

Weighted Hierarchical Mixed Topological Map: une méthode de classification hiérarchique à deux niveaux basée sur les cartes topologiques mixtes pondérées W-HMTM

Mory Ouattara

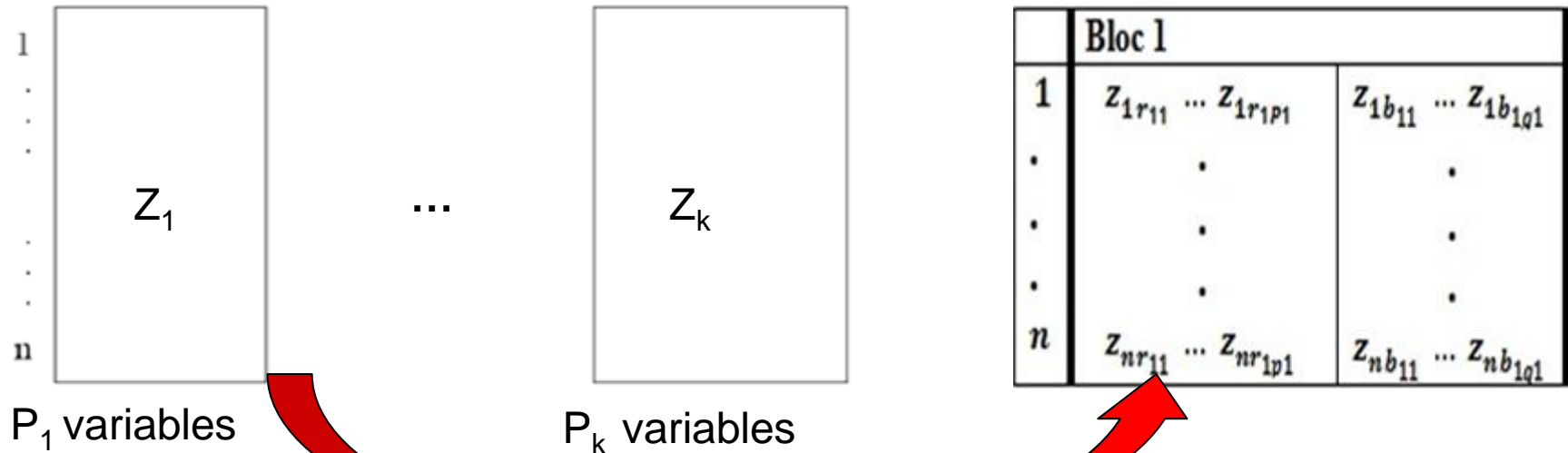
Ndèye Niang-Keita, Sylvie Thiria, Corinne Mandin, Fouad Badran

Plan

1. Introduction
2. Mixed Topological Maps (MTM)
3. Weighted Hierarchical Mixed Topological maps (WHMTM)
4. Application
5. Conclusion & Perspective

Introduction

Grand nombre de variables mixtes structurées en bloc



K tableaux à étudier

Trouver une $K + 1$ classification automatique spécifique à chaque bloc et dans le même temps une classification générale des individus qui restitue au mieux l'information de chaque bloc

Approche Tandem

Classification automatique sur composantes synthétiques obtenues par réduction du nombre de variables :

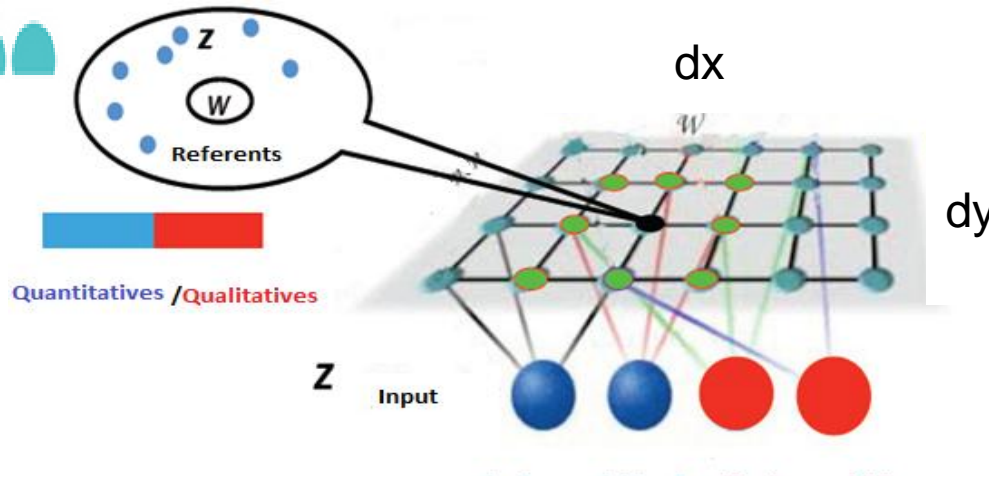
- ACP si variables quantitatives
- ACM si variables qualitatives
- AFDM : réduction par analyse factorielle des données mixtes

Approches Directes neuronales

- SOM (Kohonen, 1995)
- KACM (Cottrell, 2003)
- MTM (Lebbah, 2005)

La méthode WHMTM (Weighted-Hierarchical Mixed Topological Map) permet le traitement simultané de variables qualitatives et quantitatives ainsi que **la prise en compte de leur structuration en blocs**

Carte Topologique Mixte (Lebbah (2005))



$$P_c = \{Z \in E / \chi(Z) = c\}$$

Affectation

Carte topologique

$$D(z_i, w_c) = \|z_i - w_c\|^a = \|z_i^r - w_c^r\|^a + \|z_i^b - w_c^b\|^2 = \|z_i^r - w_c^r\|^2 + 63\mathcal{H}(z_i^r, w_c^b)$$

Foction de Voisinage

$$J_{MTM}^T(\chi, w) = \sum_{z_i \in E} \sum_{c \in C} \kappa^T(\delta(\chi(z_i), c)) * D$$

➤ Initialisation : Choix des référents initiaux W_0 et des paramètres d'apprentissages

➤ Etape itérative : les référents W^{t-1} sont connus et fixés

□ Phase d'affectation : les z_i sont affectés aux référents le plus proche

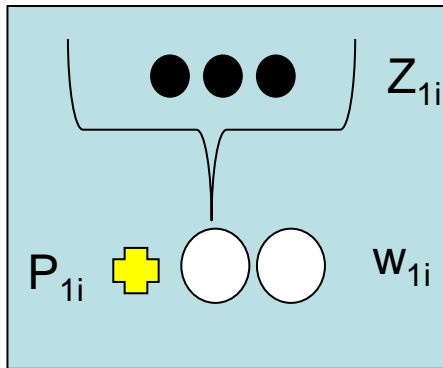
$$\chi(z_i) = \mathit{argmin}_c (\|z_i - w_c\|^2)$$

□ Phase de minimisation : les Z_i sont fixés pour le calcul des centroïdes

$$W_c^r = \frac{\sum_{z_i \in E} \kappa(\delta(\chi(z_i), r)) z_i^r}{\sum_{z_i \in E} \kappa(\delta(\chi(z_i), r))} \quad \text{pour } W^r$$

$$W_c^{bk} = \begin{cases} 0 & \text{si } \sum_{z_i \in A} \kappa(\delta(\chi(z_i), r)) (1 - z_i^{bk}) > \sum_{z_i \in A} \kappa(\delta(\chi(z_i), r)) z_i^{bk} \\ 1 & \text{sinon} \end{cases} \quad \text{pour } W^b$$

Méthode proposée: w-HMTM



Bloc 1	
1	$Z_{1r_{11}} \dots Z_{1r_{1p_1}} \quad Z_{1b_{11}} \dots Z_{1b_{1q_1}}$
.	.
.	.
.	.
n	$Z_{nr_{11}} \dots Z_{nr_{1p_1}} \quad Z_{nb_{11}} \dots Z_{nb_{1q_1}}$

...

Bloc P	
$Z_{1r_{p1}} \dots Z_{1r_{pp}}$	$Z_{1b_p} \dots Z_{1b_{qp}}$
.	.
.	.
.	.
$Z_{nr_{p1}} \dots Z_{nr_{pp}}$	$Z_{nb_1} \dots Z_{nb_{qp}}$

Niveau inférieur application de MTM à chaque blocs de variables

1	$V_{1x_1} = \begin{pmatrix} w_{11} & \dots & w_{1bx_1} \end{pmatrix}$...	$V_{1x_k} = \begin{pmatrix} w_{1b} & \dots & w_{1bx_k} \end{pmatrix}$
.	$V_{1x_1} = \begin{pmatrix} w_{11} & \dots & w_{1bx_1} \\ \vdots & & \vdots \end{pmatrix}$...	$V_{1x_k} = \begin{pmatrix} w_{1b} & \dots & w_{1bx_k} \\ \vdots & & \vdots \end{pmatrix}$
.
.
n	$V_{-n} = \begin{pmatrix} w_{n1} & \dots & w_{Nbx_1} \end{pmatrix}$...	$V_{-n_{xk}} = \begin{pmatrix} w_{N1} & \dots & w_{Nbx_k} \end{pmatrix}$

Méthode proposée: w-HMTM

Distance pondérée

$$D^2(z_i, w_j) = \sum_{k=1}^n \underline{p_{ik}} (z_{ik} - w_{jk})^2$$

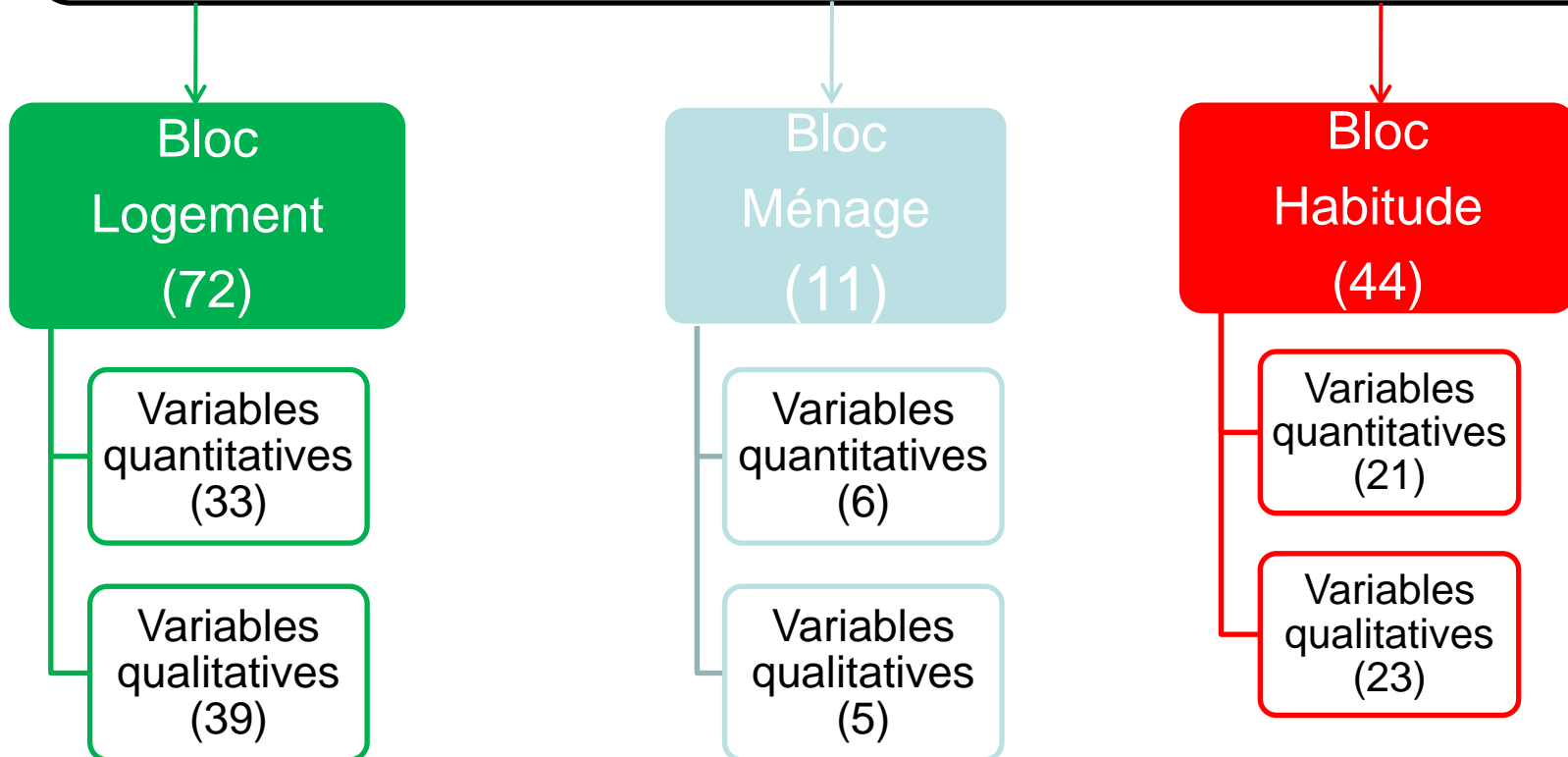
Fonction de coût $J_{WHMTM}^T(\chi, \Omega) = \sum_{V_i \in E} \sum_{c \in C} k^T(\sigma(\chi(V_i), \Omega_c)) D^2(V_i, \Omega_c)$

$$W_c^r = \frac{\sum_{V_i \in E} \sum_k p_{ik} k(\sigma(\chi(V_i), c)) V_i^r}{\sum_{V_i \in E} \sum_k p_{ik} k(\sigma(\chi(V_i), c))}$$

$$W_c^{bk} = \begin{cases} 0 & \text{si } \sum_{V_i \in E} k(\sigma(\chi(V_i), c)) \sum_k p_{ik} (1 - V_i^{bk}) > \sum_{V_i \in E} k(\sigma(\chi(V_i), c)) \sum_k p_{ik} (V_i^{bk}) \\ 1 & \text{sinon} \end{cases}$$

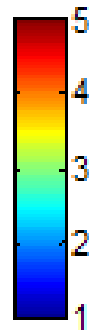
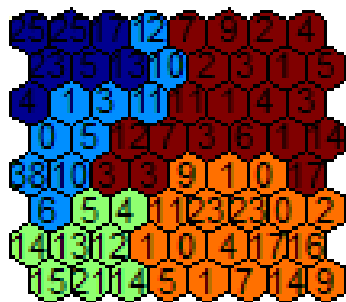
Application de w-HMTM aux données de la Compagne nationale « Logements »

Base de données
(125 variables, 567 logements)



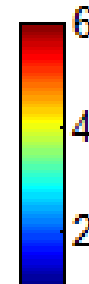
Structure des cartes au niveau des blocs de variables

Logements



DB=0,95

Ménages



DB=0,92

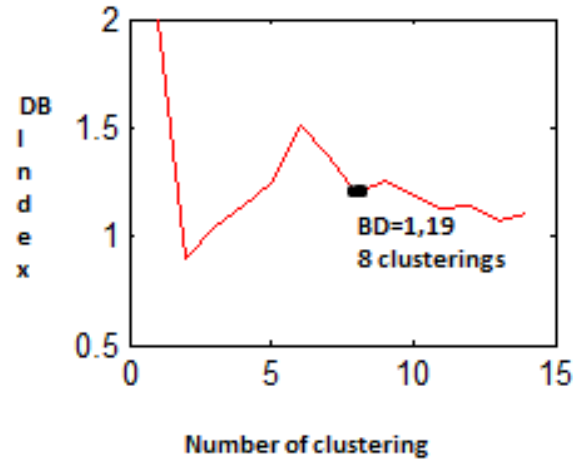
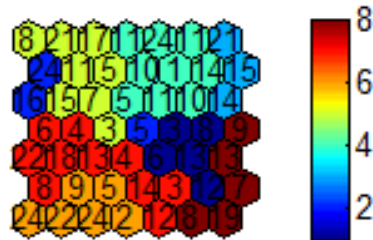
Habitudes



DB=1,15

Structure des cartes au niveau 2

W-HMTM



Contenue de la carte au niveau 2 du WHMTM

CI 1

Grandes maisons individuelles

Personnes seules

Revenus faibles

Entretiennent plus leur logement

Forte concentration en Acroléine

CI 2

Grandes maisons individuelles

Vivent en couple

Concentrations fortes en toluène

CI 3

Logements collectives et récents

Personnes jeunes

Entretiennent moyennement

+ d'enfants de plus de 10 ans

Faible concentration en benzène **et** Dichlorobenzène

CI 4

Logements collectives et anciens

Jeunes

Entretiennent faiblement leurs logements

Méthodes neuronales : W-HMTM

Rand	Log	Men	Hab	W-HMTM	MTM
Logement	1	0,69	0,70	0,82	0.73
Ménage		1	0,72	0,68	0.67
Habitudes			1	0,72	0.67
W-MTM				1	0.55
MTM					1

Indice de comparaison de Rand

Conclusion

- La méthode de classification à deux niveaux présentée ici permet une double description synthétique des données par rapport aux blocs et par rapport à l'information globale initiale contenue dans les données
- Classification finale des individus représentant un consensus entre les blocs de variables

- **Rendre les poids adaptatifs dans W-HMTM**

- **Sélectionner au niveau inférieur les variables les plus informatives**

**Merci beaucoup pour votre
attention**