Visualisation avec les cartes topologiques catégorielles

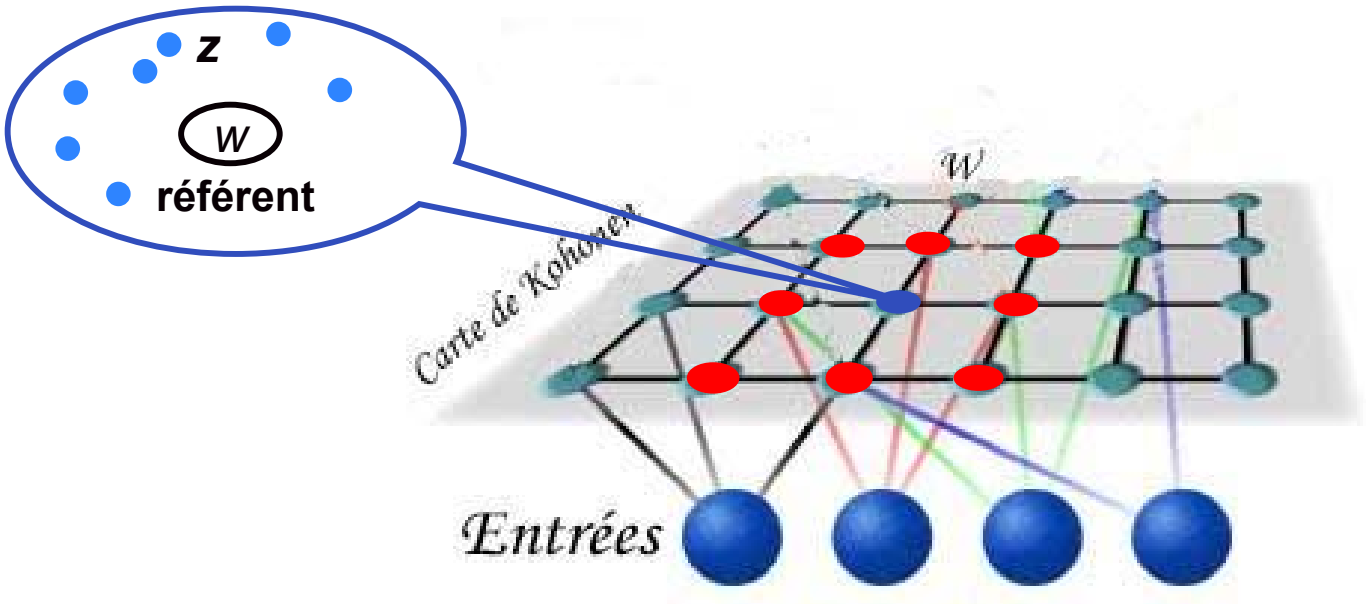
Mustapha LEBBBAH, Fouad BADRAN, Sylvie THIRIA
CNAM, paris 6,Renault

CLERMONT FERRAND 20/01/2004

PLAN

- Cartes topologiques et Données qualitatives
- Carte topologique binaire: **BTM**
- Carte topologique probabiliste: **CTM**
- Validation du CTM
- Conclusion

Cartes topologiques et données qualitatives



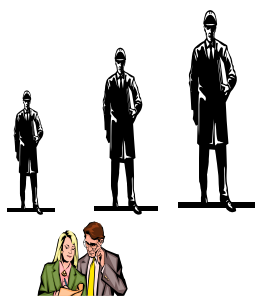
$K(\delta(\ c_1,c_2)\ )$

Fonction noyau

•Données qualitatives

Taille:

Sexe:



Variables qualitatives et codage

Taille: **Petit**, Moyen, Grand

Petit, **Moyen**, Grand

Petit, Moyen, **Grand**

Ordinale

1 0 0

1 1 0

1 1 1

Couleur : **rouge**, vert, bleu

rouge, **vert**, bleu

rouge, vert, **bleu**

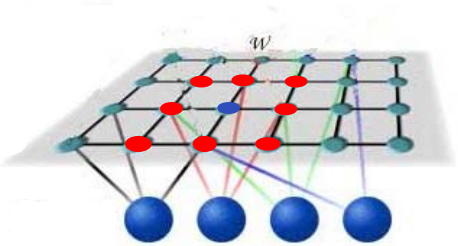
Disjonctif

1 0 0

0 1 0

0 0 1

Le modèle BTM



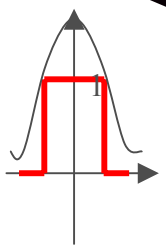
Goveart 88, Marchetti 89

Distance de Hamming

$H(.,.)$

100**1111**0**1**0**1**0

Additif , Disjonctif



BTM

1	1	1	1	1	0	0	1	1	1	0	0
1	1	0	1	1	1	1	1	0	0	0	0
1	1	1	1	0	0	0	1	1	1	1	1
1	1	1	1	1	1	0	1	1	0	0	0
1	1	1	1	1	1	0	1	1	1	1	0
1	0	0	1	1	1	0	1	1	1	0	0

K_1
K_2
K_2
K_3
K_4
...

Fonction de coût

$$E(W) = \sum_{z_i \in APP} \sum_{r \in C} K(\delta(c,r)) H(z_i, W_r)$$

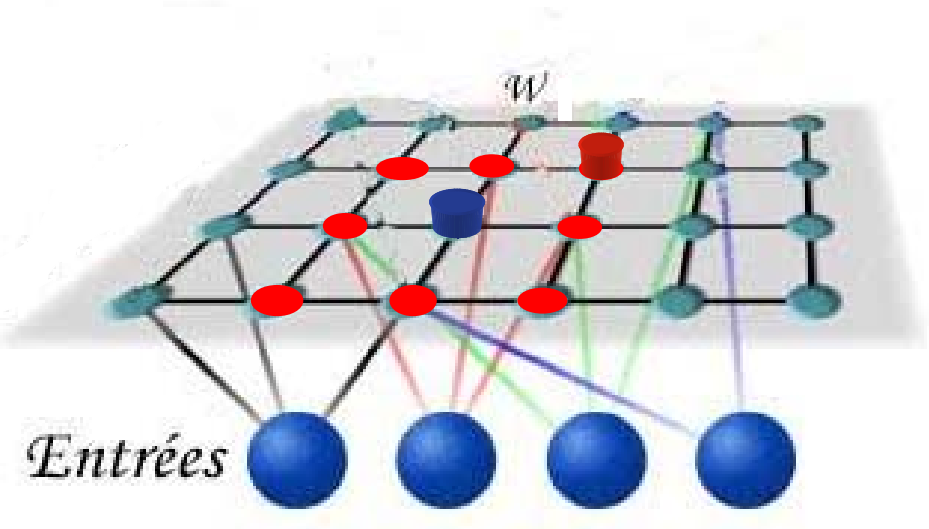
1 1 1 1 0 1 0 1 0 1 0 0

Centre médian

Minimisation utilisant les nuées dynamiques

Cartes Topologiques et Probabilité

$p(z/c_1)$ $p(c_2)$



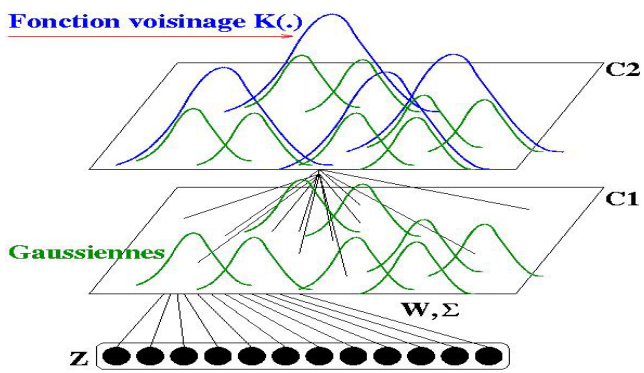
$$p(c_1/c_2) = \frac{K(\delta(c_1,c_2))}{\Sigma_c}$$

$$p(z) = \Sigma_{C2}\, p(c_2)\, p_{c_2}(z)$$

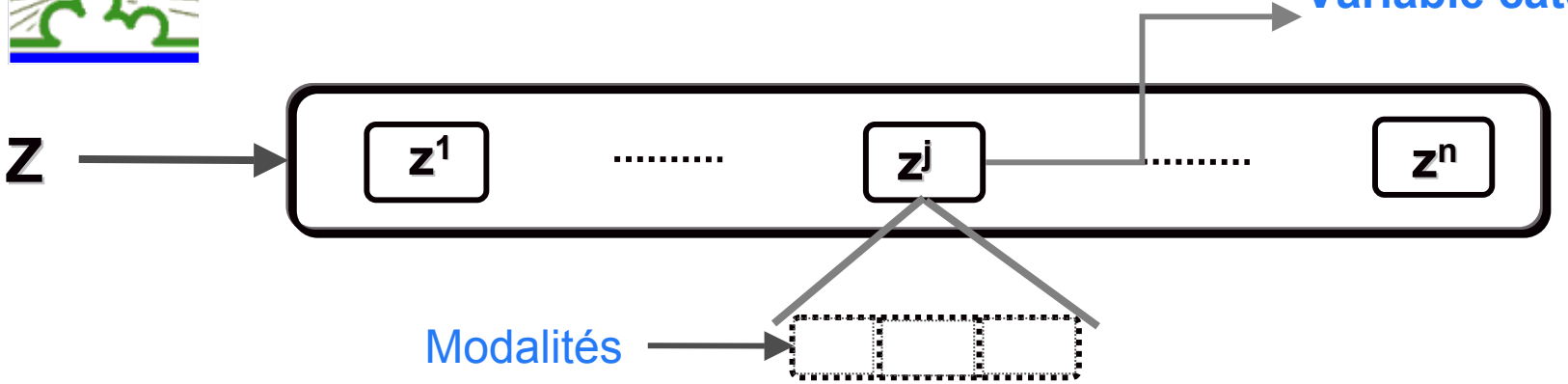
$$p_{c_2}(z) = \Sigma_{C1}\, p(c_1/c_2)\, p(z/c_1)$$

$p(z/c_1)$= **Loi Gaussienne sphérique**

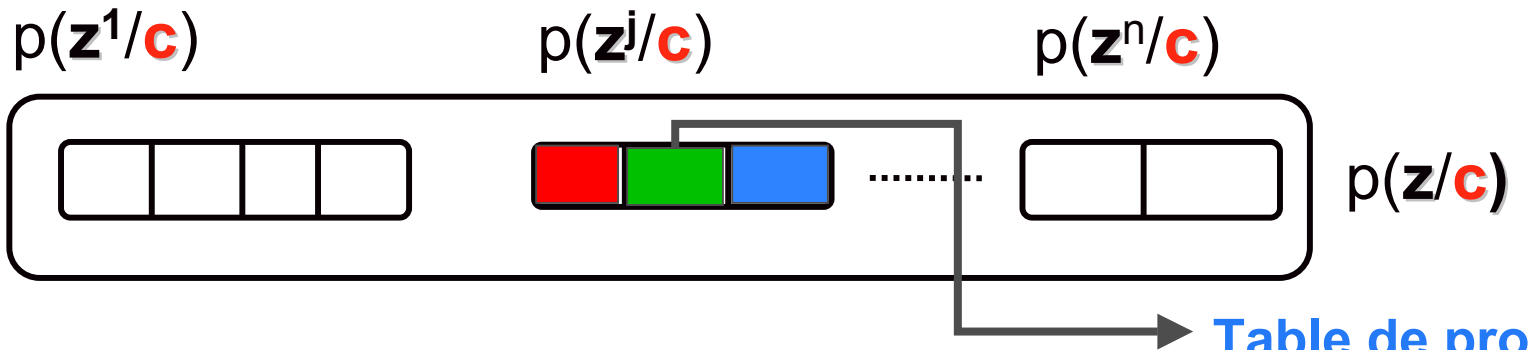
$p(z)$= **Mélange de mélanges locales de lois Gaussiennes**



Le Modèle CTM



Variables indépendantes → $p(z/c)=\prod_{j=1..n} p(z^j/c)$



$$p(z)= \Sigma_{c2} \boxed{p(c_2)} \Sigma_{c1} p(c_1/c_2) \boxed{p(z/c_1)}$$

EM

CTM= Carte Topologique+ données catégorielles

Formules obtenues:

$$p(c_2) = \frac{\sum_z p(c_2/z)}{\sum_{c2} \sum_z \sum_{c1} p(c_1,c_2/z)}$$

Probabilité a priori

$$p(z^{j,mod}/c) = \frac{\sum_{z \in zj,mod} p(c_1/z)}{\sum_z p(c_1/z)}$$

Une composante de la table de probabilités unidimensionnelle

$$p(c_1/z) = \sum_{c2} p(c_1,c_2/z)$$

$$p(c_2/z) = \sum_{c1} p(c_1,c_2/z)$$

$$p(c_1,c_2/z) = \frac{p(c_2)\ K(\delta\ (c_1,c_2))\ p(z/c_1)}{\sum_{c2} p(c_2)\ p(z/c_2)}$$

Fonction de voisinage

Prise en compte de l'ordre topologique

Validation du CTM

Base réelle belge

1106 assurés, 9 variables, Bon / Mauvais conducteur

- **U**tilité (Privé,Professionnelle)
- **S**exe (Homme, Femme, Véhicule de Société)
- **L**angue (Français, Autre)
- **Ag**e (Vieux, Moyen, Jeune)
- **Lo**calisation (Capitale, Province)
- **B**onus (1,2)
- **Po**lice (86, Autre)
- Puissance (Grande, Petite)
- **A**ge **V**éhicule (Ancien, Nouveau)

p(z/c)

U	S	Lg	Ag	Lo	B	Po	Pu	AV
Pv	?	?	?	?	?	?	?	?
Pf	?	?	?	?	?	?	?	?
	?		?					

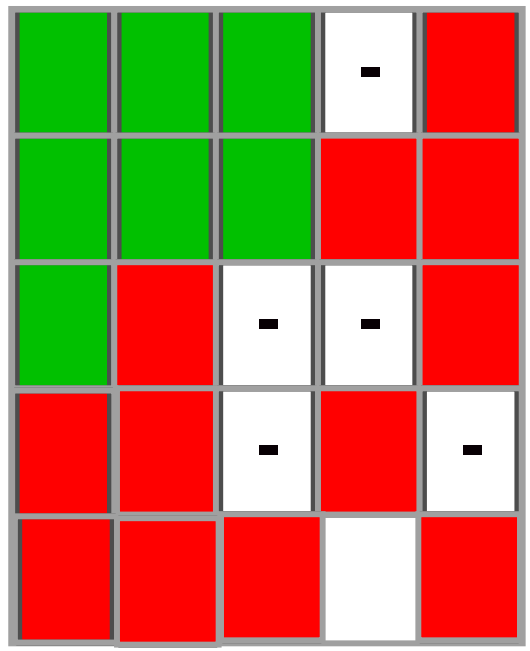
20 paramètres à estimer

Visualisation multi-dimensionnelle

H	V	-	J	H	J	F	J	-	J
-	An	-	An	-	An	Pt	Nou	Gr	-
H	-	-	J	H	J	F	J	VS	J
-	-	-	An	Gr	An	Pt	Nou	Gr	Nou
H	V	H	-	VS	J	VS	J	VS	J
Gr	-	Gr	-	Pt	Nou	Pt	Nou	Gr	Nou
-	-	H	-	VS	J	VS	J	VS	J
Gr	-	Gr	-	Pt	An	Pt	An	Gr	Nou
H	M	-	M	F	M	F	M	F	M
Gr	-	-	-	Pt	-	Pt	Nou	Pt	Nou

Représentation des 4 variables

(**Sexe**, **Age**, **Puissance**, **Age du Véhicule**,)



Affectation avec la probabilités a posteriori p(c/z)

Etiquetage

bon/mauvais

Base de Sémiométrie

Base réelle

1128 individus, 70 Mots (variables)

7 notes sémiométriques (modalités)

- Univocité sémantique
- Stabilité sémantique
- Non-consensualité

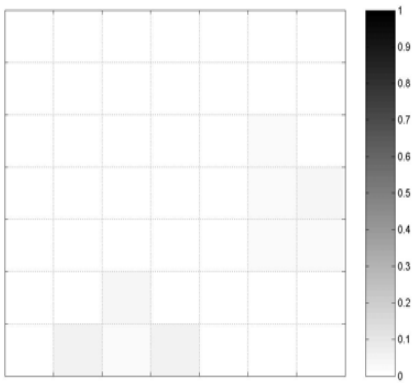
ACHETER
AMBITION
ANGOISSE
ARGENT
ASTUCIEUX
AUDACE
BIJOU
CADEAU
….

Mot → 

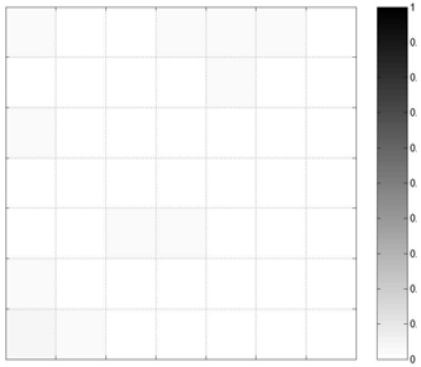
7 paramètres à estimer

Répartition de probabilité du mot : ACHETER

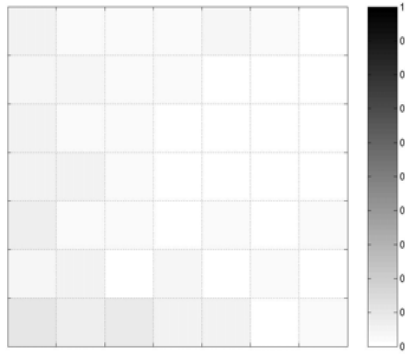
note:1..7



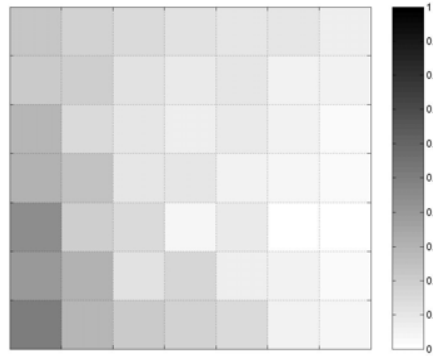
1



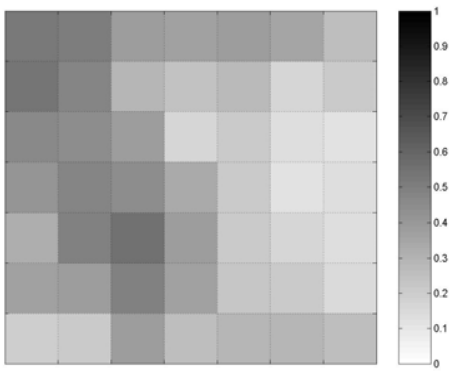
2



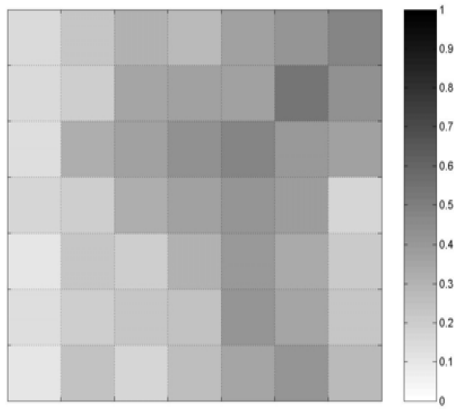
3



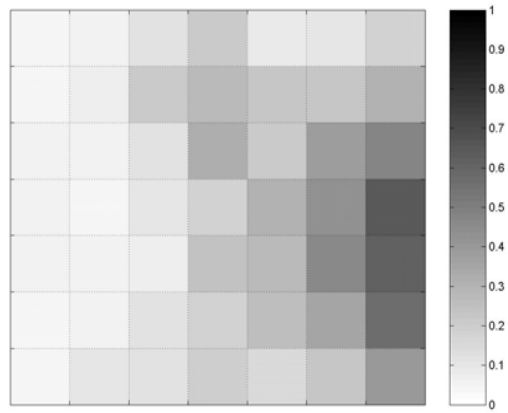
4



5



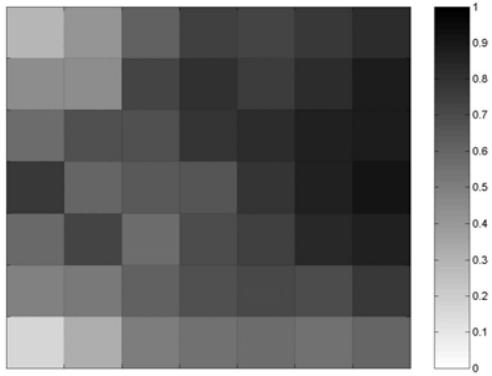
6



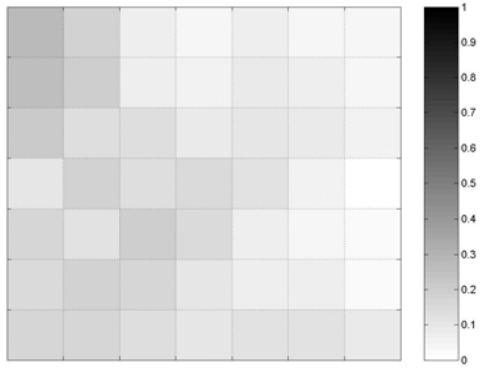
7

Répartition de probabilité du mot : Mort

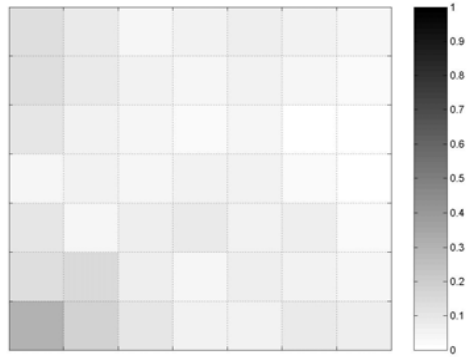
note: 1..7



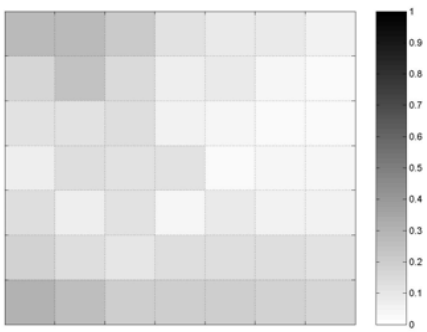
1



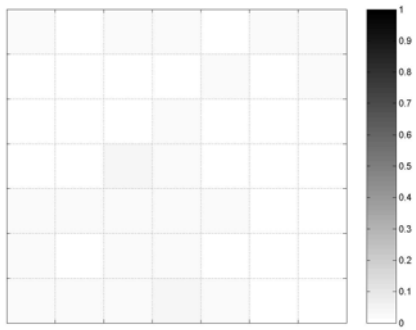
2



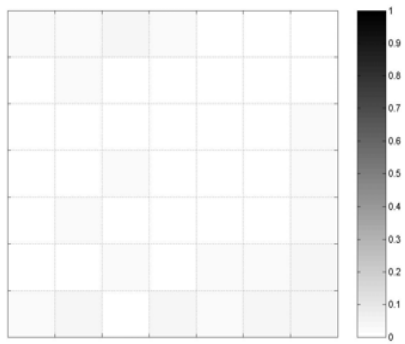
3



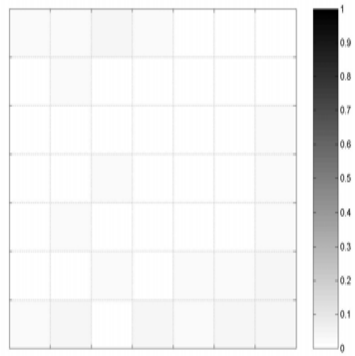
4



5

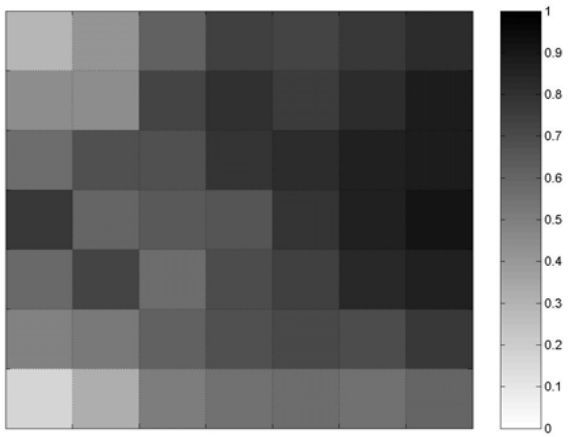


6

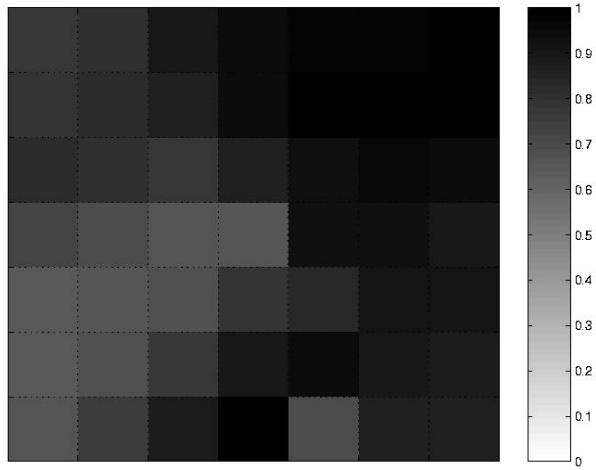


7

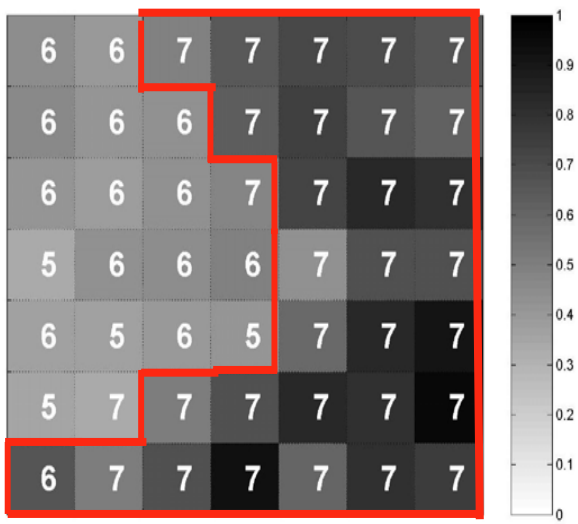
Répartition de la probabilité Maximale



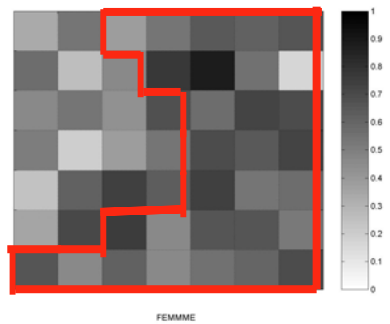
Mort (1)



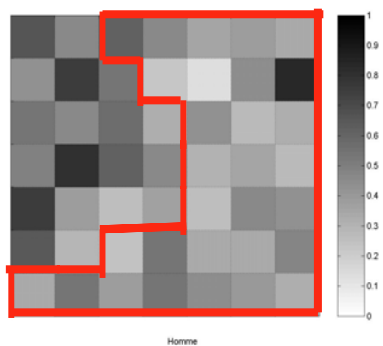
Guerre (1)



Fleur



Femme



Homme

Conclusion

CTM

- Etendre le champs d'application des cartes topologiques

- Réaliser une partition de données en des sous-ensembles homogènes

- Outils de visualisation

- Utiliser comme classifieur

Perspectives

•Utiliser une loi adaptée aux données qualitatives codées en additives

•Adapter la CAH à ce type de données

•Evolution des états sur la carte (HMM)