

Contributions aux communications multi-vues pour l'apprentissage collaboratif

Denis Maurel

10 Décembre 2018

Mme. Raja CHIKY	Directrice de thèse
M. Antoine CORNUÉJOLS	Examineur
Mme. Pascale KUNTZ	Rapporteuse
M. Sylvain LEFEBVRE	Encadrant
M. Christophe MARSALA	Président du jury
M. Marcílio DE SOUTO	Examineur
M. Jérémie SUBLIME	Encadrant
Mme. Rosanna VERDE	Rapporteuse



Introduction

Clustering collaboratif incrémental

Optimisation de paramètres pour le clustering collaboratif

Système de reconstruction collaboratif

Contributions et Perspectives

Introduction

L'apprentissage machine (ou Machine Learning en anglais) est le domaine de recherche regroupant les algorithmes et les méthodes permettant **d'apprendre automatiquement** un résultat à partir d'un ensemble de données, aussi appelés **individus**. Les trois principales problématiques du Machine Learning sont:

L'apprentissage machine (ou Machine Learning en anglais) est le domaine de recherche regroupant les algorithmes et les méthodes permettant **d'apprendre automatiquement** un résultat à partir d'un ensemble de données, aussi appelés **individus**. Les trois principales problématiques du Machine Learning sont:

- **La classification:** apprentissage des correspondances entre une donnée et son label.

L'apprentissage machine (ou Machine Learning en anglais) est le domaine de recherche regroupant les algorithmes et les méthodes permettant **d'apprendre automatiquement** un résultat à partir d'un ensemble de données, aussi appelés **individus**. Les trois principales problématiques du Machine Learning sont:

- **La classification**: apprentissage des correspondances entre une donnée et son label.
- **Le clustering**: détection de groupes d'individus similaires.

L'apprentissage machine (ou Machine Learning en anglais) est le domaine de recherche regroupant les algorithmes et les méthodes permettant **d'apprendre automatiquement** un résultat à partir d'un ensemble de données, aussi appelés **individus**. Les trois principales problématiques du Machine Learning sont:

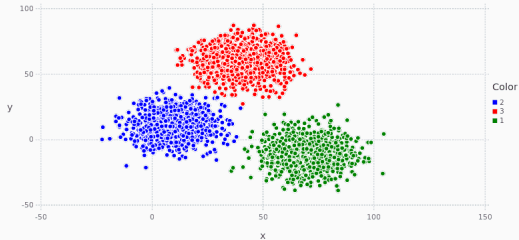
- **La classification:** apprentissage des correspondances entre une donnée et son label.
- **Le clustering:** détection de groupes d'individus similaires.
- **L'apprentissage par renforcement:** apprentissage d'un comportement permettant à un modèle de réagir à un environnement dynamique.

L'apprentissage machine (ou Machine Learning en anglais) est le domaine de recherche regroupant les algorithmes et les méthodes permettant **d'apprendre automatiquement** un résultat à partir d'un ensemble de données, aussi appelés **individus**. Les trois principales problématiques du Machine Learning sont:

- **La classification**: apprentissage des correspondances entre une donnée et son label.
- **Le clustering**: détection de groupes d'individus similaires.
- **L'apprentissage par renforcement**: apprentissage d'un comportement permettant à un modèle de réagir à un environnement dynamique.

Clustering

- Tâche d'apprentissage **non supervisée** consistant à rassembler des groupes d'individus (a.k.a. clusters) de sorte à **maximiser la similarité intra-groupe** et à **minimiser la similarité inter-groupes**.



Clustering: similarité entre individus

- La notion de **similarité** est souvent assimilée à la notion de **distance**.
- La similarité doit être **adaptée à la nature des données**.

Euclidienne	$\ a - b\ _2 = \sqrt{\sum_i (a_i - b_i)^2}$
Manhattan	$\ a - b\ _1 = \sum_i a_i - b_i $
Maximum	$\ a - b\ _\infty = \max_i a_i - b_i $
Mahalanobis	$\sqrt{(a - b)^T S^{-1} (a - b)}$
Hamming	$\text{Hamming}(a, b) = \sum_i (1 - \delta_{a_i, b_i})$

Table 1: Exemples de distances

Clustering: types de partitions

Après un clustering, on obtient une partition de l'espace ainsi que les appartenances des individus à chaque groupe de cette partition. Ces appartenances peuvent être **dures**, **molles** ou **floues**.

	c ₁	c ₂	c ₃
x ₁	1	0	0
x ₂	0	1	0
x ₃	0	0	1
x ₄	0	0	1

(a) Clustering dur

	c ₁	c ₂	c ₃
x ₁	1	1	0
x ₂	0	1	1
x ₃	0	0	1
x ₄	0	0	1

(b) Clustering mou

	c ₁	c ₂	c ₃
x ₁	0.9	0.1	0
x ₂	0	0.8	0.2
x ₃	0	0.3	0.7
x ₄	0	0	1.0

(c) Clustering flou

Table 2: Les trois principaux types d'appartenances à des clusters

On peut regrouper les algorithmes de clustering en sous-catégories suivant l'approche qu'ils utilisent:

- **Méthodes hiérarchiques:** création d'un arbre de correspondance entre les individus (Agglomerative method).
- **Méthodes de quantification de vecteurs:** définition d'individus prototypes pour synthétiser les individus en entrée (K-Means, SOM).
- **Méthodes de densité:** estimation des clusters suivant les zones les plus densément peuplées de l'espace d'entrée (DBSCAN).
- **Méthodes stochastiques:** création de modèles probabilistes définissant la probabilité d'appartenance d'un individu à un cluster donné (GMM).

On peut regrouper les algorithmes de clustering en sous-catégories suivant l'approche qu'ils utilisent:

- **Méthodes hiérarchiques:** création d'un arbre de correspondance entre les individus (Agglomerative method).
- **Méthodes de quantification de vecteurs:** définition d'individus prototypes pour synthétiser les individus en entrée (K-Means, SOM).
- **Méthodes de densité:** estimation des clusters suivant les zones les plus densément peuplées de l'espace d'entrée (DBSCAN).
- **Méthodes stochastiques:** création de modèles probabilistes définissant la probabilité d'appartenance d'un individu à un cluster donné (GMM).

Apparition d'un nouveau contexte: **un même ensemble d'individu** est décrit dans plusieurs base de données indépendantes appelées **vues**.

Problème: comment obtenir un clustering de cet ensemble d'individus ?

Deux approches:

Apparition d'un nouveau contexte: **un même ensemble d'individu** est décrit dans plusieurs base de données indépendantes appelées **vues**.

Problème: comment obtenir un clustering de cet ensemble d'individus ?

Deux approches:

- **Coopérative:** chaque vue effectue un clustering de ses données locales avant de transférer ses résultats à une entité tiers qui devra fusionner les résultats.

Apparition d'un nouveau contexte: **un même ensemble d'individu** est décrit dans plusieurs base de données indépendantes appelées **vues**.

Problème: comment obtenir un clustering de cet ensemble d'individus ?

Deux approches:

- **Coopérative:** chaque vue effectue un clustering de ses données locales avant de transférer ses résultats à une entité tiers qui devra fusionner les résultats.
- **Collaborative:** chaque vue effectue un premier clustering local, puis le modifie en fonction des résultats obtenus par les autres vues.

Apparition d'un nouveau contexte: **un même ensemble d'individu** est décrit dans plusieurs base de données indépendantes appelées **vues**.

Problème: comment obtenir un clustering de cet ensemble d'individus ?

Deux approches:

- **Coopérative:** chaque vue effectue un clustering de ses données locales avant de transférer ses résultats à une entité tiers qui devra fusionner les résultats.
- **Collaborative:** chaque vue effectue un premier clustering local, puis le modifie en fonction des résultats obtenus par les autres vues.

Clustering collaboratif: définition

Le **clustering collaboratif** est un domaine récent désignant l'ensemble des méthodes permettant à **plusieurs algorithmes de clustering** opérant sur des **sources de données différentes** de collaborer pour **améliorer localement** leurs résultats.

- Les algorithmes utilisés peuvent être **différents**.
- Les vues doivent partager soit leurs descripteurs (clustering vertical), soit **leurs individus** (clustering horizontal) pour pouvoir être comparées.

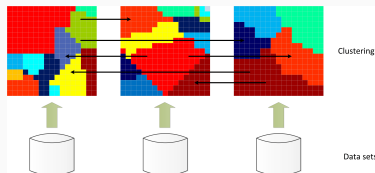


Figure 2: Illustration du principe de clustering collaboratif horizontal

Clustering collaboratif: processus

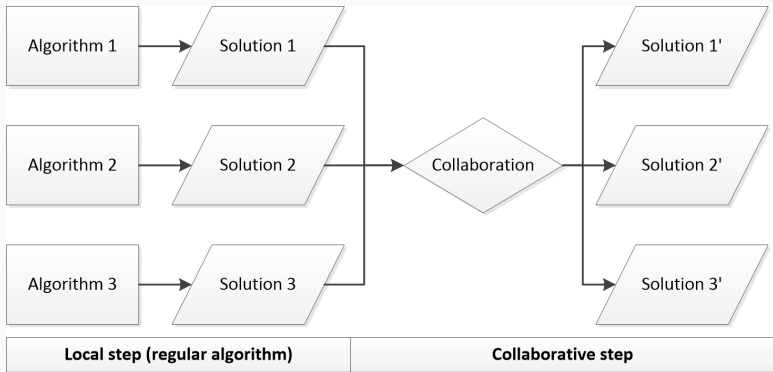


Figure 3: Processus de clustering collaboratif

Définition du critère collaboratif:

$$Q^i = \alpha_i Q_{\text{local}}^i(v_i) + Q_{\text{collab}}^i(v_i, v_{j \neq i}) \quad (1)$$

$$= \alpha_i Q_{\text{local}}^i(v_i) + \sum_{j \neq i} \beta_j^i C_j^i(v_i, v_j) \quad (2)$$

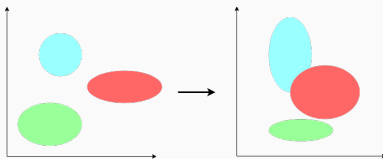
- Q_{local}^i est généralement basé sur **le critère de l'algorithme local** à optimiser.
- Q_{collab}^i se base sur **l'échange d'information entre vues**, typiquement les appartenances des individus aux clusters respectifs de chaque vue. C_j^i représente **la dissimilarité** entre les vues i et j .
- α_i et β_j^i sont définis **à la main**. L'approximation $\forall j, \beta_j^i = \alpha_i^2$ est parfois utilisée car donnant de bons résultats en pratique.

Clustering collaboratif incrémental

Objectif du clustering collaboratif: définir un ensemble de clusters suivant **la distribution** des données fournies en entrée.

Problème: il arrive que cette distribution **évolue au cours du temps**

Exemple: évolution du régime alimentaire d'un individu ou de la répartition des salaires à l'échelle d'une population.



Utilisation du clustering **incrémental**: les clusters sont appris au cours du temps afin de **s'adapter** aux éventuels changements de distribution. On utilise les N_{batch} derniers individus comme échantillon d'apprentissage.

Définition d'une méthode de clustering incrémental:

Définition d'une méthode de clustering incrémental:

- Choix de la **méthode de clustering**

Définition d'une méthode de clustering incrémental:

- Choix de la **méthode de clustering**
- **Adaptation de la méthode de clustering** pour de l'apprentissage incrémental

Définition d'une méthode de clustering incrémental:

- Choix de la **méthode de clustering**
- **Adaptation de la méthode de clustering** pour de l'apprentissage incrémental
- **Adaptation du clustering collaboratif** au modèle de clustering obtenu

Cartes Auto-Adaptatrices (SOM)

Dans notre cas, utilisation des **cartes auto-adaptatrices de Kohonen** (ou Self-Organizing Maps (SOM) en anglais) comme méthode de clustering.

- Méthode à base de **prototypes** (quantification de vecteurs)
- Permet la **visualisation** de données en hautes dimensions
- Notion de **voisinage**: utilisation d'une **fonction de température**.

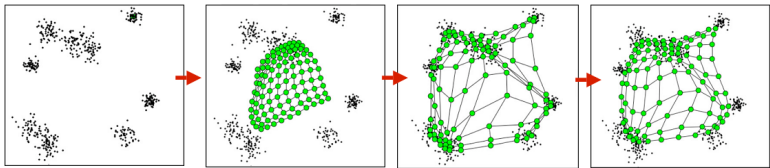
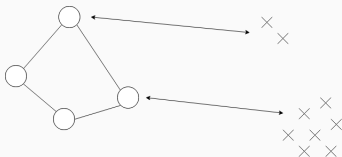


Figure 4: Exemple de SOM

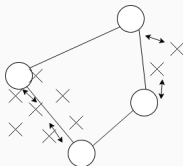
Cartes Auto-Adaptatrices (SOM)

Fonction de voisinage: $K_{i,j} = \exp \left(-\frac{d_1^2(i,j)}{\lambda(t)} \right)$

Fonction de température: $\lambda(t) = \lambda_{\min} \left(\frac{\lambda_{\max}}{\lambda_{\min}} \right)^{\frac{1}{t}}$



(a) Température élevée



(b) Température faible

Les solutions existantes pour les SOM incrémentales se basent toutes sur l'ajout de prototypes ([Papilinski2012] et [Deng2000]).

Les solutions existantes pour les SOM incrémentales se basent toutes sur l'ajout de prototypes ([Papinski2012] et [Deng2000]).

Problème 1: Est-il possible de trouver une solution **conservant la topologie de la carte** ?

Les solutions existantes pour les SOM incrémentales se basent toutes sur l'ajout de prototypes ([Papinski2012] et [Deng2000]).

Problème 1: Est-il possible de trouver une solution **conservant la topologie de la carte** ?

Problème 2: λ est **dépendante du temps**, ce qui pose problème avec le clustering incrémental.

Les solutions existantes pour les SOM incrémentales se basent toutes sur l'ajout de prototypes ([Papilinski2012] et [Deng2000]).

Problème 1: Est-il possible de trouver une solution **conservant la topologie de la carte** ?

Problème 2: λ est **dépendante du temps**, ce qui pose problème avec le clustering incrémental.

→ rendre la fonction dépendante des **individus**

$$\lambda(t) = \lambda_{\min} \left(\frac{\lambda_{\max}}{\lambda_{\min}} \right)^{\frac{1}{t}} \rightarrow \tilde{\lambda}(B, W) = \frac{1}{N_{\text{batch}}} \sum_{i=1}^{N_{\text{batch}}} \|x_i - \chi(x_i)\|_2$$
$$K_{i,j}(\lambda(t)) \rightarrow K_{i,j}(\tilde{\lambda}(B, W)) \rightarrow \tilde{K}_{i,j}$$

- **1^{ère} contrainte:** des individus correspondants doivent appartenir à des **prototypes correspondants** ou à leurs voisinages proches.

- **1^{ère} contrainte:** des individus correspondants doivent appartenir à des **prototypes correspondants** ou à leurs voisinages proches.
- **2^{ème} contrainte: même topologie des cartes** pour toutes les vues pour les rendre comparables.

- **1^{ère} contrainte:** des individus correspondants doivent appartenir à des **prototypes correspondants** ou à leurs voisinages proches.
- **2^{ème} contrainte: même topologie des cartes** pour toutes les vues pour les rendre comparables.

$$\begin{aligned} & Q_{\text{local}}^m(K_{i,j}) \quad / \quad Q_{\text{collab}}^m(K_{i,j}^m - K_{i,j}^{m'}) \\ \Rightarrow & Q_{\text{local}}^m(\tilde{K}_{i,j}) \quad / \quad Q_{\text{collab}}^m(\tilde{K}_{i,j}^m - \tilde{K}_{i,j}^{m'}) \\ \Rightarrow & \tilde{Q}_{\text{local}}^m \quad / \quad \tilde{Q}_{\text{collab}}^m \end{aligned} \quad K_{i,j} = \exp \left(-\frac{d_1^2(i,j)}{\lambda(t)} \right)$$

Le nouveau critère dépendant uniquement des N_{batch} derniers individus apparus, il est possible d'effectuer un **apprentissage collaboratif incrémental** sur l'ensemble des vues.

Test de notre méthode sur 4 jeux de données différents

- Spambase
- Waveform
- Wisconsin Breast Cancer Diagnosis (WDBC)
- Isolet

Pureté d'un prototype: classe majoritaire parmi les individus associés à ce prototype.

Pureté d'une SOM: pureté moyenne de ses prototypes.

Erreur moyenne de quantification: distance moyenne entre un prototype et les individus qui y sont associés.

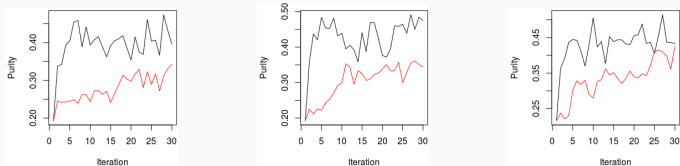
$$qe = \frac{1}{N_{batch}} \sum_{i=1}^{N_{batch}} \|x_i - \chi(x_i)\|^2$$

Table 3: Erreur de quantification moyenne pour chaque base de donnée.

	SOM incrémentales	clustering collaboratif incrémental	Gain
Spam Base	0.223	0.203	+ 8.9%
Waveform	0.197	0.24	- 21.8%
WDBC	0.183	0.18	+ 1,6%
Isolet	2.61	1.34	+ 48,7%

- **Score en quantification à long termes globalement comparables:** le clustering collaboratif n'endommage pas les résultats locaux.
- Le clustering collaboratif est sensible **à la présence de vues bruitées.**

Figure 6: Évolution des puretés pour Isolet. Les lignes rouges représentent les SOM incrémentales tandis que les lignes noires représentent les SOM incrémentales collaboratives. Chaque itération correspond à l'arrivée d'une nouvelle donnée

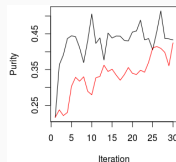


(a) $N_{\text{batch}} = 10$, View 1

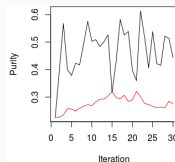
(b) $N_{\text{batch}} = 10$, View 2

(c) $N_{\text{batch}} = 10$, View 3

- Les SOM incrémentales collaboratives **apprennent plus vite que les SOM incrémentales seules**: exploitation de l'information via le partage.
- Les SOM collaboratives sont globalement **plus instables** que les SOM incrémentales.



(a) $N_{\text{batch}} = 10$, View 3



(b) $N_{\text{batch}} = 3$, View 3

Figure 8: Évolution des puretés pour Isolet: impact de la taille du batch sur la stabilité de l'apprentissage.

- La stabilité de l'apprentissage **augmente avec la taille du batch**: plus d'informations à exploiter.

Contributions

- Définition d'une **nouvelle méthode de cartes auto-adaptatives incrémentales**
- Adaptation de la méthode précédente au **clustering collaboratif**
- Analyse de l'impact de **la taille du batch sur la stabilité de l'apprentissage**.

Contributions

- Définition d'une **nouvelle méthode de cartes auto-adaptatives incrémentales**
- Adaptation de la méthode précédente au **clustering collaboratif**
- Analyse de l'impact de **la taille du batch sur la stabilité de l'apprentissage**.

Comment déterminer quelles sont les meilleures collaborations inter-vues ?

Optimisation de paramètres pour le clustering collaboratif

Rappel du critère du clustering collaboratif:

$$\begin{aligned} Q^i &= \alpha_i Q_{\text{local}}^i(v_i) + Q_{\text{collab}}^i(v_i, v_{j \neq i}) \\ &= \alpha_i Q_{\text{local}}^i(v_i) + \sum_{j \neq i} \beta_j^i C_j^i(v_i, v_j) \end{aligned}$$

Objectif: Apprendre automatiquement les α et les β en s'affranchissant de simplifications telles que $\beta = \alpha^2$.

Nous avons proposé **une nouvelle méthode de pondération** permettant d'apprendre automatiquement les poids fixant **les importances relatives des différents scores**.

La définition des α et des β passe par la résolution d'un problème sous contrainte. Les contraintes sont les suivantes :

La définition des α et des β passe par la résolution d'un problème sous contrainte. Les contraintes sont les suivantes :

$$\bullet \beta^* = \operatorname{argmin}_{\beta} \sum_{i \neq j} \beta_{i,j} c_{i,j}$$

La définition des α et des β passe par la résolution d'un problème sous contrainte. Les contraintes sont les suivantes :

- $\beta^* = \operatorname{argmin}_{\beta} \sum_{i \neq j} \beta_{i,j} c_{i,j}$
- $\forall j \quad \prod_{i \neq j} \beta_{i,j} = 1$

La définition des α et des β passe par la résolution d'un problème sous contrainte. Les contraintes sont les suivantes :

- $\beta^* = \operatorname{argmin}_{\beta} \sum_{i \neq j} \beta_{i,j} c_{i,j}$
- $\forall j \quad \prod_{i \neq j} \beta_{i,j} = 1$
- $\forall (i, j) \quad \beta_{i,j} > 0$

La définition des α et des β passe par la résolution d'un problème sous contrainte. Les contraintes sont les suivantes :

- $\beta^* = \operatorname{argmin}_{\beta} \sum_{i \neq j} \beta_{i,j} c_{i,j}$
- $\forall j \quad \prod_{i \neq j} \beta_{i,j} = 1$
- $\forall (i, j) \quad \beta_{i,j} > 0$
- Affranchissement des α : comme il s'agit de pondérations relatives, on peut fixer $\alpha = 1$.

La définition des α et des β passe par la résolution d'un problème sous contrainte. Les contraintes sont les suivantes :

- $\beta^* = \operatorname{argmin}_{\beta} \sum_{i \neq j} \beta_{i,j} c_{i,j}$
- $\forall j \quad \prod_{i \neq j} \beta_{i,j} = 1$
- $\forall (i, j) \quad \beta_{i,j} > 0$
- Affranchissement des α : comme il s'agit de pondérations relatives, on peut fixer $\alpha = 1$.

Objectif: mettre des poids plus élevés sur les meilleurs accords, tout en gardant une certaine partie des vues en désaccord pour faire évoluer l'information

Critère + contrainte d'égalité + contrainte d'inégalité

→ **méthode de Karush-Kuhn-Tucker (KKT)**

$$\mathcal{L}(\beta, \nu, \lambda) = \sum_{j=1}^J \sum_{i \neq j}^J (\beta_{i,j} c_j^i - \nu_j \ln \beta_{i,j} - \lambda_{i,j} \beta_{i,j}).$$

$$\frac{\partial \mathcal{L}}{\partial \beta_{i,j}} = 0, \quad \frac{\partial \mathcal{L}}{\partial \nu} = 0, \quad \frac{\partial \mathcal{L}}{\partial \lambda} = 0$$

Après résolution, on obtient:

$$\forall(i,j), \quad i \neq j \quad \beta_{i,j} = \frac{(\prod_{k \neq j} c_k^i)^{\frac{1}{I-1}}}{c_j^i}$$

L'importance d'une vue externe relativement à une vue locale est proportionnelle au rapport entre la moyenne géométrique de toutes les dissimilarités sur la dissimilarité entre les deux vues .

Après résolution, on obtient:

$$\forall (i, j), \quad i \neq j \quad \beta_{i,j} = \frac{(\prod_{k \neq j} c_k^i)^{\frac{1}{j-1}}}{c_j^i}$$

L'importance d'une vue externe relativement à une vue locale est proportionnelle au rapport entre la moyenne géométrique de toutes les dissimilarités sur la dissimilarité entre les deux vues .

→ **plus 2 vues sont similaires, plus elles collaboreront.**

Deux axes d'analyse:

- Analyse de **l'évolution du critère collaboratif** avec et sans apprentissage automatique des poids (critère: différence relative).
- Analyse des **valeurs prises par les β** en fin d'apprentissage. (critère: différences relatives entre les poids).

Deux axes d'analyse:

- Analyse de **l'évolution du critère collaboratif** avec et sans apprentissage automatique des poids (critère: différence relative).
- Analyse des **valeurs prises par les β** en fin d'apprentissage. (critère: différences relatives entre les poids).

Ce que l'on s'attend à trouver:

- Une **amélioration de la valeur du critère collaboratif** grâce à notre méthode
- **L'identification automatique des vues bruitées**, menant à des valeurs de β proches de 0.

Deux axes d'analyse:

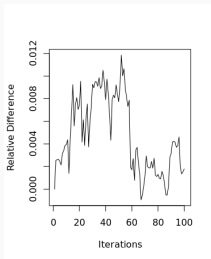
- Analyse de **l'évolution du critère collaboratif** avec et sans apprentissage automatique des poids (critère: différence relative).
- Analyse des **valeurs prises par les β** en fin d'apprentissage. (critère: différences relatives entre les poids).

Ce que l'on s'attend à trouver:

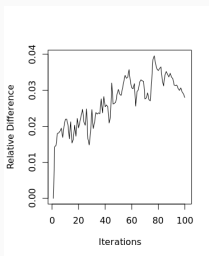
- Une **amélioration de la valeur du critère collaboratif** grâce à notre méthode
- **L'identification automatique des vues bruitées**, menant à des valeurs de β proches de 0.

À noter: pour chaque base de données, **une vue entièrement composée de bruit** a été rajoutée pour tester le second point.

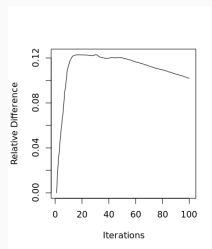
Figure 9: Différences relatives du critère collaboratif avec et sans optimisation des β tout au long du processus d'apprentissage.



(a) Spambase



(b) Isolet



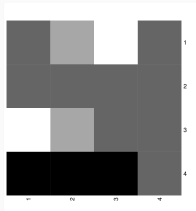
(c) VHR Strasbourg

- **Amélioration du critère** de manière significative dans la majorité des cas
- Tous les jeux de données **ne sont pas traités de la même manière** (dépend de la quantité et de la qualité de l'informations à partager).

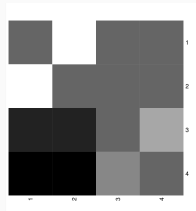
Figure 11: Cartes de chaleur des matrices de β pour chaque jeu de données.

$M(i, j)$ correspond à l'importance accordée à la Vue j par la Vue i .

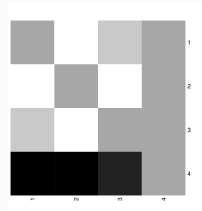
Blanc = forte collaboration - noir = faible collaboration - diagonale représente $\beta = 1$.



(a) WDBC



(b) Waveform



(c) Isolet

- **Identification des vues bruitées.**
- Apparition de **méta-clusters** (des clusters de clusters). Les vues auront tendance à se regrouper en sous-groupe mutuellement d'accord.

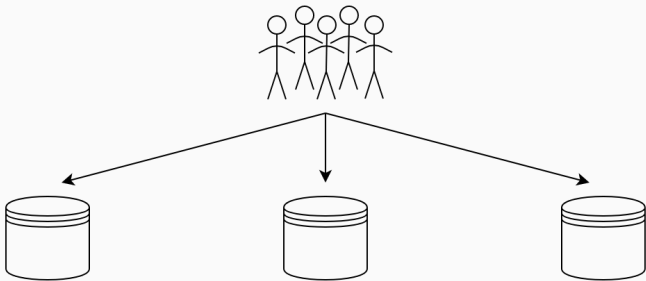
- Proposition d'un **système de pondération automatique des vues externes** pour le clustering collaboratif.
- **Amélioration** de la valeur du critère collaboratif
- **Identification des vues bruitées.**
- **(BONUS)** Apparition de **méta-clusters** de vues

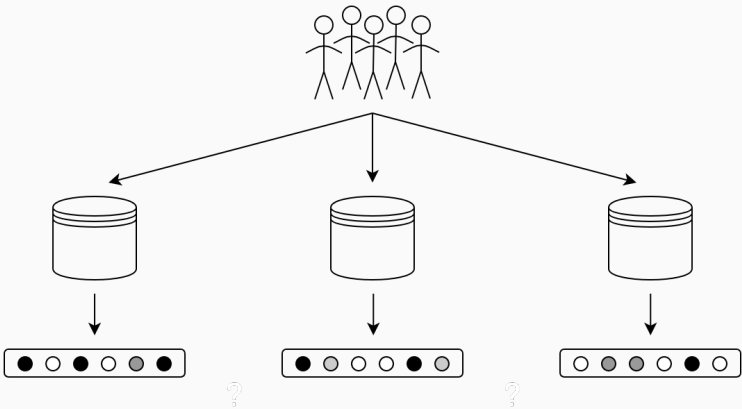
- Proposition d'un **système de pondération automatique des vues externes** pour le clustering collaboratif.
- **Amélioration** de la valeur du critère collaboratif
- **Identification des vues bruitées.**
- **(BONUS)** Apparition de **méta-clusters** de vues

**Est-il possible d'étendre l'idée de collaboration
à d'autres problèmes que le clustering ?**

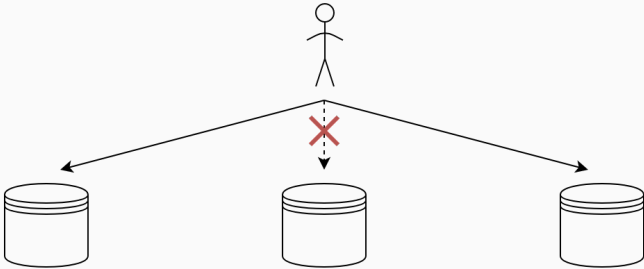
Système de reconstruction collaboratif

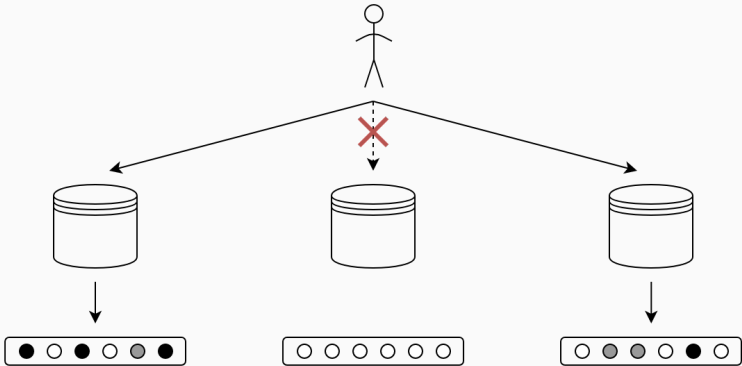


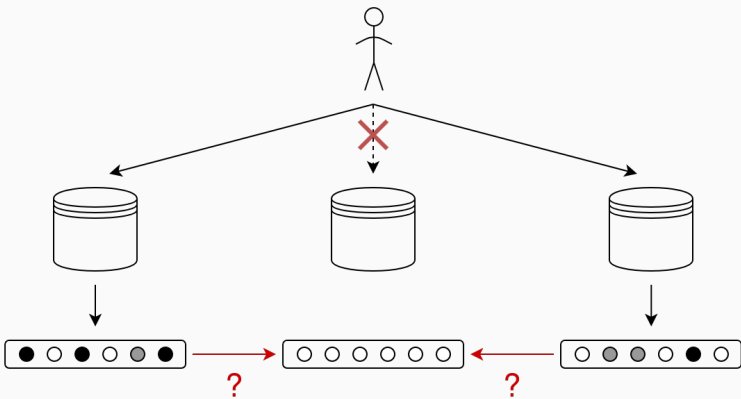












Reconstruction collaborative: définition

Comment définir le problème de **reconstruction collaborative** ?

- Application du **paradigme collaboratif** à un autre problème
- Le clustering collaboratif nécessite la **description d'un même individu dans chacune des vues**
- En pratique, les données souvent sont **soit manquantes, soit incomplètes**.

?			?		?						
		?		?			?		?		
	?	?	?	?			?				?
?					?				?	?	

La reconstruction collaborative devrait permettre d'obtenir **une approximation** d'un individu dans une vue connaissant **l'information présente dans les autres vues**.

1. **Compresser et anonymiser** les données
2. **Transférer de l'information** d'une vue externe à une vue locale
3. **Combiner les informations** en provenance de différentes sources

1. **Compresser et anonymiser** les données
2. **Transférer de l'information** d'une vue externe à une vue locale
3. **Combiner les informations** en provenance de différentes sources

Ce que nous proposons:

1. **Des réseaux de neurones** (autoencodeurs).

1. **Compresser et anonymiser** les données
2. **Transférer de l'information** d'une vue externe à une vue locale
3. **Combiner les informations** en provenance de différentes sources

Ce que nous proposons:

1. **Des réseaux de neurones** (autoencodeurs).
2. **D'autres réseaux de neurones** (perceptrons multi-couches).

1. **Compresser et anonymiser** les données
2. **Transférer de l'information** d'une vue externe à une vue locale
3. **Combiner les informations** en provenance de différentes sources

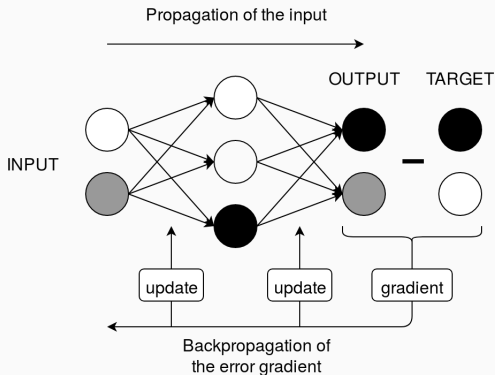
Ce que nous proposons:

1. **Des réseaux de neurones** (autoencodeurs).
2. **D'autres réseaux de neurones** (perceptrons multi-couches).
3. **Une nouvelle méthode de combinaison.**

- Réseau de neurone:
**approximateur universel
de fonction.**

- **Backpropagation:**
Méthode utilisée pour
**mettre à jour des
paramètres.**

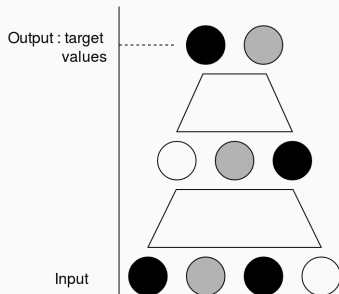
- L'erreur est propagée **de
la sortie vers l'entrée** du
réseau.



Réseaux de neurones: MLP et autoencodeurs

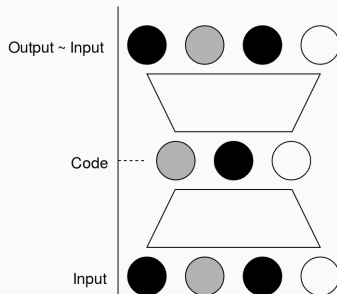
Perceptron multi-couches

- entrées et sorties **différentes**
- apprentissage **supervisé**



Autoencodeur

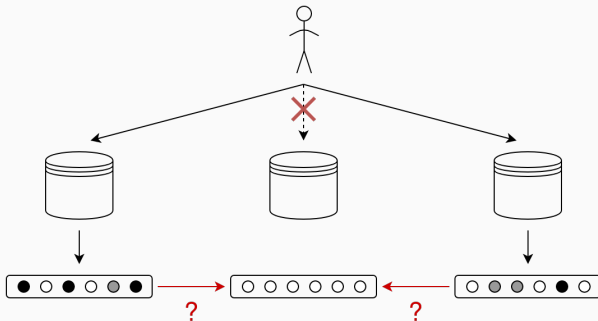
- entrées et sorties **identiques**
- apprentissage **non supervisé**



Reconstruction collaborative: rappel

Rappel de la problématique

- Un individu décrit dans **toutes les vues sauf une**
- Utilisation des informations externes pour obtenir **une approximation locale**



Reconstruction collaborative: système

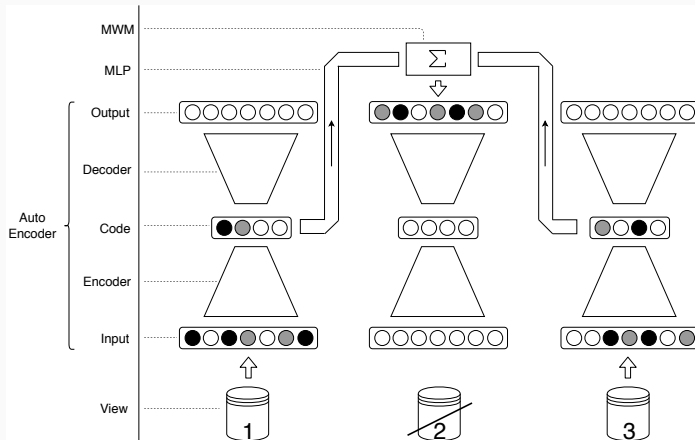


Figure 13: Architecture d'un système de reconstruction collaborative

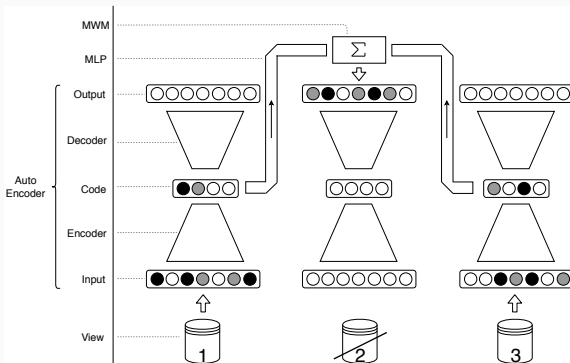
Reconstruction collaborative: système

Autoencodeurs

- Un par vue: encoder chaque espace
- Rend difficile la reconstruction des données originales sans le décodeur

Perceptron multi-couches

- Décode les représentations externes
- Passe d'un espace codé à l'espace local



- Pour chaque individu et avec N vues, on peut avoir jusqu'à $N - 1$ **reconstructions différentes**.
- Comment les combiner ?

- Pour chaque individu et avec N vues, on peut avoir jusqu'à $N - 1$ **reconstructions différentes**.
- Comment les combiner ?
- Habituellement, combinaison **pondérée** des différentes reconstructions.
 - Un scalaire par vue
 - Différentes façons d'apprendre les poids

- Pour chaque individu et avec N vues, on peut avoir jusqu'à $N - 1$ **reconstructions différentes**.
- Comment les combiner ?
- Habituellement, combinaison **pondérée** des différentes reconstructions.
 - Un scalaire par vue
 - Différentes façons d'apprendre les poids
- MAIS hypothèse forte: chaque vue contient exactement **la même information** sur chaque descripteur → irréaliste.
- À la place d'un poids unique, nous utilisons un **masque**.

Reconstruction collaborative: combinaison

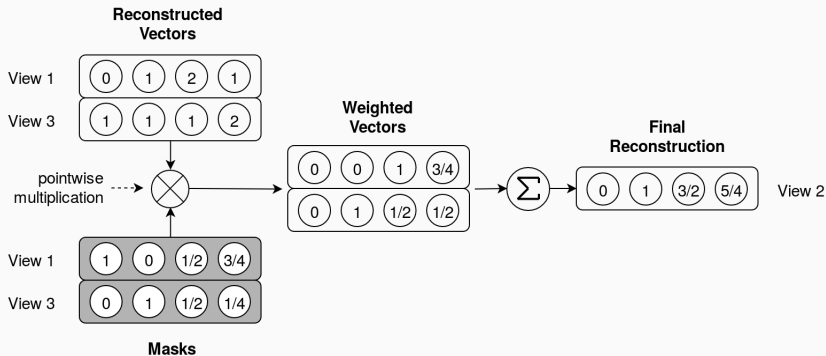
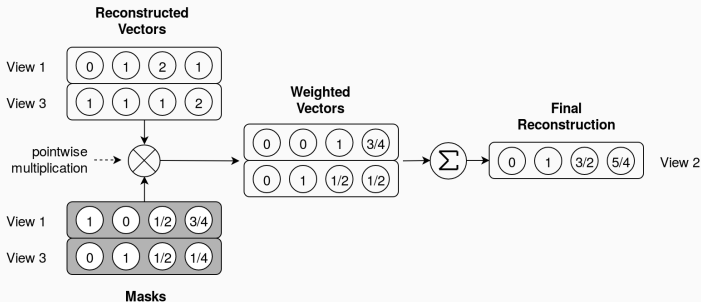


Figure 14: Pondération par masque (Masked Weighting Method en anglais).

Reconstruction collaborative: combinaison

- Chaque vue possède $N - 1$ masques, un par vue externe, utilisés pour **pondérer chaque représentation externe**.
- Chaque masque est entraîné de manière itérative
- **Avantage:** les masques peuvent se concentrer sur **les parties les mieux reconstruites** par la vue externe.



Idée de base pour la mise à jour des poids

$$E_i = \frac{1}{|V_i|} \sum_{x_i \in V_i} \|x_i - \tilde{x}_i\|^2 \quad \text{puis} \quad \frac{\partial E_i}{\partial w_{i|j}^k} = 0$$

- E_i = erreur de la i -ème vue
- V_i = i -ème vue
- x_i = individu de V_i
- $w_{i|j}^k$ = k -ème coordonnée du masque attribué à la j -ème vue

Après calcul:

$$w_{ij}^k = \frac{\sum_{x_i \in V_i} x_{ij}^k (x_i^k - \sum_{j' \in [1..N] \setminus \{i,j\}} w_{ij'}^k x_{ij'}^k)}{\sum_{x_i \in V_i} (x_{ij}^k)^2}$$

Après calcul:

$$w_{ij}^k = \frac{\sum_{x_i \in V_i} x_{ij}^k (x_i^k - \sum_{j' \in [1..N] \setminus \{i,j\}} w_{ij'}^k x_{ij'}^k)}{\sum_{x_i \in V_i} (x_{ij}^k)^2}$$

- La mise à jour d'un poids dépend de **tous les autres paramètres**
→ Définition d'une règle de mise à jour **itérative**.

- Entraînement **séquentiel**
 - les autoencodeurs pour **encoder** les données
 - les perceptrons multi couches pour **reconstruire** les individus
 - les masques pour **combiner** les échantillons reconstruits
- Une vue n'a **jamais** accès aux données originales de deux vues différentes.
 - **autoencodeurs**: données originales locales
 - **perceptrons**: données externes encodées + données originales locales
 - **masques**: échantillons reconstruits + données originales locales

Reconstruction collaborative: entraînement

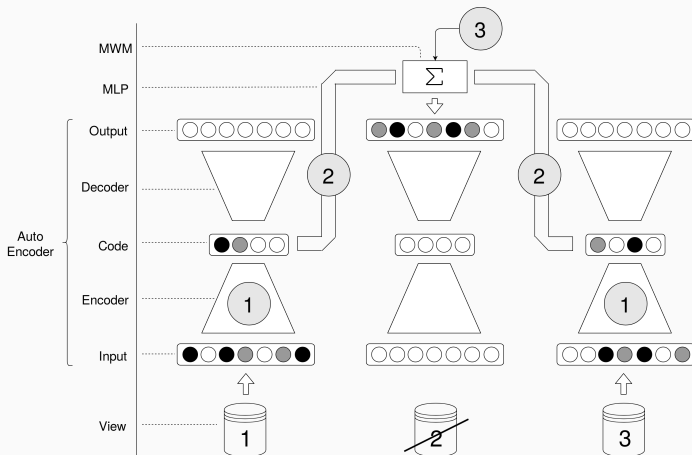


Figure 15: Apprentissage séquentiel des éléments du système

- Wisconsin Diagnostic Breast Cancer (WDBC)
 - 569 individus
 - 32 descripteurs répartis en 3 vues
- Multi-Features Digital Dataset (MFDD)
 - 2000 individus
 - 76 coefficients de Fourier
 - 216 correlations profile
 - 64 coefficients Karhunen-Love
 - **240 pixels moyennés en fenêtres de 2×3**
 - 47 moments de Zernike
- Madelon
 - 4400 individus
 - 20 (utiles) 480 (bruits) variables

- Distance quadratique moyenne

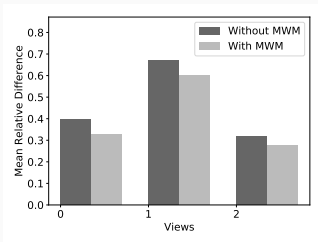
$$\text{MSE}(x, y) = \frac{1}{K} \sum_{i=1}^K (x_i - y_i)^2$$

- Différence relative moyenne

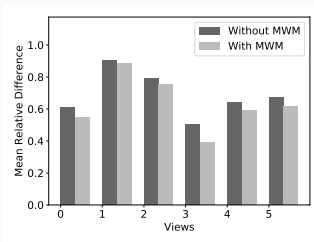
$$\text{MRD}(X, Y) = \frac{1}{K} \sum_{i=1}^K \left| \frac{x_i - y_i}{y_i} \right|$$

- Analyse visuelle d'images reconstruites (MFDD)

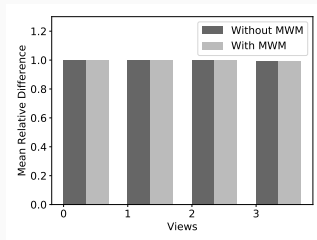
Reconstruction collaborative: résultats



(a) WDBC



(b) MFDD



(c) Madelon

Reconstruction collaborative: images originales

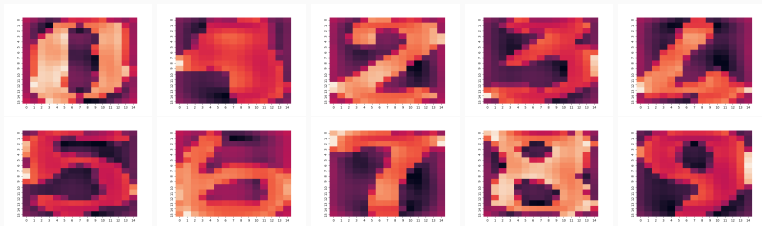


Figure 17: Échantillon d'images originales

Reconstruction collaborative: images reconstruites

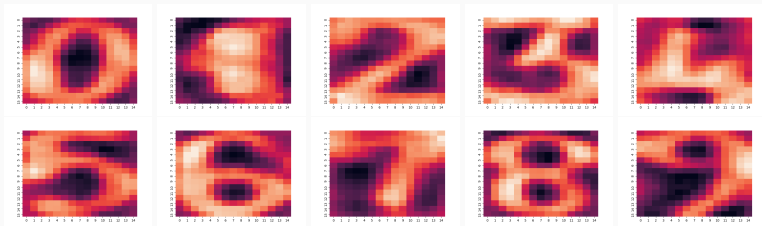


Figure 18: Échantillon d'images reconstruites de bonne qualité (90%)

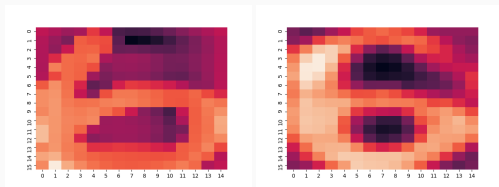


Figure 19: Capture de l'information intrinsèque.

Reconstruction collaborative: images améliorables

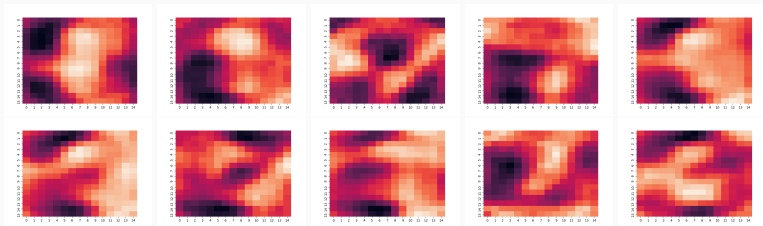
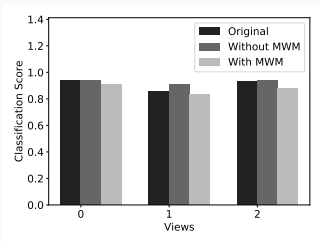


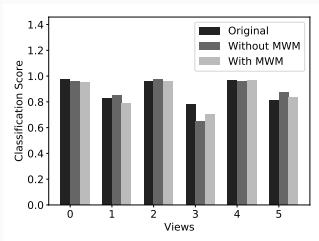
Figure 20: Échantillon d'images reconstruites de mauvaise qualité (10%)

- Les données sont reconstruites pour **ensuite** être utilisées
- Cas d'application: **la classification** (ici avec des Random Forests)
- Est-ce que les individus reconstruits sont utilisables pour des applications ultérieures ?
- **Critère: différence en classification:** différence entre le score utilisant les données originales et celui utilisant les données reconstruites.

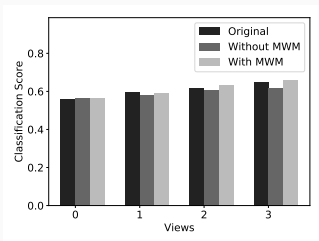
Reconstruction collaborative: un peu plus loin



(a) WDBC



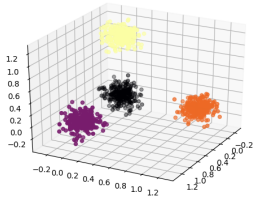
(b) MFDD



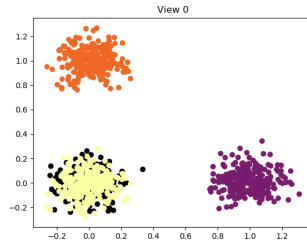
(c) Madelon

Reconstruction collaborative: efficacité de la MWM

- Nouveau jeu de données **artificiel**: Cube
- 100 individus décrits par 3 descripteurs
- chaque vue est créée en **supprimant** une des dimensions
- **Objectif**: Tester la capacité de la méthode de pondération par masque à **détecter quels descripteurs sont pertinents**.



(a) Jeu de données Cube



(b) Projection

Reconstruction collaborative: efficacité de la MWM

- La méthode de pondération par masque **améliore les résultats en reconstruction**.
- La sélection des features reconstruits **fonctionne t-elle en pratique ?**

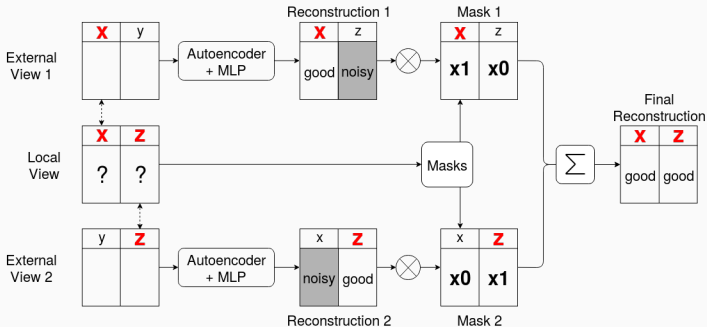


Figure 23: Combinaison de deux individus partiellement corrects

	Moyenne	Écart type
Descripteur partagé	0.920	0.026
Descripteur non partagé	0.143	0.034

Table 4: Moyenne et écart-type des paramètres des masques en fonction du descripteur qu'elles pondèrent

- Les masques arrivent à **cibler les descripteurs partagés** tout en **limitant l'utilisation des descripteurs non partagés**.
- Le faible écart type indique une certaine **stabilité** de la méthode

Contributions

- Définition d'un nouveau **cas d'application du paradigme collaboratif**.
- Définition d'un système permettant de **reconstruire de manière collaborative** et basé sur des **réseaux de neurones**.
- Définition d'une méthode permettant de **combinaison efficacement un ensemble de reconstructions**.

Résultats

- Scores **améliorables** en reconstruction.
- Reconstruction **conserve une information intrinsèque**.
- Proof of concept, mais le système global doit être **amélioré**.

Contributions et Perspectives

Dans cette thèse, nous avons proposé les trois contributions suivantes:

- Un algorithme de clustering collaboratif incrémental permettant une adaptation des clusters aux éventuelles évolutions de la distribution des données.
- Une méthode de pondération automatique des vues externes permettant d'améliorer la qualité des collaborations lors d'un clustering collaboratif.
- L'application du paradigme collaboratif au problème de la reconstruction de données manquantes.

Les apports de la thèse concernent les deux domaines suivants:

- Le clustering collaboratif
 - (article) ICONIP 2017
 - (article) EGC 2018
 - (article) IJCNN 2018
- La reconstruction collaborative
 - (journal, en cours de soumission) KAIS

- Analyse du comportement aux limites de la méthode de clustering collaboratif incrémental.
- Amélioration de chaque module utilisé pour le système de reconstruction collaborative, en particulier les liens permettant d'obtenir les reconstructions intermédiaires.
- Étude théorique et approfondissement du paradigme de reconstruction collaborative: définition des conditions de transférabilité des informations inter-vues, attestation de la qualité des individus reconstruits.

Merci pour votre attention.

Questions ?