



**HAL**  
open science

## Datamining spatial Une approche basée sur la notion de zone

Mohammed Ridouani, F Elhebil, M Dahchour, M Hassi Rahou

► **To cite this version:**

Mohammed Ridouani, F Elhebil, M Dahchour, M Hassi Rahou. Datamining spatial Une approche basée sur la notion de zone. 4 EMES JOURNEES NATIONALES SUR LES SYSTEMES INTELLIGENTS: THEORIES ET APPLICATIONS , Nov 2006, Mohammedia, Maroc. hal-01426327

**HAL Id: hal-01426327**

**<https://hal.science/hal-01426327>**

Submitted on 4 Jan 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Datamining spatial

## Une approche basée sur la notion de zone

M. RIDOUANI<sup>1</sup> F. ELHEBIL<sup>2</sup> M. DAHCHOUR<sup>3</sup> M. HASSI RAHOU<sup>4</sup>

<sup>1</sup> ESTC, Département Génie Informatique - Casablanca - Maroc.  
mohammed\_ridouani@yahoo.fr

<sup>2,3</sup> Institut National des Postes et Télécommunications (INPT), Rabat, Maroc.  
E-Mail : elhebil@inpt.ac.ma, dahchour@inpt.ac.ma

<sup>4</sup> Agence Nationale de Réglementation des Télécommunications (ANRT), Rabat, Maroc.  
E-Mail : rahou@anrt.ac.ma

### Résumé :

La fouille de données spatiales est une extension de la fouille de données classiques prenant en compte les relations spatiales entre objets. Les règles d'associations correspondantes se basent essentiellement sur la notion de *distance*. Dans cet article, nous proposons une approche d'extraction des règles d'association qui se base plutôt sur la notion de *zone*. Nous avons défini les algorithmes requis et avons implémenté une application informatique mettant en évidence l'intérêt de notre approche.

### Mots-clés :

Datamining spatial, Datamining zonal, règles d'association zonales.

### Abstract:

The spatial datamining is an extension of the traditional datamining taking into account the spatial relationships between objects. The related association rules are based on the notion of distance. In this article, we propose an approach of extracting association rules which is based rather on the notion of zone. We defined the necessary algorithms and implemented an application showing the interest of our approach.

### Keywords:

Spatial datamining, zonal datamining, zonal association rules.

## 1 Introduction

La fouille de données ou datamining vise à exploiter les grandes quantités de données collectées dans divers domaines. Elle désigne le processus d'extraction de connaissances implicites, précédemment inconnues et potentiellement utiles à partir de données. Ces connaissances apprises peuvent être sous

diverses formes : règles, modèles, concepts, etc.

La fouille de données *spatiales* (FDS) s'intéresse à l'analyse des bases de données spatiales [7, 9, 15, 10]. Elle constitue une extension de la fouille de données classiques [4,6] en ce sens qu'elle prend en compte les relations spatiales entre objets. Il est important de souligner que l'extraction des règles d'association spatiales est indissociable de la notion de *distance*.

Ce travail, effectué en collaboration avec l'ANRT (Agence Nationale de Réglementation des Télécommunications), introduit la notion de datamining zonal. Cette approche consiste à segmenter la région à étudier en plusieurs zones bien délimitées (p.ex., ville, quartier, une grande surface, etc.). Les règles d'association correspondantes sont dites *zonales* et se basent sur l'appartenance des objets à ces zones. Elles permettent ainsi de trouver la dépendance entre les objets sans devoir calculer la distance entre eux.

Pour appréhender davantage notre approche nous avons pris l'exemple de la ville de Mohammedia et développé une application qui consiste à extraire et à interpréter les règles d'association caractérisant les cybercafés de la ville. Il s'agit de déterminer le type et le nombre d'objets spatiaux qui déterminent le nombre des cybers. Ces objets spatiaux peuvent être des établissements, des

centres de culture et loisirs, des complexes médicaux, des réseaux routiers, des zones industrielles, ou autres.

La suite de cet article est structurée comme suit. La section 2 rappelle quelques définitions de base. Dans la section 3 nous présentons les motivations de la démarche de datamining zonal. La section 4 est consacrée à la présentation des différents algorithmes qui ont été élaborés. La section 5 présente l'application qui a été développée pour simuler la répartition spatiale des cybercafés de la ville de Mohammedia.

## 2 Définitions

Dans cette section nous rappelons quelques définitions de base utiles pour appréhender la suite de ce travail.

**Définition 1 (Item) :** Un item est tout objet, article ou attribut appartenant à un univers  $I = \{x_1, x_2, \dots, x_n\}$

**Définition 2 (Itemset) :** Un itemset est tout ensemble d'items  $X$  de  $I$ . Lorsque le cardinal de l'itemset est égal à  $k$  nous parlons de  $k$ -itemset.

**Définition 3 (Transaction) :** Une transaction est un itemset identifiée de manière univoque par un identificateur. L'ensemble de tous les identificateurs de transactions tids sera désigné par l'ensemble  $\tau$ .

**Définition 4 (TidSet) :** On appelle tidset tout sous-ensemble d'identificateurs de transactions (tids) de  $\tau$ .

**Définition 5 (Base transactionnelle) :** Une base transactionnelle  $D$  est un ensemble de couples formés d'un identificateur de transaction tid et de la transaction proprement dite :  $D = \{(y, X_y) / y \in \tau, X_y \subseteq \tau\}$

**Définition 7 (Représentation d'une base transactionnelle) :** Une base de données transactionnelle peut être représentée sous forme horizontale, verticale ou booléenne.

**Définition 8 (Fréquence) :** La fréquence d'un itemset  $X$ , noté  $freq(X)$ , est le nombre de

transactions de  $D$  contenant  $X$  :

$$freq(X) = |\{(y, X_y) \in D / X \subseteq X_y\}| = |t(X)|$$

**Définition 9 (Support) :** Le support d'un itemset  $X$ , noté  $sup(X)$  (noté  $\sigma(X)$ ) est la proportion de transactions de  $D$  contenant  $X$  :

$$sup(X) = \frac{|\{(y, X_y) \in D / X \subseteq X_y\}|}{|D|} = \frac{|t(X)|}{|D|}$$

**Définition 10 (Itemset fréquent) :** Etant donné un seuil, appelé support minimum, un itemset  $X$  est dit fréquent (relativement à  $\gamma$ ) dans une base de transactions  $D$ , si son support dépasse un seuil fixé a priori appelé support minimum et noté  $\gamma$ .

$X$  est fréquent ssi  $sup(X) \geq \gamma$

**Définition 11 (Règle d'association) :** Soit  $X$ , un itemset et  $A$  un sous-ensemble de  $X$ . Une règle d'association est une règle de la forme  $A \rightarrow X$ ,  $A$  exprimant le fait que les items de  $A$  tendent à apparaître avec ceux de  $X$ .  $A$  s'appelle la prémisse de la règle et  $X$  la conséquence.

**Définition 12 (Confiance) :** La confiance d'une règle d'association  $A \rightarrow C$  notée  $conf(A \rightarrow C)$ , représente la proportion de transactions couvrant  $A$  qui couvrent aussi  $C$ . On peut écrire :

$$conf(A \rightarrow C) = \frac{|t(A \cup C)|}{|t(A)|}, \text{ ou}$$

$$conf(A \rightarrow C) = \frac{supp(A \cup C)}{supp(A)} \text{ avec } 0 \leq conf(A \rightarrow C) \leq 1$$

**Définition 13 (Règle d'association solide) :** Une règle d'association  $A \rightarrow C$  est dite solide si étant donné un support minimum  $\gamma$ , l'itemset  $A \cup C$  est fréquent et si sa confiance dépasse un seuil donné, fixé a priori, appelé le seuil de confiance minimum  $\phi$ .

$A \rightarrow C$  est solide ssi  $supp(A \cup C) \geq \gamma$  et  $conf(A \rightarrow C) \geq \phi$

**Définition 14 (Étapes d'extraction des règles d'association) :** Etant donné un support minimum  $\gamma$  et une confiance minimum  $\phi$ , le processus d'extraction des règles d'association peut être effectué en deux étapes :

- Trouver les itemsets fréquents :

$$F = \{X \subseteq \tau / sup(X) \geq \gamma\}$$

- Trouver les règles solides :

$$R = \{r : (A \rightarrow C) / (A \cup C) \subseteq F \wedge (A \cap C) = \emptyset \wedge \text{conf}(A \rightarrow C) \geq \theta\}$$

techniques d'extraction des règles d'association spatiales.

**Définition 15 (Règle d'association spatiale) :**

Une règle d'association spatiale est de type :  $X \rightarrow Y(s,c)$  avec  $s$  comme support et  $c$  la confiance telles que  $X$  et  $Y$  sont des ensembles de prédicats spatiaux (p.ex., *intersecte*, *disjoint*, *à gauche de*, etc.) et non spatiaux (p.ex., *catégorie-âge* ( $x, jeune$ )).

Exemple :

$$\text{est-une}(x, \text{maison}) \wedge \text{près-de}(x, \text{plage}) \rightarrow \text{est-chère}(x) (90\%)$$

### 3 Position du problème

Il ressort d'une étude réalisée pour le compte de l'ANRT [2] que les cybercafés jouent un rôle important dans le développement du marché de l'internet. Dés lors, le problème qui se pose est d'identifier le type d'objets spatiaux et leur nombre qui déterminent le nombre des cybercafés. Pour cela, on s'appuie sur une base de données spatiales et on fait recours à la FDS afin d'extraire des règles d'associations pertinentes. Ces règles se basent généralement sur la notion de *distance*. Or la plupart des informations spatiales sont stockées dans les bases de données sous forme d'adresses postales. Ceci pose le problème de géocodage : l'opération qui consiste généralement à associer une adresse postale à un couple de coordonnées ( $X, Y$ ).

Par ailleurs, la plupart des techniques d'extraction des règles d'association spatiales se basent sur l'algorithme Apriori [1] qui a montré ces limites comme précisé dans [3,5]. D'autres algorithmes avancés ont été proposés notamment dans l'extraction des représentations condensées des itemsets fréquents : la génération des itemsets maximaux (ex : MaxMiner [3] et Pincer-Search [8]) et des itemsets fermés (ex: Close [11], A-Close, Close+ [12], Charm [14], Closet [13]).

Ces algorithmes alternatifs visent les bases de transactions et ne sont pas supportés par les

## 4 Datamining zonal

L'objectif du datamining zonal que nous proposons est de contourner le problème de géocodage en définissant des zones ainsi que l'appartenance d'objets spatiaux à ces zones, et par la suite, établir la dépendance entre ces objets spatiaux, sans devoir calculer les coordonnées ( $X, Y$ ) de chaque objet spatial et la distance entre les différents objets spatiaux. Pour supporter pleinement les algorithmes alternatifs à celui d'Apriori (voir section 3), nous proposons de transformer la base des données spatiales en une base de transactions.

### 4.1 Démarche d'extraction des règles d'association zonales

Notre démarche est composée de quatre algorithmes :

- Algorithme 0 : Etape préliminaire (définition des zones et localisation des objets spatiaux)
- Algorithme 1 : Préparation des ingrédients qui seront utilisés pour préétablir la BDT
- Algorithme 2 : Préparation de la BDT
- Algorithme 3 : Extraction des règles d'association zonales solides

#### Notations

Symbole	Signification
$Z_i$	la $i^{\text{ème}}$ zone
$N_z$	Le nombre total de zones
$OBJ_i$	le $i^{\text{ème}}$ Objet spatial
$N_o$	Le nombre total d'objets spatiaux
$I_{nm}$	le $n^{\text{ème}}$ Intervalle du la $m^{\text{ème}}$ objet
$K_i$	Le nombre d'intervalles pour
$/* \dots */$	Commentaire
$Z-OBJ_i$	la $i^{\text{ème}}$ table Zone-Objet
$NOZ_{ii}$	Le nombre d'objets spatiaux
$N_{cz}$	Le nombre de caractéristiques de
$C-Z_i$	la $i^{\text{ème}}$ caractéristique des zones
BDT	Base de transaction

### Algorithme 0 : Etape préliminaire

#### Entrée 0 :

Base de donnée brute

#### Sortie 0:

La table zone

Les tables d'objets spatiaux

#### Début

01 : définir les zones

02 : créer la table zone

03 : localiser chaque objet spatial (chercher la zone où appartient chaque objet spatial)

04 : créer les tables d'objets spatiaux

#### Fin

### Algorithme 1 : Préparation des ingrédients qui seront utilisés pour préétablir la BDT

#### Entrée 1 :

La table zone

Les tables d'objets spatiaux

#### Sortie 1:

Les tables Zone-Objet

Les statistiques sur les tables Zone-Objet

#### Début

00 : /\*.....Création et chargement des tables Zone-Objet .....\*/

01 : Pour i allant de 1 à  $N_O$  Faire

02 : Création de la table Zone-Objet : Z- $OBJ_i$

03 : /\*.....Chargement de la table Zone-Objet : Z- $OBJ_i$ .....\*/

04 : Pour j allant de 1 à  $N_Z$  Faire

05 : calculer le nombre d'objets spatiaux ( $OBJ_j$ ) dans chaque zone  $Z_j$  :  $NOZ_{ij}$

06 : charger ce nombre ( $NOZ_{ij}$ ) dans la table Z- $OBJ_i$

07 : Fin pour

08 : Fin pour

09 : /\*.....Calcule des statistiques sur les tables Zone-Objet.....\*/

10 : Pour i allant de 1 à  $N_O$  Faire

11 : calculer des statistiques sur la table Zone-Objet : Z- $OBJ_i$

12 : Fin pour

#### Fin

### Algorithme 2 : Préparation de la BDT

#### Entrée 2

Les intervalles de variation du nombre d'objets spatiaux :  $I_{nm}$

Les tables Zone-Objet : Z- $OBJ_i$

Les  $N_{CZ}$  caractéristiques des zones supplémentaires : C- $Z_i$

#### Sortie 2:

Base de transaction : BDT

#### Début

01 : Créer le fichier de la base de transaction

02 : /\*.....Chargement du fichier de la BDT par des transactions.....\*/

03 : Pour j allant de 1 à  $N_Z$  Faire

04 : /\*Chargement de la  $j^{ème}$  transaction par des informations sur les objets spatiaux\*/

05 : Pour i allant de 1 à  $N_O$  Faire

06 : Pour r allant de 1 à  $k_i$  Faire

07 : Si  $NOZ_{ij}$  appartient à  $I_{ri}$  alors écrire 1 dans la base de transaction

08 : Si non alors écrire 0 dans la base de transaction

09 : Finsi

10 : Fin pour

11 : Fin pour

12 : /\*Chargement de la  $j^{ème}$  transactions par des caractéristiques sur la zone  $Z_j$ \*/

13 : Pour x allant de 1 à  $N_{CZ}$  Faire

14 : Si la zone  $Z_j$  a une caractéristique C- $Z_x$  alors écrire 1

15 : Si non alors écrire 0

16 : Finsi

17 : Fin pour

18 : Retour à la ligne

/\* pour écrire la prochaine transaction\*/

19 : Fin pour

#### Fin

### Algorithme 3 : Extraction des règles d'association zonales solides

#### Entrée 3

Base de transaction

Support minimum  $\gamma$ , confiance minimum  $\Phi$ .

#### Sortie3

Règles d'association zonales solides

#### Début

01 : Algorithme d'extraction des règles d'association : génération des itemsets fréquents

02 : Algorithme d'extraction des règles d'association : génération des règles d'association

#### FIN

## 4.2 Les tables et concepts associés

### Structures des tables Zone, Objet spatial et Zone-Objet :

**Zone**(Id\_Zone, Nom-Zone, C-Z1 : caractéristique-zone 1, ..., C-ZNcz : caractéristique-zone Ncz)

**Objet spatial OBJi**(Id OBJi, Nom-OBJi, Caractéristique-OBJi 1, ..., Caractéristique-OBJi 2, Id-Zone)

**Zone-Objet : Z-OBJi**(Id\_Zone, Nom-Zone, Nombre OBJi : NOZij)

### Exemples de statistiques sur les tables Zone-Objet :

- Savoir le nombre maximal ( $\max_i$ ) et le nombre minimal ( $\min_i$ ) d'apparitions de chaque objet spatial OBJ<sub>i</sub> par zone Z<sub>j</sub> ( $\max_i = \sup(\text{NOZ}_{ij} / j=1,2,\dots, N_Z)$  et  $\min_i = \min(\text{NOZ}_{ij} / j=1,2,\dots, N_Z)$ ).
- Calculer la moyenne d'apparitions de chaque objet spatial OBJ<sub>i</sub> par zone Z<sub>j</sub>

$$\text{moy}_i = \sum_{j=1}^{N_k} \text{NOZ}_{ij} / N_Z$$

### Les intervalles de variation du nombre d'objets spatiaux : $I_{nm}$

On prend par exemple : l'objet spatial 1 OBJ<sub>1</sub> ( $m=1$ ) avec  $\max_1=20$  et  $\min_1 = 0$  ( $0 \leq \text{NOZ}_{1j} \leq 20$ )

On peut prendre par exemple les intervalles des entiers  $I_{nm}$  suivants :  $I_{11} = [0,7[$

$$I_{21} = [7,16[ \quad I_{31} = [16,20[$$

### Chargement du fichier de la BDT par des transactions

On prend par exemple 3 zones ( $N_Z=3$ ) avec 2 caractéristiques ( $N_{CZ}=2$ ) et 2 objets spatiaux ( $N_O=2$ )

### Les tables zone

Table zone 1		Table zone 2		Table zone 3	
Id zone1	1	Id zone2	2	Id zone3	3
C-Z <sub>1</sub>	0	C-Z <sub>1</sub>	1	C-Z <sub>1</sub>	0
C-Z <sub>2</sub>	1	C-Z <sub>2</sub>	1	C-Z <sub>2</sub>	0

C-Z<sub>i</sub>=0 signifie que la zone n'a pas la caractéristique C-Z<sub>i</sub>

C-Z<sub>i</sub> =1 signifie que la zone a la caractéristique C-Z<sub>i</sub>

### Les tables Zone-Objet : Z-OBJ<sub>i</sub>

Table Zone-Objet : Z-OBJ1				Table Zone-Objet : Z-OBJ2			
Id_zone1	1	2	3	Id_zone2	1	2	3
Nom-zone	Z1	Z2	Z3	Nom-zone	Z1	Z2	Z3
NOZ1j	15	8	11	NOZ2j	6	8	2

$$\max_1=15, \min_1=8, \max_2=8, \min_2=2$$

Les intervalles des entiers  $I_{n1}$  ( $k_1=4$ ) :  $I_{11} = [8,10[$ ,  $I_{21} = [10,12[$ ,  $I_{31} = [12,14[$ ,  $I_{41} = [14,15[$

Les intervalles des entiers  $I_{n2}$  ( $k_2=3$ ) :  $I_{12} = [2,4[$ ,  $I_{22} = [4,6[$ ,  $I_{32} = [6,8[$ .

### Base de transaction

	OBJ <sub>1</sub>				OBJ <sub>2</sub>			C-Z <sub>i</sub>	
	I <sub>11</sub>	I <sub>21</sub>	I <sub>31</sub>	I <sub>41</sub>	I <sub>12</sub>	I <sub>22</sub>	I <sub>32</sub>	C-Z <sub>1</sub>	C-Z <sub>2</sub>
Z <sub>1</sub>	0	0	0	1	0	0	1	0	1
Z <sub>2</sub>	1	0	0	0	0	0	1	1	1
Z <sub>3</sub>	0	1	0	0	1	0	0	0	0

## 5 Application

L'objectif de l'application est d'extraire les règles d'association caractérisant les cybercafés de la ville de Mohammedia. Il s'agit de déterminer le type et le nombre d'objets spatiaux qui déterminent le nombre de cybers. Nos sources de données sont une base de données (cybers) provenant de l'ANRT et une carte vectorielle de la ville Mohammedia. Nous nous intéressons à trois objets spatiaux : Cybers (OBJ1), Établissements (écoles, collèges, lycées, facultés) (OBJ2) et Complexes médicaux

(cliniques, hôpitaux, centres médicaux) (OBJ3). Le nombre total d'objets spatiaux est donc  $N_O=3$ . Le nombre total de zones est  $N_Z=36$ . La zone est définie comme une *adresse*

L'intervalle $I_{mn}$	Numéro	Item	Nombre d'objet spatiaux est :	Catégories d'objets spatiaux
$I_{11} = [0,3[$	1	X <sub>1</sub>	faible	Cybers
$I_{21} = [3,5[$	2	X <sub>2</sub>	moyen	
$I_{31} = [5,7[$	3	X <sub>3</sub>	fréquent	
$I_{12} = [0,2[$	4	X <sub>4</sub>	faible	Etablissements
$I_{22} = [2,3[$	5	X <sub>5</sub>	moyen	
$I_{32} = [3,4[$	6	X <sub>6</sub>	fréquent	
$I_{13} = [0,1[$	7	X <sub>7</sub>	faible	Complexes médicaux
$I_{23} = [1,2[$	8	X <sub>8</sub>	moyen	
$I_{33} = [2,2]$	9	X <sub>9</sub>	fréquent	

*postale* et le nombre de caractéristique des zones est  $N_{CZ}=0$

**Les structures des tables d'objets spatiaux sont :**

**Cyber**(id\_cyber, nom\_cyber, adresse\_cyber, types\_accès, nom\_zone)  
**Etablissement**(id\_etab, nom\_etab, adresse\_etab, catégorie\_etab, nom\_zone)  
**ComplexeMédical**(id\_méd, nom\_méd, adresse\_méd, catégorie\_méd, nom\_zone)  
**Zone**(id\_zone, nom\_zone)

**Les structures des tables Zone-Objet (Z-OBJi) sont :**

**Zone-Cyber:Z-OBJ1**(id\_zone, nom\_zone :Zj, nbr-Cyber : NOZ1j)  
**Zone-Etablissement:Z-OBJ2**(id\_zone, nom\_zone :Zj, nbr-Etablissement : NOZ2j)  
**Zone-ComplexeMédical:Z-OBJ3**(id\_zone, nom\_zone :Zj, nbr-ComplexeMédical : NOZ3j)

**Les intervalles de variation du nombre d'objets spatiaux :  $I_{nm}$**

Les intervalles des entiers  $I_{n1}$  ( $k_1=3$ ) :  $I_{11} = [0,3[$ ,  $I_{21} = [3,5[$ ,  $I_{31} = [5,7[$

Les intervalles des entiers  $I_{n2}$  ( $k_2=3$ ) :  $I_{12} = [0,2[$ ,  $I_{22} = [2,3[$ ,  $I_{32} = [3,4[$ .

Les intervalles des entiers  $I_{n3}$  ( $k_3=3$ ) :  $I_{13} = [0,1[$ ,  $I_{23} = [1,2[$ ,  $I_{33} = [2,2]$ .

Le tableau ci-dessous présente le numéro et l'item associé à chaque intervalle de variation du nombre d'objets spatiaux par zone afin de simplifier et comprendre les itemsets fréquents et les règles d'association zonales, générées ultérieurement par le programme d'extraction des règles d'association.

**La base de transaction**

Cybers			Etablissements			Complexes		
X1	X2	X3	X4	X5	X6	X7	X8	X9
<b>0</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>
1	0	0	0	0	1	0	1	0
1	0	0	1	0	0	0	1	0
1	0	0	1	0	0	0	0	1
1	0	0	1	0	0	0	1	0
1	0	0	1	0	0	0	1	0
1	0	0	0	0	1	0	0	1
1	0	0	1	0	0	0	1	0

Le « 0 » indique l'absence de l'item et le « 1 » indique son existence.

Par exemple, la première ligne (en gras) du tableau ci-dessus illustre que dans la 1<sup>ère</sup> zone :

- Le nombre de cybers appartient à  $I_{21} = [3,5[$  (numéro 2)  $\Rightarrow$  le nombre est moyen
- Le nombre des établissements appartient à  $I_{12} = [0,2[$  (numéro 4)  $\Rightarrow$  le nombre est faible
- Le nombre de complexes médicaux appartient à  $I_{23} = [1,2[$  (numéro 8)  $\Rightarrow$  le nombre est moyen

**Outil de génération des règles d'associations zonales**

Pour extraire les motifs fréquents et les règles d'association, nous avons implémenté l'algorithme *Apriori* en java. L'outil de

génération de règles réalisé est donné à la figure 1.

Pour générer les règles d'association :

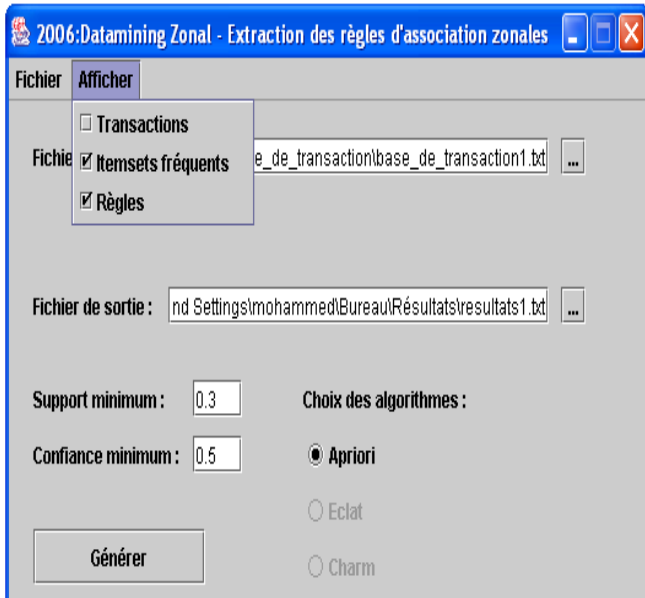


Figure 1. L'outil de génération des règles d'associations

- Charger le fichier source (base de transactions) en utilisant l'entrée prévue à cet effet.
- Préciser le fichier de sortie où les règles d'association seront enregistrées
- Saisir le support minimum et la confiance minimum (par défaut : support minimum = 0.1 et confiance minimum = 0.5).
- Choisir l'algorithme (p.ex., *Apriori*)
- Choisir le format d'affichage des résultats à l'aide de l'entrée « Afficher » du menu. Les modes d'affichage sont : Transactions, itemssets fréquents ou règles.
- Cliquer sur 'Générer'

Dans la suite, nous donnons les résultats obtenus pour un nombre de zones = 36. Support minimum = 0.3 et Confiance minimum = 0.5

### ***Interprétation des résultats concernant la liste des itemsets fréquents***

- 78% de zones de la ville Mohammadia ont un nombre faible de cybers
- 64% de zones de la ville Mohammadia contiennent un nombre faible de cybers et un nombre faible d'établissements
- 58% de zones de la ville Mohammadia contiennent un nombre faible de cybers et un nombre faible de complexes médicaux
- 50% de zones de la ville Mohammadia contiennent un nombre faible de cybers , un nombre faible d'établissements et un nombre faible de complexes médicaux.

### ***Interprétation des résultats concernant les règles d'association***

- Dans les zones où on a un nombre faible d'établissements, on trouve aussi un nombre faible de cybers avec un pourcentage de 79.31%.
- Dans les zones où on a un nombre faible de complexes médicaux, on trouve aussi un nombre faible de cybers avec un pourcentage de 75%.
- Dans les zones où on a un nombre faible d'établissements et un nombre faible de complexes médicaux, on trouve aussi un nombre faible de cybers avec un pourcentage de 78.26%.

## **6 Conclusion**

Dans cet article, nous avons proposé une approche d'extraction des règles d'association qui se base sur la notion de *zone* (qui peut être d'une forme quelconque) au lieu de la notion de distance utilisée par les techniques d'extraction des règles d'association spatiales classiques. Cette technique est particulièrement intéressante pour contourner le problème de géocodage. Nous avons également proposé de transformer la base de données spatiale en une base de transactions afin de supporter les algorithmes d'extraction de règles d'associations qui ont été conçus



dans le contexte des bases de transactions. Notre démarche a été illustrée par une application réalisée à l'aide d'un outil de génération de règles d'association que nous avons développé à cet effet. Les résultats obtenus sont très encourageants. Pour consolider davantage ces résultats, nous envisageons d'appliquer cette démarche de datamining zonal à d'autres domaines.

## 7 Bibliographie

- [1] R. Agrawal, T. Imielinski et A.N. Swami. Mining association rules between sets of items in large databases. *In: Proc. of the 1993 ACM SIGMOD Int. Conf. on Management of Data*, pp. 207-216, 1993.
- [2] ANRT, Marché de l'Internet au Maroc : Etat des lieux et perspectives de développement, Décembre 2004.
- [3] R. J. Bayardo. Efficiently mining long patterns from databases. *In ACM SIGMOD Conf. Management of Data*, June 1998.
- [4] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy. *Advances in Knowledge Discovery and Data Mining*, AAAI Press / MIT Press (1996).
- [5] J. Han, M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, San Francisco, CA, 2001.
- [6] K. Koperski, J. Han. Discovery of spatial association rules in geographic information databases, *4th Int. Symp. Advances in Spatial Databases, SSD, vol.951*, pp. 47-66. Springer-Verlag, 1995.
- [7] R. Laurini, D. Thompson, *Fundamentals of Spatial Information Systems*, Academic Press, London, UK (1994).
- [8] D.I. Lin, Z.M. Kedem. Pincer-Search: A New Algorithm for Discovering the Maximum Frequent Set, *Lecture Notes in Computer Science*, 1997.
- [9] P.A. Longley, M.F. Goodchild, D.J. Maguire, D.W. Rhind. *Geographic Information Systems*, Volume 1, Wiley (1999).
- [10] R. Marghoubi, *Fouille de données spatiales, Utilisation des Treillis de Galois pour l'extraction des règles d'association spatiales, application au domaine du Service Universel des Télécommunications*, thèse de doctorat, sous la direction de A. Boulmakoul, mai 2006.
- [11] N. Pasquier, Y. Bastide, R. Taouil, L. Lakhal. Pruning closed itemset lattices for associations rules, *BDA 1998*.
- [12] N. Pasquier, Y. Bastide, R. Touil, L. Lakhal. Discovering frequent closed itemsets for association rules. *In Proc. of 7th Int. Conf. on Database Theory (ICDT'99)*, LNCS, volume 1540, Springer-Verlag, p. 398-416, Jan. 1999.
- [13] J. Pei, J. Han, R. Mao, S. Nishio, S. Tang, D. Yang. Closet: An efficient algorithm for mining frequent closed itemsets. *In Proc. of the ACM-SIGMOD DMKD'00*, Dallas, Texas, USA, p. 21-30, 2000.
- [14] M. J. Zaki, C. J. Hsiao. Charm: An efficient algorithm for closed itemset mining. *In Proc. of the 2nd SIAM Int. Conf. on Data Mining, Arlington, Virginia, USA*, p. 34-43, April 2002.
- [15] K. Zeitouni, L. Yeh, Les bases de données spatiales et le data mining spatial, *Revue internationale de géomatique, Numéro spécial « Data mining spatial »*, Vol. 9, N° 4 (99). Egalement dans les actes des *Journées sur le Data Mining spatial et l'analyse du risque*, Versailles (2000).