R_V^{[ii]}(\chi, w) = \alpha_{[ii]}^{[jj]} \sum_{i=1}^N \sum_{j=1}^{|w|} \mathcal{K}_{\sigma(j, \chi(x_i))}^{[ii]} \|x_i^{[ii]} - w_j^{[ii]}\|^2$$

$$+ (\alpha_{[ii]}^{[jj]})^2 \sum_{i=1}^N \sum_{j=1}^{|w|} \left(\mathcal{K}_{\sigma(j, \chi(x_i))}^{[ii]} - \mathcal{K}_{\sigma(j, \chi(x_i))}^{[jj]} \right)^2 \|x_i^{[ii]} - w_j^{[ii]}\|^2$$

Terme de Collaboration Horizontale

La collaboration se fait au niveau de **fond de voisinage**

! On ne peut pas calculer la distance entre les prototypes des différents sites.

*Contexte**Clustering**Collaboratif**à base de**Prototypes**Collaboration**sous cadre**Déterministe**SOM**Collaboration à
base de SOM**Estimation de
la Confiance**Expérimentation**Collaboration**sous cadre**Probabiliste**Collaboration**sous cadre**Bayésien**Bilan &**Perspectives*

$$\begin{aligned}
 R_V^{[ii]}(\chi, w) = & \alpha_{[ii]}^{[jj]} \sum_{i=1}^N \sum_{j=1}^{|w|} \mathcal{K}_{\sigma(j, \chi(x_i))}^{[ii]} \|x_i^{[ii]} - w_j^{[ii]}\|^2 \\
 & + (\alpha_{[ii]}^{[jj]})^2 \sum_{i=1}^{N^{[ii]}} \sum_{j=1}^{|w|} \left(\mathcal{K}_{\sigma(j, \chi(x_i))}^{[ii]} - \mathcal{K}_{\sigma(j, \chi(x_i))}^{[jj]} \right)^2 \|w_j^{[ii]} - w_j^{[jj]}\|^2
 \end{aligned}$$

Contexte

Clustering

Collaboratif

à base de

Prototypes

Collaboration

sous cadre

Déterministe

SOM

Collaboration à
base de SOMEstimation de
la Confiance

Expérimentation

Collaboration

sous cadre

Probabiliste

Collaboration

sous cadre

Bayésien

Bilan &

Perspectives

$$R_V^{[ii]}(\chi, w) = \alpha_{[ii]}^{[jj]} \sum_{i=1}^N \sum_{j=1}^{|w|} \mathcal{K}_{\sigma(j, \chi(x_i))}^{[ii]} \|x_i^{[ii]} - w_j^{[ii]}\|^2 + (\alpha_{[ii]}^{[jj]})^2 \sum_{i=1}^N \sum_{j=1}^{|w|} \left(\mathcal{K}_{\sigma(j, \chi(x_i))}^{[ii]} - \mathcal{K}_{\sigma(j, \chi(x_i))}^{[jj]} \right)^2 \|w_j^{[ii]} - w_j^{[jj]}\|^2$$

Ajouter une étape dans la phase globale : Calcul automatique des coefficients de collaboration, i.e. confiance.

$$\alpha_{[ii]}^{[jj]}(t+1) = \alpha_{[ii]}^{[jj]}(t) + \frac{\sum_{i=1}^N \sum_{j=1}^{|w|} \mathcal{K}_{\sigma(j, \chi(x_i))}^{[ii]}}{2 \sum_{i=1}^N \sum_{j=1}^{|w|} \left(\mathcal{K}_{\sigma(j, \chi(x_i))}^{[ii]} - \mathcal{K}_{\sigma(j, \chi(x_i))}^{[jj]} \right)^2}$$

Contexte

*Clustering
Collaboratif
à base de
Prototypes*

*Collaboration
sous cadre
Déterministe*

SOM

*Collaboration à
base de SOM*

*Estimation de
la Confiance*

Expérimentation

*Collaboration
sous cadre
Probabiliste*

*Collaboration
sous cadre
Bayésien*

*Bilan &
Perspectives*

Pour avoir des scénarios de données distribuées



ID	Att1	Att2	Att3	AttD
1						
..						
m						
m+1						
m+2						
..						
N						

Verticale

Contexte

*Clustering Collaboratif
à base de
Prototypes*

*Collaboration
sous cadre
Déterministe*

SOM

*Collaboration à
base de SOM*

*Estimation de
la Confiance*

Expérimentation

*Collaboration
sous cadre
Probabiliste*

*Collaboration
sous cadre
Bayésien*

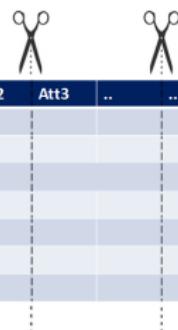
*Bilan &
Perspectives*

Pour avoir des scénarios de données distribuées



ID	Att1	Att2	Att3	AttD
1						
..						
m						
m+1						
m+2						
..						
N						

Verticale



ID	Att1	Att2	Att3	AttD
1						
..						
m						
m+1						
m+2						
..						
N						

Horizontale

Résultats

Contexte

Clustering

Collaboratif
à base de
PrototypesCollaboration
sous cadre

Déterministe

SOM

Collaboration à
base de SOMEstimation de
la Confiance

Expérimentation

Collaboration
sous cadre

Probabiliste

Collaboration
sous cadre

Bayésien

Bilan &
Perspectives

	Données ¹	Carte	Collaboration Horizontale		
			Pureté	<i>qe</i>	α
Wdbc 569×32	SOM_1 SOM_2 $SOM_{1 \leftarrow 2}$ $SOM_{2 \leftarrow 1}$	SOM_1	94.95	1.99	1.74
		SOM_2	97.27	2.07	
		$SOM_{1 \leftarrow 2}$	95.77	1.84	2.12
		$SOM_{2 \leftarrow 1}$	97.32	1.94	
Isolet 150×52	SOM_1 SOM_2 $SOM_{1 \leftarrow 2}$ $SOM_{2 \leftarrow 1}$	SOM_1	81.20	12.61	2.05
		SOM_2	95.12	14.45	
		$SOM_{1 \leftarrow 2}$	81.39	12.21	1.86
		$SOM_{2 \leftarrow 1}$	96.06	14.18	
Madelon 4400×500	SOM_1 SOM_2 $SOM_{1 \leftarrow 2}$ $SOM_{2 \leftarrow 1}$	SOM_1	60.88	15.58	1.65
		SOM_2	62.64	15.50	
		$SOM_{1 \leftarrow 2}$	61.01	15.48	1.79
		$SOM_{2 \leftarrow 1}$	63.57	15.40	
SpamBase 4601×56	SOM_1 SOM_2 $SOM_{1 \leftarrow 2}$ $SOM_{2 \leftarrow 1}$	SOM_1	83.86	3.45	1.92
		SOM_2	85.72	2.55	
		$SOM_{1 \leftarrow 2}$	84.17	3.23	1.59
		$SOM_{2 \leftarrow 1}$	83.59	2.41	
Waveform 5000×40	SOM_1 SOM_2 $SOM_{1 \leftarrow 2}$ $SOM_{2 \leftarrow 1}$	SOM_1	81.64	1.98	1.2
		SOM_2	40.21	2.41	
		$SOM_{1 \leftarrow 2}$	62.47	2.14	5.9
		$SOM_{2 \leftarrow 1}$	54.63	2.27	

1. Données réelles du site UCI, sauf Madelon et Waveform

Résultats

Contexte

Clustering Collaboratif à base de Prototypes

Collaboration sous cadre

Déterministe

SOM

Collaboration à base de SOM

Estimation de la Confiance

Expérimentation

Collaboration sous cadre Probabiliste

Collaboration sous cadre Bayésien

Bilan & Perspectives

	Données ¹	Carte	Collaboration Horizontale		
			Pureté	<i>qe</i>	α
Wdbc 569 × 32	SOM₁	94.95	1.99		
		SOM₂	97.27	2.07	
	SOM _{1 ← 2}	95.77	1.84		1.74
	SOM _{2 ← 1}	97.32	1.94		2.12
Isolet 150 × 52	SOM₁	81.20	12.61		
		SOM ₂	95.12	14.45	
	SOM _{1 ← 2}	81.39	12.21		2.05
	SOM _{2 ← 1}	96.06	14.18		1.86
Madelon 4400 × 500	SOM₁	60.88	15.58		
		SOM ₂	62.64	15.50	
	SOM _{1 ← 2}	61.01	15.48		1.65
	SOM _{2 ← 1}	63.57	15.40		1.79
SpamBase 4601 × 56	SOM₁	83.86	3.45		
		SOM ₂	85.72	2.55	
	SOM _{1 ← 2}	84.17	3.23		1.92
	SOM _{2 ← 1}	83.59	2.41		1.59
Waveform 5000 × 40	SOM₁	81.64	1.98		
		SOM ₂	40.21	2.41	
	SOM _{1 ← 2}	62.47	2.14		1.2
	SOM _{2 ← 1}	54.63	2.27		5.9

1. Données réelles du site UCI, sauf Madelon et Waveform

Résultats

Contexte

Clustering Collaboratif à base de Prototypes

Collaboration sous cadre

Déterministe

SOM

Collaboration à base de SOM

Estimation de la Confiance

Expérimentation

Collaboration sous cadre

Probabiliste

Collaboration sous cadre

Bayésien

Bilan & Perspectives

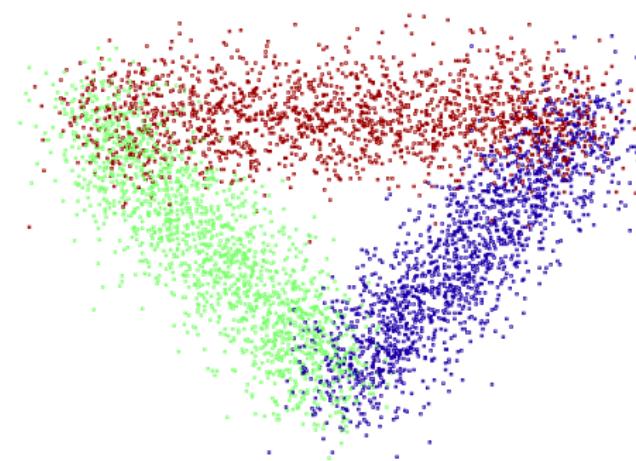
	Données ¹	Carte	Collaboration Horizontale		
			Pureté	<i>qe</i>	α
Wdbc 569 × 32	SOM ₁	94.95	1.99		
	SOM ₂	97.27	2.07		
	SOM _{1 ← 2}	95.77	1.84	1.74	
	SOM _{2 ← 1}	97.32	1.94	2.12	
Isolet 150 × 52	SOM ₁	81.20	12.61		
	SOM ₂	95.12	14.45		
	SOM _{1 ← 2}	81.39	12.21	2.05	
	SOM _{2 ← 1}	96.06	14.18	1.86	
Madelon 4400 × 500	SOM ₁	60.88	15.58		
	SOM ₂	62.64	15.50		
	SOM _{1 ← 2}	61.01	15.48	1.65	
	SOM _{2 ← 1}	63.57	15.40	1.79	
SpamBase 4601 × 56	SOM ₁	83.86	3.45		
	SOM ₂	85.72	2.55		
	SOM _{1 ← 2}	84.17	3.23	1.92	
	SOM _{2 ← 1}	83.59	2.41	1.59	
Waveform 5000 × 40	SOM ₁	81.64	1.98		
	SOM ₂	40.21	2.41		
	SOM _{1 ← 2}	62.47	2.14	1.2	
	SOM _{2 ← 1}	54.63	2.27	5.9	

1. Données réelles du site UCI, sauf Madelon et Waveform

*Contexte**Clustering
Collaboratif
à base de
Prototypes**Collaboration
sous cadre
Déterministe**SOM**Collaboration à
base de SOM**Estimation de
la Confiance**Expérimentation**Collaboration
sous cadre
Probabiliste**Collaboration
sous cadre
Bayésien**Bilan &
Perspectives*

Waveform

- 5000 observations.
- 40 variables, dont 21 pertinentes et 19 bruitées.
- 3 clusters



Résultats de collaboration horizontale, base Waveform

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

SOM

Collaboration à
base de SOM

Estimation de
la Confiance

Expérimentation

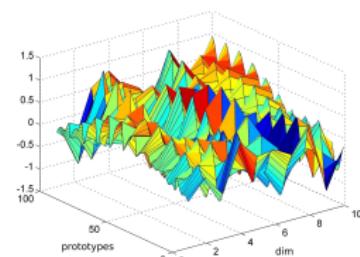
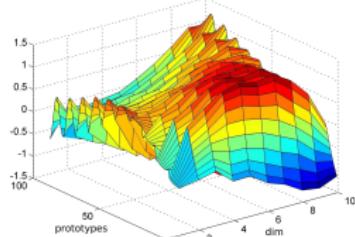
Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Variables pertinentes
pureté=81.64%

Variables bruitées
pureté=40.21%



Les valeurs de pureté plus élevées
indiquent une meilleure classification.

Résultats de collaboration horizontale, base Waveform

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

SOM

Collaboration à
base de SOM

Estimation de
la Confiance

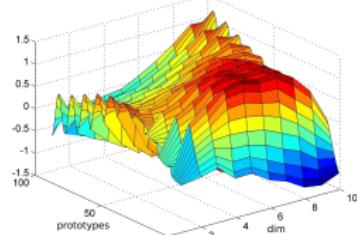
Expérimentation

Collaboration
sous cadre
Probabiliste

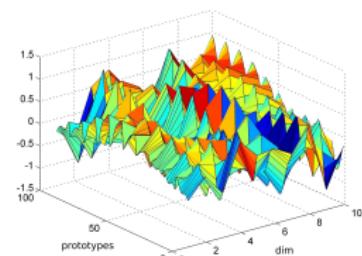
Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Variables pertinentes
pureté=81.64%



Variables bruitées
pureté=40.21%



Les valeurs de pureté plus élevées
indiquent une meilleure classification.

Résultats de collaboration horizontale, base Waveform

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

SOM

Collaboration à
base de SOM

Estimation de
la Confiance

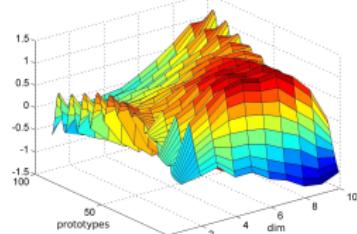
Expérimentation

Collaboration
sous cadre
Probabiliste

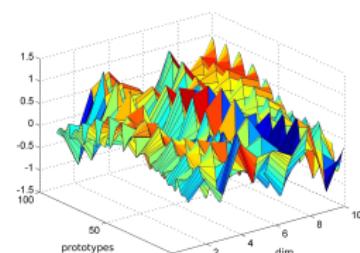
Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Variables pertinentes
pureté=81.64%



Variables bruitées
pureté=40.21%



Les valeurs de pureté plus élevées
indiquent une meilleure classification.

Résultats de collaboration horizontale, base Waveform

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

SOM

Collaboration à
base de SOM

Estimation de
la Confiance

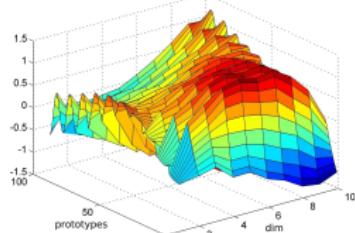
Expérimentation

Collaboration
sous cadre
Probabiliste

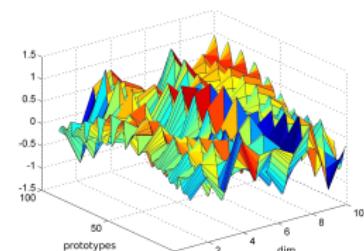
Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

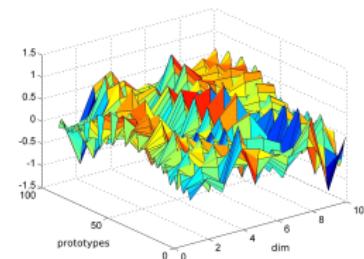
Variables pertinentes
pureté=81.64%



Variables bruitées
pureté=40.21%



Les valeurs de pureté plus élevées
indiquent une meilleure classification.



pureté=54.63%
 $\alpha = 5.9$

Résultats de collaboration horizontale, base Waveform

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

SOM

Collaboration à
base de SOM

Estimation de
la Confiance

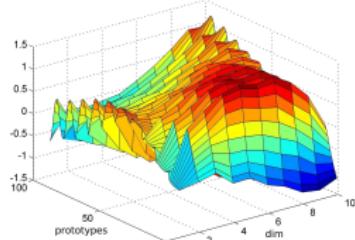
Expérimentation

Collaboration
sous cadre
Probabiliste

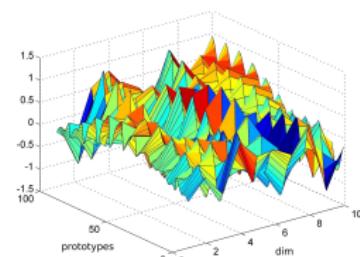
Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Variables pertinentes
pureté=81.64%



Variable bruitées
pureté=40.21%



Les valeurs de pureté plus élevées
indiquent une meilleure classification.

Résultats de collaboration horizontale, base Waveform

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

SOM

Collaboration à
base de SOM

Estimation de
la Confiance

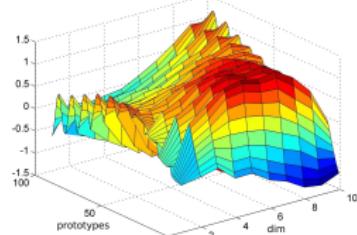
Expérimentation

Collaboration
sous cadre
Probabiliste

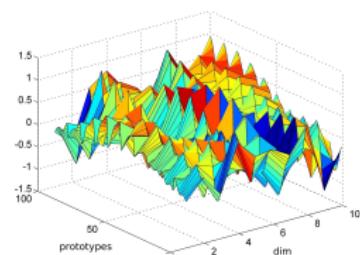
Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Variables pertinentes
pureté=81.64%



Variable bruitées
pureté=40.21%



Les valeurs de pureté plus élevées
indiquent une meilleure classification.

Résultats de collaboration horizontale, base Waveform

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

SOM

Collaboration à
base de SOM

Estimation de
la Confiance

Expérimentation

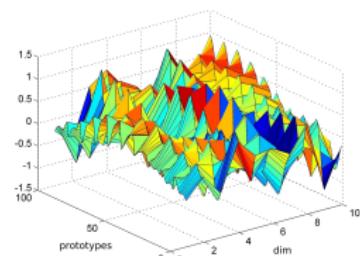
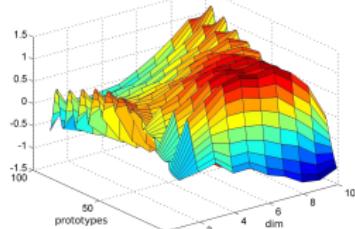
Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Variables pertinentes
pureté=81.64%

Variable bruitées
pureté=40.21%



Les valeurs de pureté plus élevées
indiquent une meilleure classification.

Résultats de collaboration horizontale, base Waveform

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

SOM

Collaboration à
base de SOM

Estimation de
la Confiance

Expérimentation

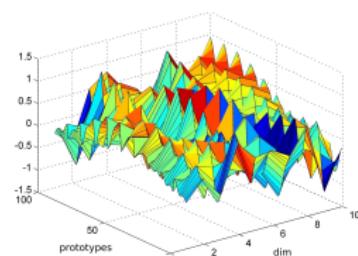
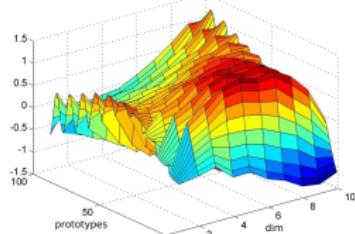
Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

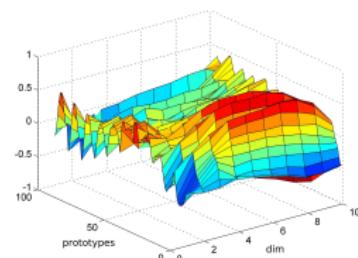
Bilan &
Perspectives

Variables pertinentes
pureté=81.64%

Variable bruitées
pureté=40.21%



Les valeurs de pureté plus élevées
indiquent une meilleure classification.



pureté=62.47%
 $\alpha = 1.2$

*Contexte**Clustering**Collaboratif**à base de**Prototypes**Collaboration**sous cadre**Déterministe**SOM**Collaboration à**base de SOM**Estimation de**la Confiance**Expérimentation**Collaboration**sous cadre**Probabiliste**Collaboration**sous cadre**Bayésien**Bilan &**Perspectives*

- ! SOM ne définit pas un modèle à densité dans l'espace de données.
- ! Convergence non garantie.
- ! Manque de théorie pour le choix de l'initialisation des paramètres.

*Contexte**Clustering**Collaboratif**à base de**Prototypes**Collaboration**sous cadre**Déterministe**SOM**Collaboration à**base de SOM**Estimation de**la Confiance**Expérimentation**Collaboration**sous cadre**Probabiliste**Collaboration**sous cadre**Bayésien**Bilan &**Perspectives*

- ! SOM ne définit pas un modèle à densité dans l'espace de données.
- ! Convergence non garantie.
- ! Manque de théorie pour le choix de l'initialisation des paramètres.

Solution proposée : Utiliser un modèle **probabiliste**.

Plan de la présentation

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

GTM
EM
Co-GTM

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

1 Contexte

2 Clustering Collaboratif à base de Prototypes

3 Collaboration sous cadre Déterministe

4 Collaboration sous cadre Probabiliste

5 Collaboration sous cadre Bayésien

6 Bilan & Perspectives

Generative Topographic Mapping (GTM) [Bishop95]

Contexte
Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

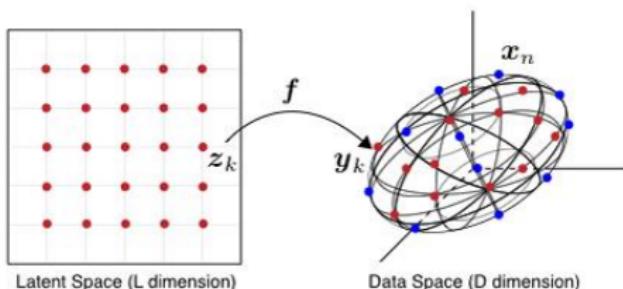
GTM

EM

Co-GTM

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives



Generative Topographic Mapping (GTM) [Bishop95]

Contexte
Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

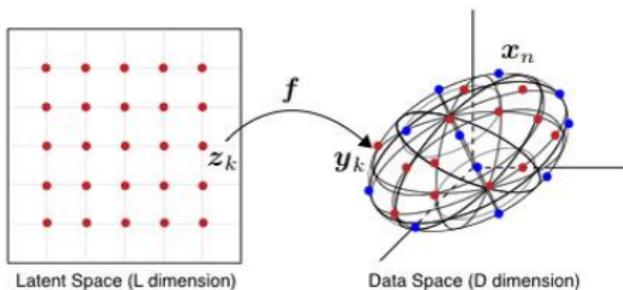
GTM

EM

Co-GTM

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives



- Transformation de l'espace latente en espace de données, contrôlée par une matrice de poids W et des fonctions de base $\Phi(z)$, formant les prototypes :

$$y = \textcolor{red}{y}(z, W) = W\Phi(z)$$

Contexte
Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

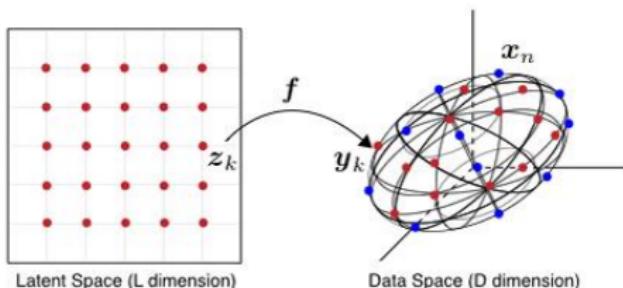
Collaboration
sous cadre
Probabiliste

GTM
EM

Co-GTM

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives



- Transformation de l'espace latente en espace de données, contrôlée par une matrice de poids W et des fonctions de base $\Phi(z)$, formant les prototypes :

$$y = \textcolor{red}{y}(z, W) = W\Phi(z)$$

- Distribution Gaussienne des données :

$$p(\textcolor{red}{x}_n|z, W, \beta) = \mathcal{N}(\textcolor{red}{y}(z, W), \beta)$$

Contexte
Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

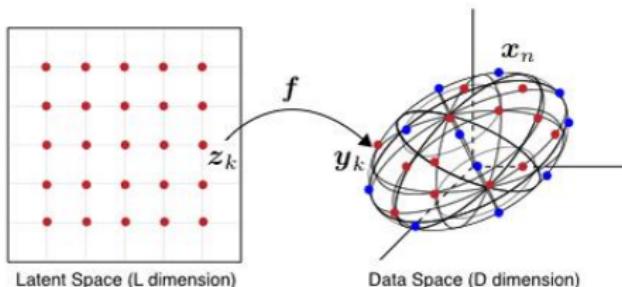
Collaboration
sous cadre
Probabiliste

GTM
EM

Co-GTM

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives



- Transformation de l'espace latente en espace de données, contrôlée par une matrice de poids W et des fonctions de base $\Phi(z)$, formant les prototypes :

$$y = \textcolor{red}{y}(z, W) = W\Phi(z)$$

- Distribution Gaussienne des données :

$$p(\textcolor{red}{x}_n|z, W, \beta) = \mathcal{N}(\textcolor{red}{y}(z, W), \beta)$$

⇒ Mélange de Gaussiennes de Vraisemblance $\mathcal{L}(\theta)$ avec $\theta = \{W, \beta\}$ les paramètres à apprendre.

*Contexte**Clustering**Collaboratif
à base de
Prototypes**Collaboration
sous cadre
Déterministe**Collaboration
sous cadre
Probabiliste**GTM**EM**Co-GTM**Collaboration
sous cadre
Bayésien**Bilan &
Perspectives*

- Vraisemblance de l'échantillon X :

$$p(X; \theta) = \int p(X|\theta)p(\theta)d\theta \text{ (2)}, \quad p(\theta|X) \text{ (3)} = \frac{p(X; \theta)p(\theta)}{p(X)}$$

- **But** : Maximiser la **vraisemblance** de X par rapport à l'ensemble de paramètres θ .

(2). Sommation si les variables aléatoires sont discrètes.
(3). Théorème de Bayes.

Contexte

Clustering

Collaboratif
à base de
PrototypesCollaboration
sous cadre
DéterministeCollaboration
sous cadre
Probabiliste

GTM

EM

Co-GTM

Collaboration
sous cadre
BayésienBilan &
Perspectives

- Vraisemblance de l'échantillon X :

$$p(X; \theta) = \int p(X|\theta)p(\theta)d\theta \text{ (2)}, \quad p(\theta|X) \text{ (3)} = \frac{p(X; \theta)p(\theta)}{p(X)}$$

- But** : Maximiser la **vraisemblance** de X par rapport à l'ensemble de paramètres θ .

! En présence des variables latentes/cachées :

$$p(X; \theta) = \iint p(X, Z|\theta)p(\theta) dZ d\theta$$

! En général, pas de solution analytique à ce problème.

(2). Sommation si les variables aléatoires sont discrètes.

(3). Théorème de Bayes.

Contexte

Clustering

Collaboratif
à base de
PrototypesCollaboration
sous cadre
DéterministeCollaboration
sous cadre
Probabiliste

GTM

EM

Co-GTM

Collaboration
sous cadre
BayésienBilan &
Perspectives

- Vraisemblance de l'échantillon X :

$$p(X; \theta) = \int p(X|\theta)p(\theta)d\theta \text{ (2)}, \quad p(\theta|X) \text{ (3)} = \frac{p(X; \theta)p(\theta)}{p(X)}$$

- But** : Maximiser la **vraisemblance** de X par rapport à l'ensemble de paramètres θ .

! En présence des variables latentes/cachées :

$$p(X; \theta) = \iint p(X, Z|\theta)p(\theta) dZ d\theta$$

! En général, pas de solution analytique à ce problème.

- Solution** → Algorithme **EM** = approche **itérative** pour un optimum local.

(2). Sommation si les variables aléatoires sont discrètes.

(3). Théorème de Bayes.

Etapes E & M pour le calcul des paramètres

Etape E - Calcul des probabilités a posteriori

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

GTM

EM

Co-GTM

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

$$\begin{aligned} r_{in} &= p(z_i|x_n, W_{old}, \beta_{old}) \\ &= \frac{p(x_n|z_i, W_{old}, \beta_{old})}{\sum_{i'=1}^K p(x_n|z'_i, W_{old}, \beta_{old})} \end{aligned}$$

La probabilité que la donnée x_n a été générée par la composante Gaussienne i .

Etapes E & M pour le calcul des paramètres

Etape E - Calcul des probabilités a posteriori

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

GTM

EM

Co-GTM

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

$$\begin{aligned} r_{in} &= p(z_i|x_n, W_{old}, \beta_{old}) \\ &= \frac{p(x_n|z_i, W_{old}, \beta_{old})}{\sum_{i'=1}^K p(x_n|z'_i, W_{old}, \beta_{old})} \end{aligned}$$

La probabilité que la donnée x_n a été générée par la composante Gaussienne i .

Etape M - Mise à jour des paramètres

Borne inférieure : $\mathbb{E}[\mathcal{L}_{comp}(W, \beta)] = \sum_{n=1}^N \sum_{i=1}^K r_{in} \ln\{p(x_n|z_i, W, \beta)\}$

Paramètres : $\left\{ \begin{array}{l} \Phi^T G \Phi \boxed{W}_{new}^T = \Phi^T R X \\ \boxed{\beta}_{new} = \frac{1}{ND} \sum_{n=1}^N \sum_{i=1}^K r_{in} \|x_n - W^{new} \phi(z_i)\|^2 \end{array} \right.$

Etapes E & M pour le calcul des paramètres

Etape E - Calcul des probabilités a posteriori

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

GTM

EM

Co-GTM

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

$$r_{in} = p(z_i|x_n, W_{old}, \beta_{old})$$

$$= \frac{p(x_n|z_i, W_{old}, \beta_{old})}{\sum_{i'=1}^K p(x_n|z'_i, W_{old}, \beta_{old})}$$

La probabilité que la donnée x_n a été générée par la composante Gaussienne i .

Etape M - Mise à jour des paramètres

Introduire terme de collaboration ici

Borne inférieure : $\mathbb{E}[\mathcal{L}_{comp}(W, \beta)] = \sum_{n=1}^N \sum_{i=1}^K r_{in} \ln\{p(x_n|z_i, W, \beta)\}$

Paramètres : $\left\{ \begin{array}{l} \Phi^T G \Phi \boxed{W}_{new}^T = \Phi^T R X \\ \boxed{\beta}_{new} = \frac{1}{ND} \sum_{n=1}^N \sum_{i=1}^K r_{in} \|x_n - W^{new} \phi(z_i)\|^2 \end{array} \right.$

*Contexte**Clustering Collaboratif à base de Prototypes**Collaboration sous cadre Déterministe**Collaboration sous cadre Probabiliste**GTM**EM**Co-GTM**Collaboration sous cadre Bayésien**Bilan & Perspectives*

Approche Verticale

$$\begin{aligned} \mathcal{L}^{ver}[ii] = & \mathbb{E}[\mathcal{L}_{comp}(W^{[ii]}, \beta^{[ii]})] - \\ & \sum_{[jj]=1, [jj] \neq [ii]}^P \alpha_{[ii]}^{[jj]} \sum_{n=1}^{N[ii]} \sum_{i=1}^K r_{in} \frac{\beta^{[ii]}}{2} \boxed{\|W^{[ii]} \phi^{[ii]}(z_i) - W^{[jj]} \phi^{[jj]}(z_i)\|^2} \end{aligned}$$

Contexte

Clustering
Collaboratif
à base de
PrototypesCollaboration
sous cadre
DéterministeCollaboration
sous cadre
Probabiliste

GTM

EM

Co-GTM

Collaboration
sous cadre
BayésienBilan &
Perspectives

Approche Verticale

$$\mathcal{L}^{ver}[ii] = \mathbb{E}[\mathcal{L}_{comp}(W^{[ii]}, \beta^{[ii]})] -$$

$$\sum_{[jj]=1, [jj] \neq [ii]}^P \alpha_{[ii]}^{[jj]} \sum_{n=1}^{N[ii]} \sum_{i=1}^K r_{in} \frac{\beta^{[ii]}}{2} \boxed{\|W^{[ii]} \phi^{[ii]}(z_i) - W^{[jj]} \phi^{[jj]}(z_i)\|^2}$$

Approche Horizontale

$$\mathcal{L}^{hor}[ii] = \mathbb{E}[\mathcal{L}_{comp}(W^{[ii]}, \beta^{[ii]})] -$$

$$\sum_{[jj]=1, [jj] \neq [ii]}^P \alpha_{[ii]}^{[jj]} \sum_{n=1}^N \sum_{i=1}^K \frac{\beta^{[ii]}}{2} \boxed{(r_{in}^{[ii]} - r_{in}^{[jj]})^2} \|x_n - W^{[ii]} \phi^{[ii]}(z_i)\|^2$$

*Contexte**Clustering**Collaboratif**à base de**Prototypes**Collaboration**sous cadre**Déterministe**Collaboration**sous cadre**Probabiliste**GTM**EM**Co-GTM**Collaboration**sous cadre**Bayésien**Bilan &**Perspectives*

- ! Optimisation vraisemblance = critère non borné
- ! Le calcul ne tient pas compte de la complexité du modèle.
- ! Risque de sur-apprentissage.

*Contexte**Clustering**Collaboratif**à base de**Prototypes**Collaboration**sous cadre**Déterministe**Collaboration**sous cadre**Probabiliste**GTM**EM**Co-GTM**Collaboration**sous cadre**Bayésien**Bilan &**Perspectives*

- ! Optimisation vraisemblance = critère non borné
- ! Le calcul ne tient pas compte de la complexité du modèle.
- ! Risque de sur-apprentissage.

Solution proposée : GTM sous cadre d'**Approximation Variationnelle**.

Plan de la présentation

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Inférence
Variationnelle
VB-GTM

Comparaison
F-VBGTM

Collaboration à
base de
VBGTM

Bilan &
Perspectives

1 Contexte

2 Clustering Collaboratif à base de Prototypes

3 Collaboration sous cadre Déterministe

4 Collaboration sous cadre Probabiliste

5 Collaboration sous cadre Bayésien

6 Bilan & Perspectives

*Contexte**Clustering
Collaboratif
à base de
Prototypes**Collaboration
sous cadre
Déterministe**Collaboration
sous cadre
Probabiliste**Collaboration
sous cadre
Bayésien**Inférence
Variationnelle**VB-GTM**Comparaison**F-VBGTM**Collaboration à
base de
VBGTM**Bilan &
Perspectives*

Soit Z les variables latentes, X les données et θ les paramètres.
La Log-Vraisemblance s'écrit :

$$\ln p(\mathbf{X}) = \ln \int p(X, Z, \theta) dZ d\theta$$

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Inférence
Variationnelle

VB-GTM

Comparaison

F-VBGTM

Collaboration à
base de
VBGTM

Bilan &
Perspectives

Soit Z les variables latentes, X les données et θ les paramètres.
La Log-Vraisemblance s'écrit :

$$\begin{aligned}\ln p(X) &= \ln \int p(X, Z, \theta) dZ d\theta \\ &= \ln \int q(Z, \theta) \frac{p(X, Z, \theta)}{q(Z, \theta)} dZ d\theta\end{aligned}$$

*Contexte**Clustering**Collaboratif
à base de
Prototypes**Collaboration
sous cadre
Déterministe**Collaboration
sous cadre
Probabiliste**Collaboration
sous cadre
Bayésien**Inférence
Variationnelle**VB-GTM**Comparaison**F-VBGTM**Collaboration à
base de
VBGTM**Bilan &
Perspectives*

Soit Z les variables latentes, X les données et θ les paramètres.
La Log-Vraisemblance s'écrit :

$$\begin{aligned}\ln p(X) &= \ln \int p(X, Z, \theta) dZ d\theta \\ &= \ln \int q(Z, \theta) \frac{p(X, Z, \theta)}{q(Z, \theta)} dZ d\theta \\ &\geq \int q(Z, \theta) \ln \frac{p(X, Z, \theta)}{q(Z, \theta)} dZ d\theta\end{aligned}$$

*Contexte**Clustering Collaboratif à base de Prototypes**Collaboration sous cadre Déterministe**Collaboration sous cadre Probabiliste**Collaboration sous cadre Bayésien**Inférence Variationnelle**VB-GTM**Comparaison**F-VBGTM**Collaboration à base de VBGTM**Bilan & Perspectives*

Soit Z les variables latentes, X les données et θ les paramètres.
La Log-Vraisemblance s'écrit :

$$\begin{aligned}\ln p(X) &= \ln \int p(X, Z, \theta) dZ d\theta \\ &= \ln \int q(Z, \theta) \frac{p(X, Z, \theta)}{q(Z, \theta)} dZ d\theta \\ &\geq \int q(Z, \theta) \ln \frac{p(X, Z, \theta)}{q(Z, \theta)} dZ d\theta\end{aligned}$$

Approximation simple et factorisée de $q(Z, \theta) \approx q_Z(Z)q_{\theta}(\theta)$:

$$\begin{aligned}\ln p(X) &\geq \int q_Z(Z)q_{\theta}(\theta) \ln \frac{p(X, Z, \theta)}{q_Z(Z)q_{\theta}(\theta)} dZ d\theta \\ &= \mathcal{F}(q_Z(Z), q_{\theta}(\theta), X).\end{aligned}$$

Contexte

Clustering

Collaboratif
à base de
PrototypesCollaboration
sous cadre
DéterministeCollaboration
sous cadre
ProbabilisteCollaboration
sous cadre
BayésienInférence
Variationnelle

VB-GTM

Comparaison

F-VBGTM

Collaboration à
base de
VBGTMBilan &
Perspectives

Soit Z les variables latentes, X les données et θ les paramètres.
La Log-Vraisemblance s'écrit :

$$\begin{aligned}\ln p(X) &= \ln \int p(X, Z, \theta) dZ d\theta \\ &= \ln \int q(Z, \theta) \frac{p(X, Z, \theta)}{q(Z, \theta)} dZ d\theta \\ &\geq \int q(Z, \theta) \ln \frac{p(X, Z, \theta)}{q(Z, \theta)} dZ d\theta\end{aligned}$$

Approximation simple et factorisée de $q(Z, \theta) \approx q_Z(Z)q_{\theta}(\theta)$:

$$\begin{aligned}\ln p(X) &\geq \int q_Z(Z)q_{\theta}(\theta) \ln \frac{p(X, Z, \theta)}{q_Z(Z)q_{\theta}(\theta)} dZ d\theta \\ &= \mathcal{F}(q_Z(Z), q_{\theta}(\theta), X).\end{aligned}$$

Maximisation de $\mathcal{F} \implies$ Maximisation de **Vraisemblance**.

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Inférence
Variationnelle

VB-GTM

Comparaison

F-VBGTM

Collaboration à
base de
VBGTM

Bilan &
Perspectives

Maximisation de \mathcal{F} itérative :

$$\text{VB-E : } q_Z^{(\text{new})}(Z) \approx \exp \left[\int p(Z, X|\theta) q_\theta^{(\text{old})}(\theta) \right]$$

$$\text{VB-M : } q_\theta^{(\text{new})}(\theta) \approx p(\theta) \exp \left[\int p(Z, X|\theta) q_Z^{(\text{new})}(Z) \right]$$

La complexité du modèle est intégrée dans \mathcal{F} .

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Inférence
Variationnelle

VB-GTM

Comparaison

F-VBGTM

Collaboration à
base de
VBGTM

Bilan &
Perspectives

Maximisation de \mathcal{F} itérative :

$$\text{VB-E : } q_Z^{(\text{new})}(Z) \approx \exp \left[\int p(Z, X|\theta) q_\theta^{(\text{old})}(\theta) \right]$$

$$\text{VB-M : } q_\theta^{(\text{new})}(\theta) \approx p(\theta) \exp \left[\int p(Z, X|\theta) q_Z^{(\text{new})}(Z) \right]$$

La complexité du modèle est intégrée dans \mathcal{F} .

Maximiser \mathcal{F} est équivalent à minimiser la divergence KL entre l'*a posteriori* approchée, $q_Z(Z)q_\theta(\theta)$, et la vraie *a posteriori*, $p(Z, \theta|X)$.

Variational Bayesian GTM (VBGTM) [Olier10]

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Inférence
Variationnelle

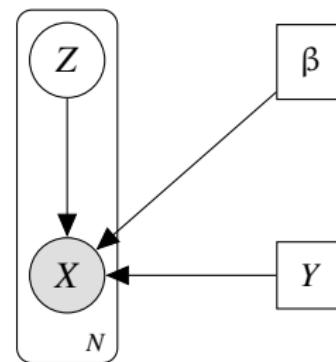
VB-GTM

Comparaison

F-VBGTM

Collaboration à
base de
VBGTM

Bilan &
Perspectives



Contexte

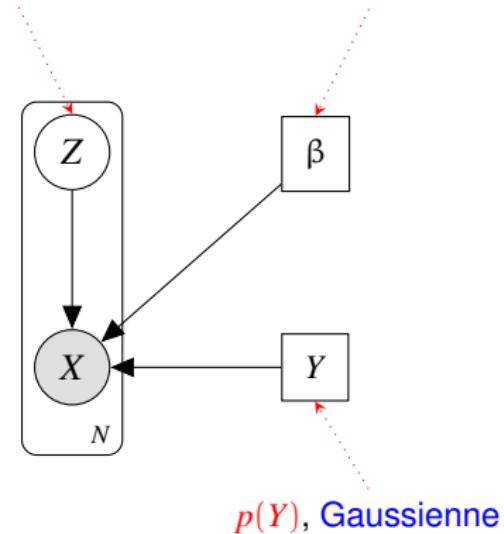
Clustering

Collaboratif
à base de
PrototypesCollaboration
sous cadre
DéterministeCollaboration
sous cadre
ProbabilisteCollaboration
sous cadre
BayésienInférence
Variationnelle

VB-GTM

Comparaison

F-VBGTM

Collaboration à
base de
VBGTMBilan &
Perspectives $p(Z)$, Multinomiale $p(\beta)$, Gamma $p(Y)$, GaussienneProposer des **a priori**, les **conjugués** des lois de la famille exponentielle.

Contexte

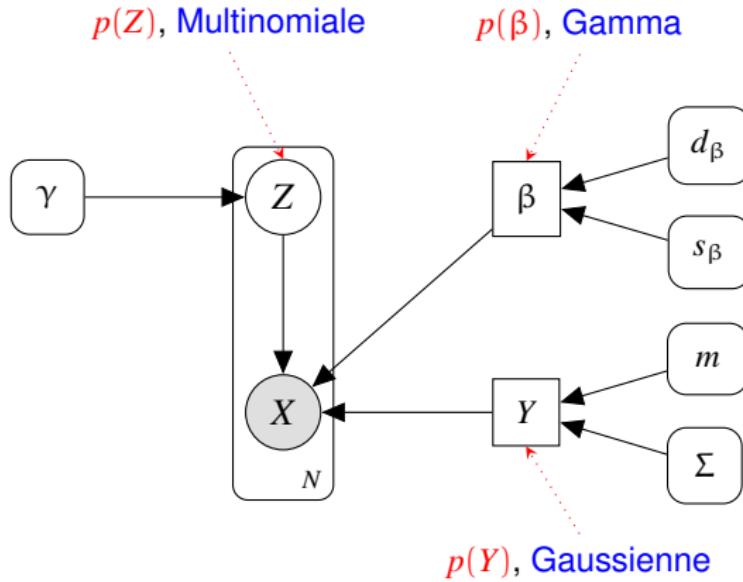
Clustering

Collaboratif
à base de
PrototypesCollaboration
sous cadre
DéterministeCollaboration
sous cadre
ProbabilisteCollaboration
sous cadre
BayésienInférence
Variationnelle

VB-GTM

Comparaison

F-VBGTM

Collaboration à
base de
VBGTMBilan &
Perspectives

Des hyperparamètres apparaissent.

Contexte

Clustering
Collaboratif
à base de
PrototypesCollaboration
sous cadre
DéterministeCollaboration
sous cadre
ProbabilisteCollaboration
sous cadre
BayésienInférence
Variationnelle

VB-GTM

Comparaison

F-VBGTM

Collaboration à
base de
VBGTMBilan &
Perspectives

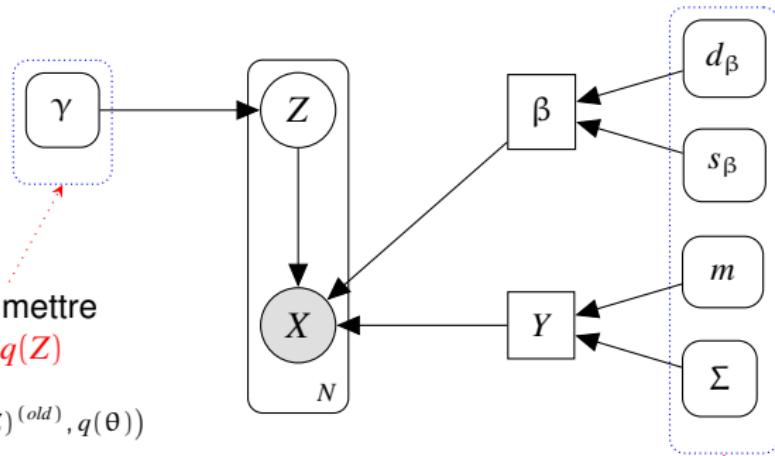
EM-Variationnel

VBE : mettre
à jour $q(Z)$

$$\underset{q(Z)}{\operatorname{argmax}} F(q(Z)^{(old)}, q(\theta))$$

VBM : mettre à jour $q(\theta)$

$$\underset{q(\theta)}{\operatorname{argmax}} F(q(Z)^{(new)}, q(\theta))$$



Comparison des résultats sur la base *Wine* (178×13 , qualité de vin, 3 clusters).

Contexte

Clustering

Collaboratif

à base de

Prototypes

Collaboration

sous cadre

Déterministe

Collaboration

sous cadre

Probabiliste

Collaboration

sous cadre

Bayésien

Inférence

Variationnelle

VB-GTM

Comparaison

F-VBGTM

Collaboration à

base de

VBGTM

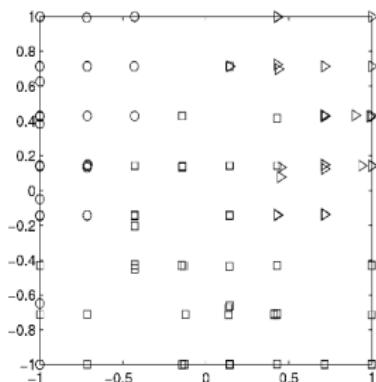
Bilan &

Perspectives

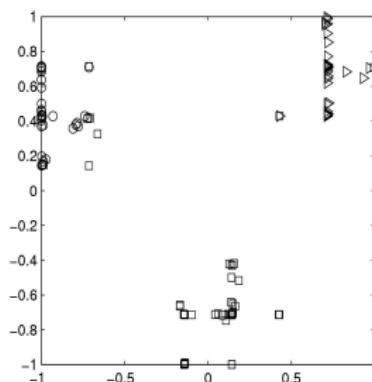
*Contexte**Clustering**Collaboratif
à base de
Prototypes**Collaboration
sous cadre
Déterministe**Collaboration
sous cadre
Probabiliste**Collaboration
sous cadre
Bayésien**Inférence
Variationnelle
VB-GTM**Comparaison
F-VBGTM**Collaboration à
base de
VBGTM**Bilan &
Perspectives*

Comparison des résultats sur la base *Wine* (178×13 , qualité de vin, 3 clusters).

Visualisation par projection de moyenne des apostérioris dans l'espace latente.



GTM

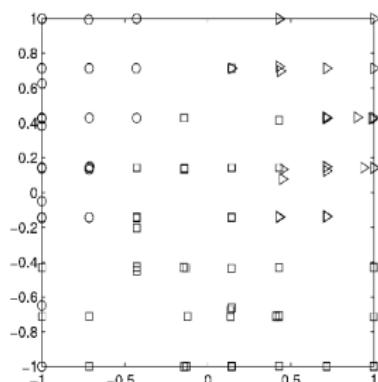


VBGTM

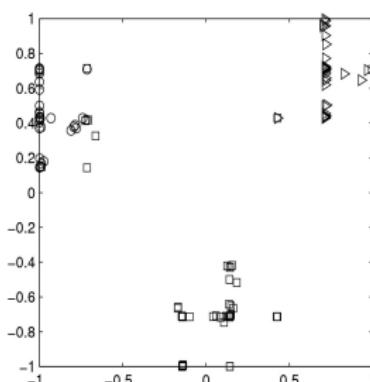
Contexte

Clustering
Collaboratif
à base de
PrototypesCollaboration
sous cadre
DéterministeCollaboration
sous cadre
ProbabilisteCollaboration
sous cadre
BayésienInférence
Variationnelle
VB-GTMComparaison
F-VBGTMCollaboration à
base de
VBGTMBilan &
PerspectivesComparison des résultats sur la base *Wine* (178×13 , qualité de vin, 3 clusters).

Visualisation par projection de moyenne des apostérioris dans l'espace latente.



GTM



VBGTM

- ! Les étiquettes dans les figures sont les **vraies étiquettes**.
- ! VBGTM ne **classifie** pas les données.

Extension : Clustering flou de VBGTM

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

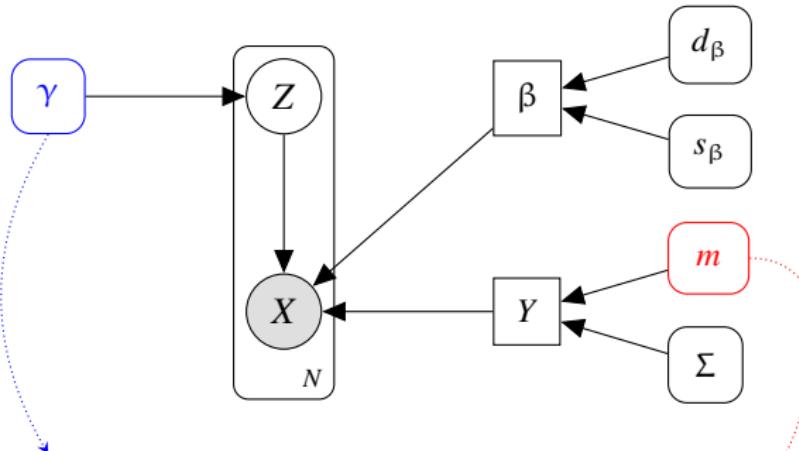
Collaboration
sous cadre
Bayésien

Inférence
Variationnelle
VB-GTM

Comparaison
F-VBGTM

Collaboration à
base de
VBGTM

Bilan &
Perspectives



Les probabilités a posteriori → degré
d'appartenance.

Les centres des Gaussiens dans l'espace de données.

Extension : Clustering flou de VBGTM

Données



dimension D

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Inférence
Variationnelle

VB-GTM

Comparaison

F-VBGTM

Collaboration à
base de
VBGTM

Bilan &
Perspectives

Extension : Clustering flou de VBGTM

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Inférence
Variationnelle

VB-GTM

Comparaison

F-VBGTM

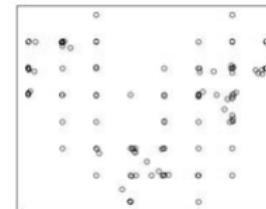
Collaboration à
base de
VBGTM

Bilan &
Perspectives

Données



Etape 1 : Modèle VBGTM



dimension D

Extension : Clustering flou de VBGTM

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Inférence
Variationnelle
VB-GTM

Comparaison
F-VBGTM

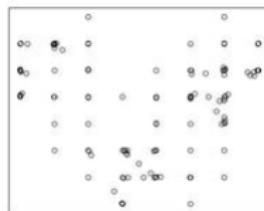
Collaboration à
base de
VBGTM

Bilan &
Perspectives

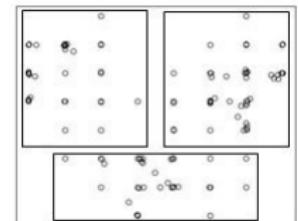
Données



Etape 1 : Modèle VBGTM



Etape 2 : F-VBGTM



dimension D

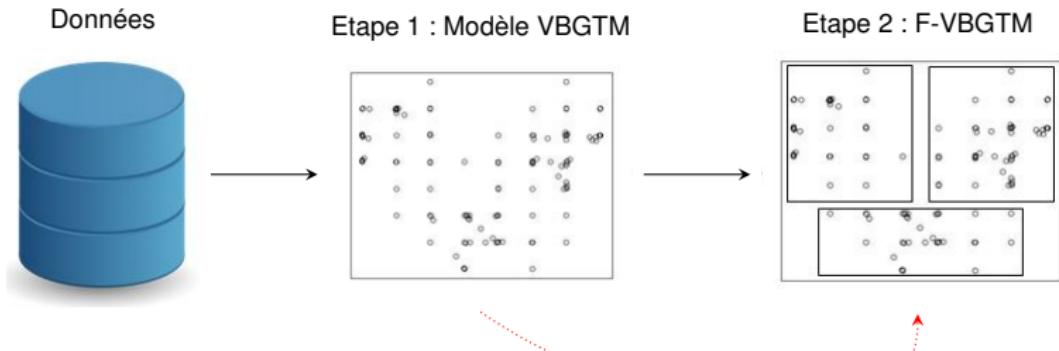
Extension : Clustering flou de VBGTM

Contexte

Clustering
Collaboratif
à base de
PrototypesCollaboration
sous cadre
DéterministeCollaboration
sous cadre
ProbabilisteCollaboration
sous cadre
BayésienInférence
Variationnelle
VB-GTM

Comparaison

F-VBGTM

Collaboration à
base de
VBGTMBilan &
Perspectives

- FCM sur les centres des Gaussiens.
 - Prototypes.
 - Degrés d'appartenance de ces centres.
- Théorème de Bayes : Degrés d'appartenance des données aux prototypes.

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Inférence
Variationnelle

VB-GTM

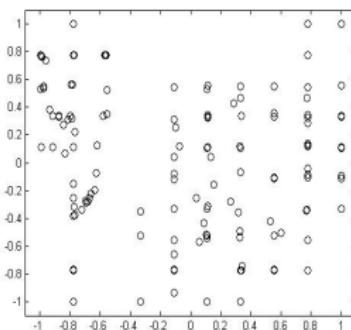
Comparaison

F-VBGTM

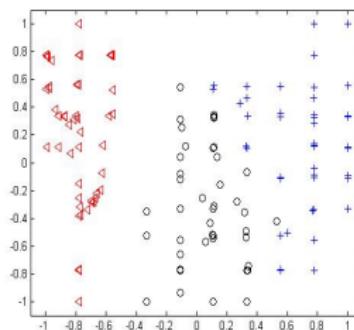
Collaboration à
base de
VBGTM

Bilan &
Perspectives

F-VBGTM : Visualisation + Classification.



VBGTM



F-VBGTM

Résultats sur la base *Iris* (150×4 , genre de fleurs, 3 clusters).

Contexte

*Clustering
Collaboratif
à base de
Prototypes*



$D[1]$

*Collaboration
sous cadre
Déterministe*



$D[2]$

*Collaboration
sous cadre
Probabiliste*

*Collaboration
sous cadre
Bayésien*

*Inférence
Variationnelle
VB-GTM
Comparaison
F-VBGTM*

Données

Distribuées



$D[P]$

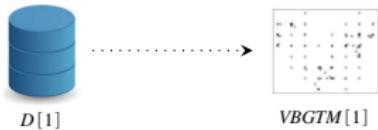
*Collaboration à
base de
VBGTM*

*Bilan &
Perspectives*

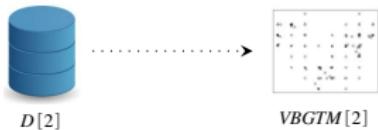
Collaboration à base de VBGTM : Co-F-VBGTM [Ghassany13]

Contexte

Clustering
Collaboratif
à base de
Prototypes



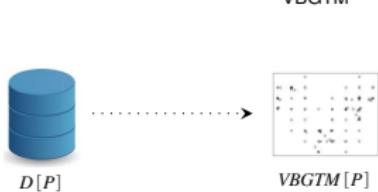
Collaboration
sous cadre
Déterministe



Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Inférence
Variationnelle
VB-GTM
Comparaison
F-VBGTM



Collaboration à base de VBGTM : Co-F-VBGTM [Ghassany13]

Contexte

*Clustering
Collaboratif
à base de
Prototypes*

*Collaboration
sous cadre
Déterministe*

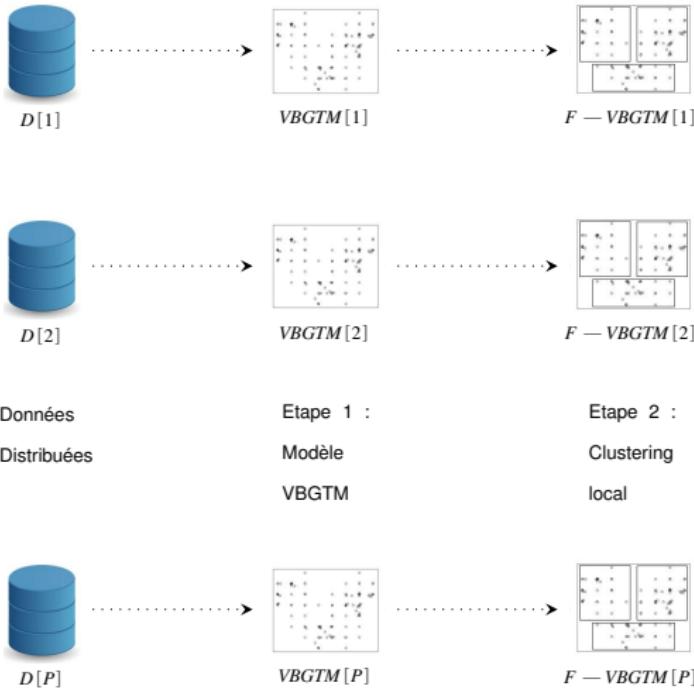
*Collaboration
sous cadre
Probabiliste*

*Collaboration
sous cadre
Bayésien*

*Inférence
Variationnelle
VB-GTM
Comparaison
F-VBGTM*

*Collaboration à
base de
VBGTM*

*Bilan &
Perspectives*



Collaboration à base de VBGTM : Co-F-VBGTM [Ghassany13]

Contexte
Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

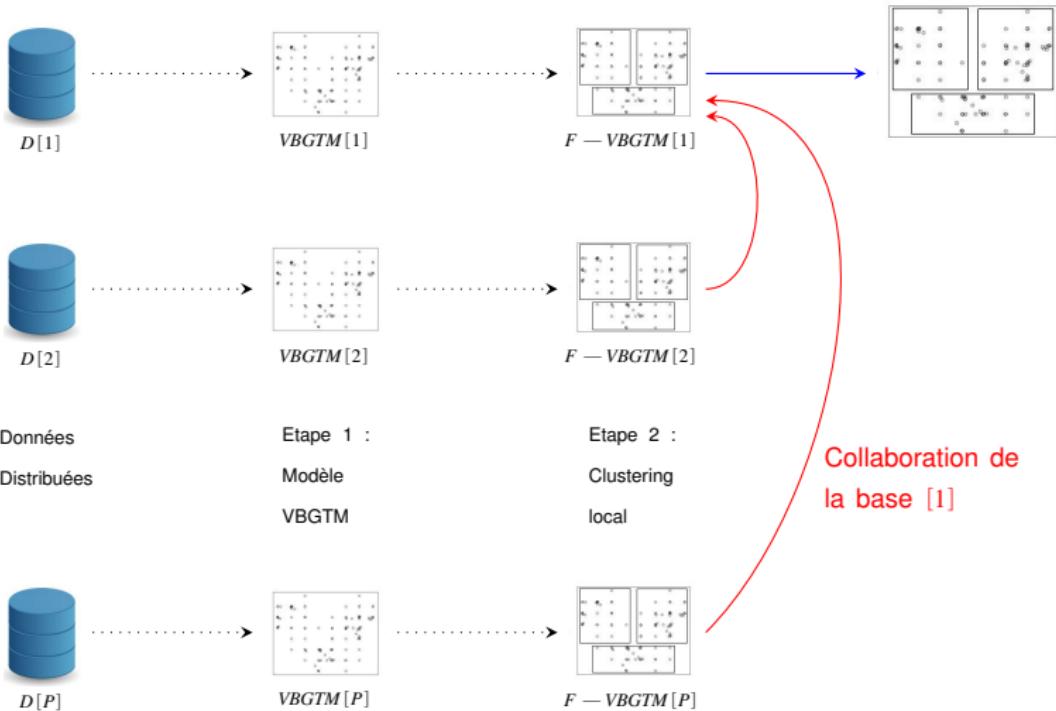
Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Inférence
Variationnelle
VB-GTM
Comparaison
F-VBGTM

Collaboration à
base de
VBGTM

Bilan &
Perspectives



Collaboration à base de VBGTM : Co-F-VBGTM [Ghassany13]

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

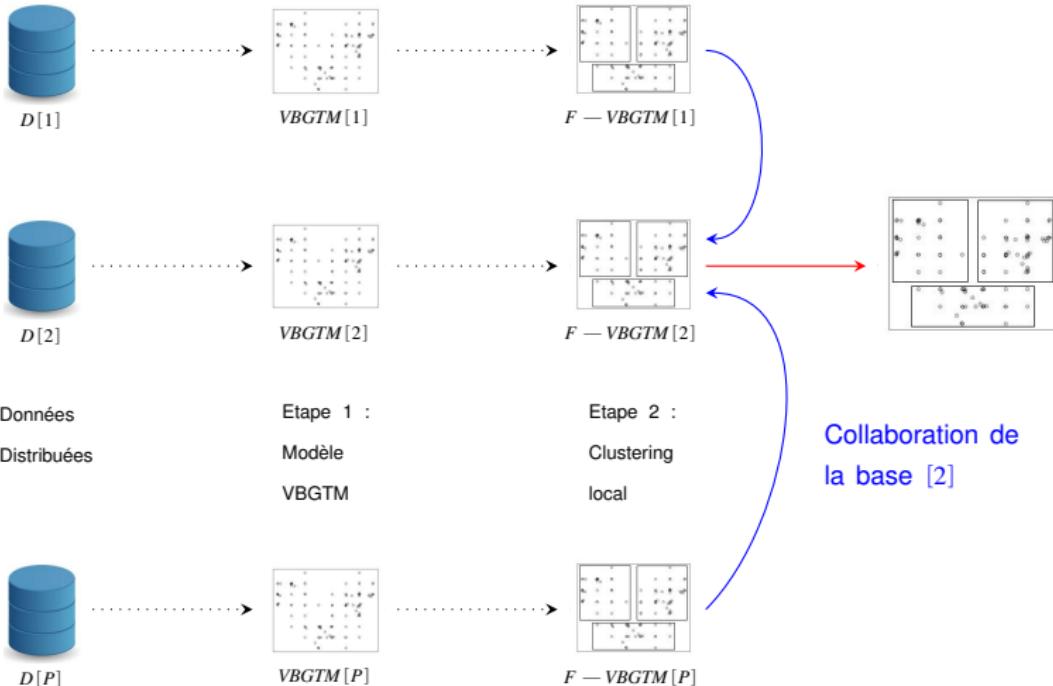
Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Inférence
Variationnelle
VB-GTM
Comparaison
F-VBGTM

Collaboration à
base de
VBGTM

Bilan &
Perspectives



Résultats de collaboration horizontale, base Waveform

Contexte
Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Inférence
Variationnelle

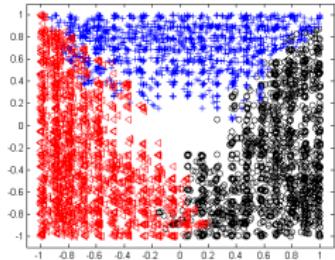
VB-GTM

Comparaison
F-VBGTM

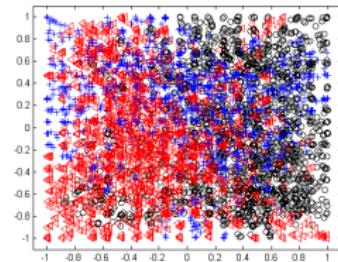
Collaboration à
base de
VBGTM

Bilan &
Perspectives

Variables pertinentes



Variable bruitées



Résultats de collaboration horizontale, base Waveform

Contexte
Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Inférence
Variationnelle

VB-GTM

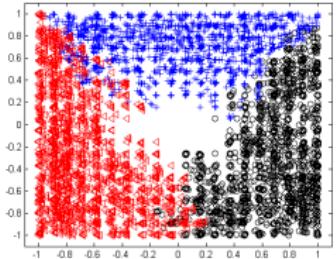
Comparaison

F-VBGTM

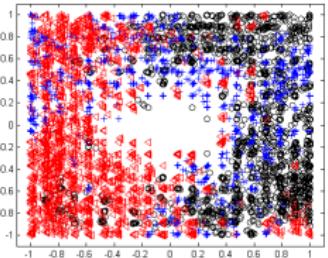
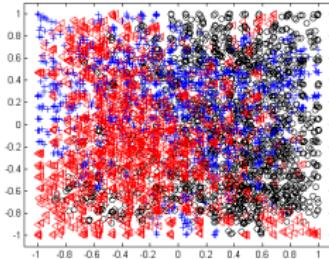
Collaboration à
base de
VBGTM

Bilan &
Perspectives

Variables pertinentes



Variable bruitées



Résultats de collaboration horizontale, base Waveform

Contexte
Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Inférence
Variationnelle

VB-GTM

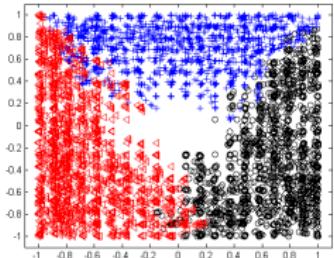
Comparaison

F-VBGTM

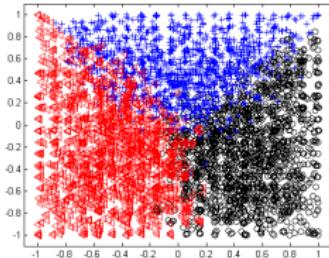
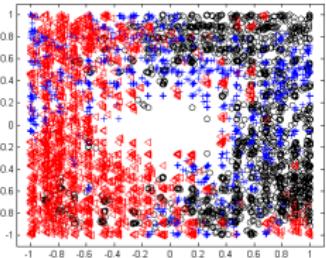
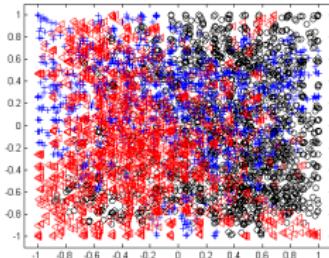
Collaboration à
base de
VBGTM

Bilan &
Perspectives

Variables pertinentes



Variable bruitées



Plan de la présentation

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

1 Contexte

2 Clustering Collaboratif à base de Prototypes

3 Collaboration sous cadre Déterministe

4 Collaboration sous cadre Probabiliste

5 Collaboration sous cadre Bayésien

6 Bilan & Perspectives

Contexte

Clustering

Collaboratif

à base de

Prototypes

Collaboration

sous cadre

Déterministe

Collaboration

sous cadre

Probabiliste

Collaboration

sous cadre

Bayésien

Bilan &

Perspectives

- Nouveaux algorithmes de Clustering Collaboratif.
- Deux approches : Horizontale et Verticale.
- Calcul automatique des coefficients de collaboration.
- Algorithmes basés sur des approches à base de prototypes :
 - > Approche déterministe.
 - > Approche probabiliste.
 - > Approche variationnelle.

*Contexte**Clustering Collaboratif à base de Prototypes**Collaboration sous cadre Déterministe**Collaboration sous cadre Probabiliste**Collaboration sous cadre Bayésien**Bilan & Perspectives*

- Faire collaborer plusieurs méthodes de clustering.
- Considérer que le nombre de clusters est différent entre les bases.
- Proposer un algorithme de collaboration “en-ligne”.
- Fusionner les cartes (SOM, GTM, VBGTM) après la collaboration pour construire un consensus.
- Etudier l'impact de la diversité entre les différents modèles sur la qualité de la collaboration.
- Utiliser la diversité pour guider la collaboration et la rendre sélective.

*Contexte**Clustering Collaboratif
à base de Prototypes**Collaboration sous cadre Déterministe**Collaboration sous cadre Probabiliste**Collaboration sous cadre Bayésien**Bilan & Perspectives*

Nous travaillons sur certaines de ces perspectives dans le cadre du projet ANR COCLICO⁽⁴⁾, en collaboration avec : ICUBE⁽⁵⁾, AgroParisTech⁽⁶⁾, LIVE⁽⁷⁾ et Espace Dev⁽⁸⁾.



(4). COLlaboration, CLassification, Incrémentalité et COnnaissances.

(5). Laboratoire des sciences de l'Ingénieur, de l'Informatique et de l'Imagerie, Université de Strasbourg

(6). Institut des sciences et industries du vivant et de l'environnement

(7). Laboratoire Image, Ville, Environnement, Université de Strasbourg

(8). L'espace au service du développement, Université Montpellier 2

Contexte

Clustering
Collaboratif
à base de
Prototypes

Collaboration
sous cadre
Déterministe

Collaboration
sous cadre
Probabiliste

Collaboration
sous cadre
Bayésien

Bilan &
Perspectives

Revues :

- *Collaborative Clustering Using Prototype-Based Techniques.* International Journal of Computational Intelligence and Applications 11, 03 (2012), 1250017.
- *Collaborative Fuzzy Clustering of Variational Bayesian GTM.* International Journal of Computational Intelligence and Applications (2013), submitted.

Conférences :

- *Learning Confidence Exchange in Collaborative Clustering.* In Neural Networks (IJCNN), The 2011 International Joint Conference on (2011), pp. 872–879.
- *Collaborative Generative Topographic Mapping.* In Neural Information Processing, vol. 7664 of Lecture Notes in Computer Science. Proc of ICONIP'12. Springer Berlin Heidelberg, 2012, pp. 591–598.(2012)
- *Collaborative Multi-View Clustering.* In Neural Networks (IJCNN), The 2013 International Joint Conference on (2013), pp. 872–879.
- *Apprentissage de la confiance des échanges en classification collaborative non supervisée,* in Proc. of CAP, Conférence Francophone d'Apprentissage , Chambéry, (2011).