

Extraction de connaissances des bases de données spatiales en accidentologie routière

Florence Huguenin-Richard^{*}, Sylvain Lassarre^{**}, Laurent Yeh[☒], Karine Zeitouni[☒]

^{*} *THEMA - Université de Franche-Comté*
32, rue Mégevand - 25030 Besançon Cedex

^{**} *DERA / INRETS*
2, Avenue Malleret-Joinville - 94114 Arcueil

[☒] *PRISM - Université de Versailles Saint-Quentin en Yvelines*
45, Avenue des Etats Unis - 78035 Versailles Cedex

1. Introduction

De nos jours, les applications de Système d'Information Géographique (**SIG**) aux problèmes d'aménagement et de gestion des territoires sont en pleine expansion. Ainsi, le nombre de bases de données géographiques et le volume de données associées ne cessent d'augmenter, notamment grâce aux techniques nouvelles de géocodage par adressage postal d'une information jusque là a-spatiale (accidents de la route, lieu de crime...).

L'accidentologie, discipline qui s'intéresse à l'analyse des accidents et du risque routier, connaît depuis quelques années une utilisation accrue de l'information sur la localisation géographique. En effet, la diffusion par le ministère de l'Équipement de Concerto - un applicatif basée sur le SIG *MapInfo* et dédié à la connaissance de l'insécurité routière - permet une généralisation de cette approche géographique.

Les possibilités de stockage, d'affichage, d'interrogation et de sélection offertes par les SIG facilitent un accès rapide et une exploitation visuelle des données. Ces fonctionnalités de base des SIG constituent déjà une réelle avancée technique pour le suivi de l'insécurité routière sur un territoire.

Cependant, la richesse de ces bases de données géographiques est souvent sous-exploitée. En effet, l'analyste se trouve très vite submergé par la masse considérable d'information archivée :

- Comment synthétiser les données ?
- Comment extraire les cartes pertinentes ?
- Comment générer de l'information utile sans multiplier le nombre de manipulations des données ?

De plus, les SIG bureautique sont encore aujourd'hui faibles en fonctions d'analyse statistique. Cette insuffisance conduit souvent à la nécessité de compléter les SIG par d'autres outils. Ce qui amène à des solutions plus complexes et moins intégrées. Par

ailleurs, même si l'organisation en couches thématiques superposées est la règle dans les SIG, l'analyse simultanée de ces couches reste limitée.

La suite du papier est organisée comme suite : la section suivante décrit les objectifs du projet. La section 3 présente les données utilisées dans le projet. La section 4 présente la méthode de généralisation intégrée au prototype. Puis nous présenterons dans la section 5 la génération de matrices de voisinages. Cette section sera suivie de la conclusion.

2. Objectifs du projet et moyens mis en œuvre

Les objectifs du projet sont de développer et d'intégrer dans un environnement SIG des outils d'exploration et d'analyse des données spatiales. Ce projet se situe dans le contexte plus générale du *Data Mining* (*Fouille de données* en français) et plus spécifiquement dans le champ de recherche en *Data Mining Spatial*. Ces objectifs sont menés en collaboration entre les informaticiens, les géographes et les accidentologues. Ils visent à faciliter l'extraction automatique ou semi-automatique de connaissances pertinentes à l'accidentologue à partir :

- de la composante spatiale et sémantique de l'information sur les accidents ;
- des relations de voisinage que les différents objets géographiques modélisés dans la géobase entretiennent entre eux dans la réalité.

Le terme de Data Mining est souvent employé pour désigner un « *ensemble d'outils permettant d'extraire automatiquement des connaissances intéressantes et intelligibles dans les bases de données (règles, régularités, patterns, associations...) et de découvrir des modèles implicites* » [FRANCO 97]. Ainsi, l'objectif principal est de découvrir des modèles difficiles à mettre en évidence, soit en raison du volume important de données, soit à cause de la quantité de variables à considérer, soit enfin que ces modèles sont imprévisibles et ne sont jamais envisagés par l'analyste même à titre d'hypothèses à vérifier [ZEITOUNI 99].

Le Data Mining se présente comme un catalogue de méthodes, réalisant des tâches fonctionnelles et qui peuvent être classées en deux familles. La première est orientée vers la découverte et l'exploration (description synthétique, classification, recherche de tendances) et la seconde réunit des analyses dans un objectif décisionnel ou prédictif (recherche de règles, modélisation). Hormis l'aspect fonctionnel, vu qu'il opère sur de gros volumes de données, le Data Mining suscite des questions d'optimisation du temps d'exécution. En outre, le processus de Data Mining n'est pas une exploitation simple des données mais un processus complexe faisant appel à des techniques de visualisation, de gestion de bases de données, d'analyse de données statistiques et de l'intelligence artificielle.

Le Data Mining Spatial hérite du Data Mining, mais en plus, il exploite les relations spatiales contenues explicitement ou implicitement dans la base de données. Ce domaine est au carrefour des techniques du Data Mining, de bases de données, de systèmes d'information géographique et de la statistique spatiale. Cependant en raison du volume des données spatiales plus important et de la complexité des traitements pour la prise en compte de relations spatiales, les performances et l'optimisation des algorithmes sont des problèmes importants à prendre en compte.

Ainsi, différents algorithmes ont été étudiés dans le SIGArcView sur lequel est implantée la base de données géographiques des accidents de la route. Ces techniques de Data Mining directement intégrées permettent d'augmenter les capacités d'extraction de connaissances : le fameux « abstraire » de la règle des « six A » définissant les principales fonctionnalités d'un SIG (« *Acquérir, Archiver, Accéder, Afficher, Analyser, Abstraire* » [DENEGRE 96]). Ils permettent à l'analyste de mener de manière automatique et conviviale différentes opérations sur la base de données géographiques. Ces opérations consistent à :

- généraliser l'information sémantique et spatiale ;
- synthétiser et caractériser l'information initiale afin d'en tirer les principales propriétés ;
- formaliser les relations sémantiques et spatiales entre les objets géographiques d'un même thème (contiguïté entre les communes) ou entre des thèmes différents (distance entre les accidents et les écoles).

3. La base de données géographiques sur les accidents de la route

En France, les accidents corporels de la circulation font l'objet d'un recueil systématique de données, renseignant les caractéristiques de l'événement et de l'infrastructure sur laquelle il est survenu, sa gravité, les véhicules et les usagers impliqués (Bordereaux d'Analyse des Accidents Corporels). Cette base de données est la source officielle de l'insécurité routière. Son exploitation permet le suivi de l'évolution du risque routier, qui repose sur un état des faits passés. Actuellement, la tendance en accidentologie est une généralisation des bases de données spatiales et leur intégration au sein de système d'information géographique. Chaque accident est alors un événement ponctuel, localisé par ses coordonnées géographiques. Cette information comprend une composante topologique (localisation de l'objet) et une composante sémantique (propriétés, données attributaires attachées). Cependant, le coût du géocodage nécessaire à l'acquisition de ces données explique en partie que toutes les collectivités territoriales ne possèdent pas encore de telles bases.

La communauté urbaine de Lille est un territoire fortement urbanisé qui regroupe quatre-vingt-sept communes dont les villes de Lille, Roubaix, Tourcoing et Villeneuve-d'Ascq. Depuis le début des années 1980, une base de données spatiales sur les accidents corporels a été développée. Elle recense et localise de façon très précise le lieu de chaque événement. Aujourd'hui, cette base de données est riche de près de 30000 cas et décrit une vingtaine de variables. Chaque accident est localisé par une paire de coordonnées spatiales, le numéro de tronçon, de carrefour ou de commune sur lequel il est survenu. Des renseignements d'un niveau de détail parfois très fin sur l'heure, la date, le jour, le mois... sont portés. Les données exposent aussi les types d'usagers impliqués (dix classes différentes), la gravité, la cause de l'accident (cinquante possibilités), la luminosité, l'état de la chaussée...

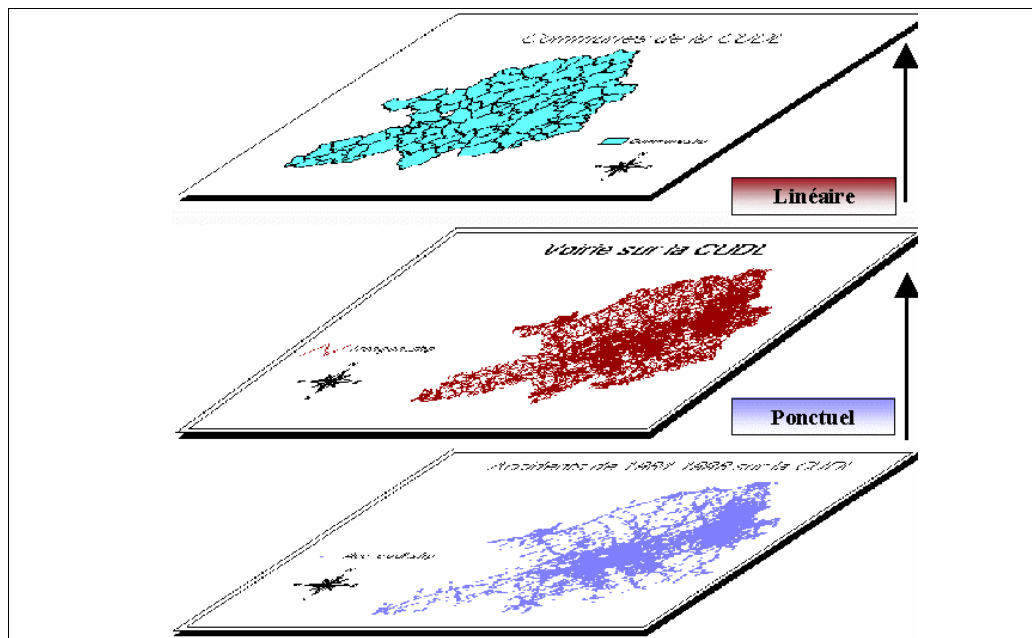
La base de données géographiques utilisée pour ce projet contient des données spatiales de type vecteur et est représentée dans un modèle relationnel. L'environnement de l'accident est renseigné par différentes couches thématiques :

- le réseau de circulation est décrit par un ensemble de 37612 polygones ou tronçons de longueur variable, qui correspondent à une portion de voie entre deux carrefours. Les informations attachées concernent le type de voie, le sens de circulation, l'état de la chaussée, l'état des trottoirs, le trafic, le type d'urbanisation environnant... Une couche carrefour a été extraite à partir de la géométrie du thème voie ;
- les limites administratives concernent le découpage en commune et en quartier pour les unités urbaines les plus importantes. Des informations sur la population, le niveau d'équipement des ménages, les pratiques de mobilité peuvent être attachées ;
- l'environnement urbain est modélisé par la connaissance de la localisation d'un ensemble très varié d'objets géographiques (écoles, bâtiments publics, zones commerciales, zones d'activités, espaces verts et terrain de jeu, trame du bâti...).

Dans cette base de données, des relations spatiales entre les entités existent. Certaines sont explicites. Ainsi dans la table attributaire «accidents», une relation spatiale sémantique est décrite entre la localisation précise d'un accident, le tronçon auquel il se rattache et la commune sur laquelle il est survenu. D'autres sont implicites et doivent être établies, telles que les relations spatiales entre les objets géographiques :

- relations au sein d'un même thème (contiguïté entre les tronçons) ;
- relations inter thèmes (si il existe deux thèmes, l'un pour les accidents et l'autre pour les écoles, la relation inter- thèmes est par exemple la distance entre les accidents et les écoles).

Figure 1. Présentation du modèle cartographique du SIG sur les accidents de la route



4. Généralisation d'information et caractérisation

L'objectif d'une opération de généralisation est d'extraire d'une base de données des informations généralisées ou synthétisées. C'est un moyen efficace pour résumer le contenu d'une base. Il existe deux cas de généralisation qui peuvent être sémantique ou spatiale.

4.1 Généralisation sémantique versus spatiale

Dans le premier cas, l'analyste cherche à résumer une information attributive très détaillée en un ensemble de nouvelles variables plus synthétique. Par exemple, l'attribut « jour » à sept possibilités contenues dans le fichier « accidents » peut être généralisé en un nouvel attribut à deux alternatives : semaine et week-end comme illustré par la figure 2a. Ce traitement utilise une méthode dite d'induction orientée attribut [HAN 92] afin de produire automatiquement un niveau de détail souhaité.

Dans le cas d'une généralisation spatiale, l'information est agrégée à un niveau d'échelle géographique plus haut. Le passage d'un niveau d'observation du risque à un autre, par exemple des communes à un département, est une opération plus connue en géographie sous le terme d'agrégation de données. Il s'agit de regrouper l'information contenue par un ensemble d'unités spatiales à une échelle donnée (la densité d'accidents par tronçon à l'échelle de la commune) en un ensemble plus agrégé (la densité d'accidents en fonction de la longueur du réseau par commune à l'échelle de la communauté urbaine de Lille). Cette opération nécessite une fonction pouvant être intégrée dans un SIG. Néanmoins, des difficultés apparaissent lorsque les données à agréger sont de nature qualitative ou lorsque la masse d'information est trop importante, ce qui multiplie le nombre de manipulations des tables.

Commentaire [y1] : Page: 7

Du point de vue des traitements, dans les deux cas, l'opération s'accompagne :

- d'une fusion des objets et des variables par re-calcul des valeurs prises (somme, moyenne... pour les valeurs quantitatives ; pourcentage, comptage... pour les valeurs qualitatives) ;
- d'un changement de données (nature, forme de la distribution statistique...) ;
- d'un apport de connaissances puisque des informations nouvelles sont acquises.

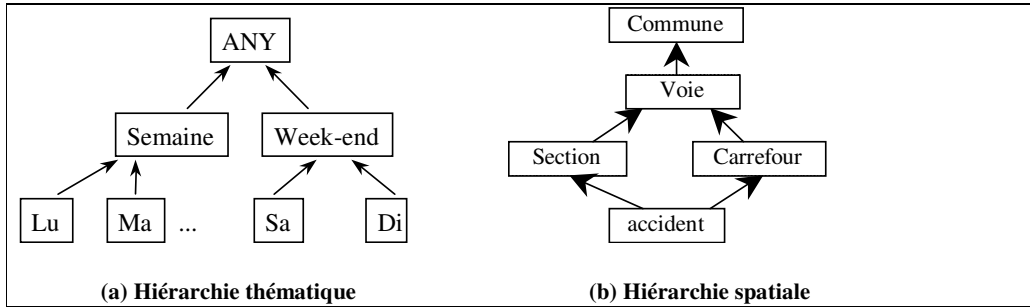
Les méthodes de généralisation d'information utilisées reposent sur la formalisation d'une hiérarchie de concepts (voir figure 2). Les attributs constituant la base de données sont organisés sous une forme hiérarchique qui exprime des connaissances synthétiques et des relations au sein de cette base. La hiérarchie décrit le passage des concepts les plus spécifiques aux concepts plus généraux de niveau supérieur. Deux types de hiérarchie de concepts sont formalisés :

- des hiérarchies thématiques, explicitement exprimées par l'expert ou implicitement contenues dans la base de données. La généralisation peut alors être faite automatiquement ;
- des hiérarchies spatiales, telles qu'un découpage administratif. Dans ce cas, elle peut être déduite par application d'opérateurs spatiaux sur le SIG (inclusion, intersection, plus proche voisin...). Mais, assez souvent ces relations sont représentées dans le schéma de la base de données. Par exemple, on connaît d'emblée le rattachement d'un accident à tel tronçon ou à telle commune. Ce cas est intéressant à considérer car il permet l'économie de l'opérateur spatial.

4.2 Mise en oeuvre

Dans notre implémentation, trois algorithmes de généralisation ont été intégrés dans *ArcView* : une généralisation à dominante non spatiale reposant sur la définition d'une hiérarchie de concept sémantique, une généralisation à dominante spatiale reposant sur une hiérarchie de concept spatiale (utilisant les opérateurs spatiaux du logiciel) et une généralisation à dominante spatiale reposant sur une hiérarchie de concept structurelle (utilisant les relations spatiales décrites dans la table).

Figure 2. Présentation de hiérarchie de concepts



4.3 Généralisation pour l'accidentologie

La hiérarchie spatiale utilisée pour l'analyse des accidents de la route se compose de cinq niveaux différents qui sont autant d'échelle d'observation du risque. À petite échelle, l'individu statistique est une commune, des sections de route sur les grands axes de communication ; à grande échelle, c'est un tronçon de voie, un carrefour, un quartier, la maille d'une grille. La hiérarchie sémantique est d'autant plus intéressante lorsque les modalités prises par un attribut sont importantes en nombre, comme c'est le cas pour la variable « cause ». Nous proposons en exemple une hiérarchie de concept sémantique appliquée sur la table « accidents ». La définition des hiérarchies de concepts est totalement ouverte. Ainsi, l'utilisateur peut au gré de ses besoins ajouter ou supprimer un niveau de détail.

Un certain nombre de calculs spécifiques à l'analyse du risque routier ont été programmés et intégrés dans les scripts de généralisation d'information. Ainsi, il est possible d'opérer de façon automatique :

- des comptages, des sommes, des moyennes... (par exemple le nombre de tués par type d'utilisateur par commune) ;
- des proportions (proportion d'accidents de piéton par commune par rapport à l'ensemble des accidents) ;
- des taux de risque (nombre d'accidents par rapport à la population par commune) ;
- des taux de gravité (nombre de victimes par rapport au nombre d'accidents).

Figure 3. Présentation de la fenêtre de dialogue

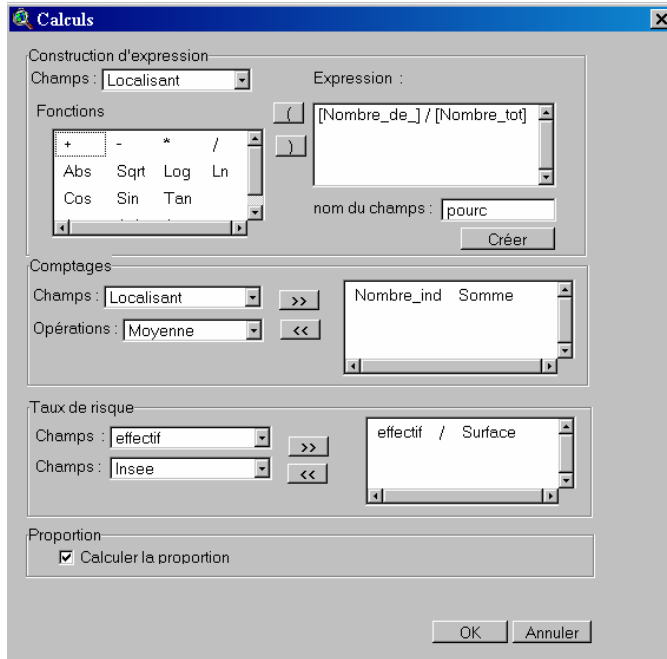
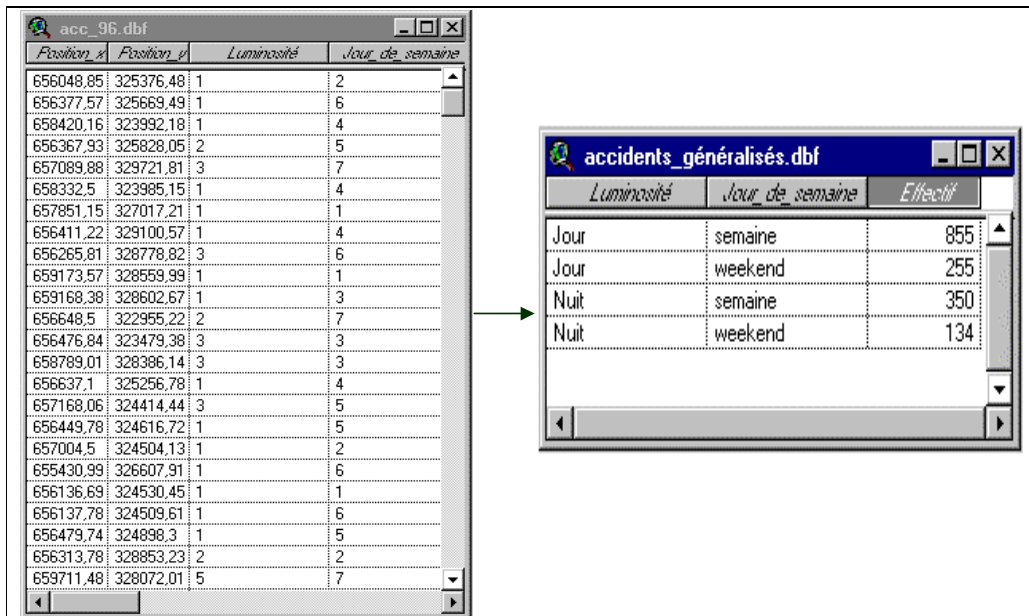


Figure 4. Présentation du résultat d'une généralisation sémantique



L'application à notre base de données spatiale a montré son importance dans le processus d'analyse. En effet, la généralisation tend à réduire le nombre de modalités des variables en vue d'une étude particulière et restitue l'information nouvelle sous forme de

tableaux ou de cartes (par exemple, la figure 4. montre la généralisation des données selon les hiérarchies sur les attributs « date » et « luminosité » des accidents).

4.4 La caractérisation

Une autre approche pour synthétiser les données spatiales est la caractérisation. Cette approche consiste à déduire de règles caractéristiques d'une partie de la base de données à analyser [Ester 98]. Ces règles sont définies par les caractéristiques qui sont plus fréquentes pour un sous ensemble de la base de données par rapport à leur fréquence relative dans le reste de la base. La spécificité de ces règles caractéristiques en base de données géographiques est qu'elles découvrent en plus de la fréquence de ces propriétés leur niveau d'extension aux voisins. Pour cela, elles nécessitent la définition d'un graphe (ou matrice) de voisinage. Plus précisément, étant donné un sous-ensemble S d'objets à analyser, les paramètres (i) de signifiante (c'est à dire le seuil de fréquence relative dans S), (ii) de confiance qui donne la proportion d'objets dans S qui satisfont le seuil de signifiante dans leur voisinage et (iii) l'extension maximum aux voisins, cette méthode découvre les propriétés $p_i = (\text{attribut}, \text{valeur})$, la fréquence relative (au-dessus du seuil de signifiante) et le nombre n_i de voisins auxquels s'étend cette fréquence de la propriété. La caractérisation peut s'exprimer par une règle :

$$S \rightarrow p_1(n_1, \text{freq}_1) \text{ et } \dots p_k(n_k, \text{freq}_k)$$

Un exemple dans notre application est la caractérisation des accidents mortels par rapport à l'ensemble des accidents ? La méthode met en évidence qu'ils se produisent plus fréquemment la nuit et le week-end et que les causes sont humaines ou indéterminées. Cette dernière propriété s'étend aux voisins, ce qui veut dire que localement, les zones où se produisent ces accidents graves sont des zones fortement accidentogènes. Cette méthode est complémentaire dans le processus de diagnostic des lieux dangereux en sécurité routière.

5. La génération de matrices de voisinage et la jointure

Les données géographiques sont caractérisées dans la réalité par des relations spatiales, le plus souvent modélisées de manière implicite dans SIG. Or, ces relations présentent une grande importance dans l'analyse de phénomène tel que le risque routier, car la prise en compte de l'influence des unités spatiales voisines, et analyse portant sur la localisation d'accidents à distance d'un point source d'insécurité tel que les écoles, ont un impact sur l'étude des phénomènes.

5.1 Matrice de voisinage et jointure spatiale

Pour déterminer ces liens au sein d'une base de données, il est nécessaire de construire des matrices de voisinage. Pour cela l'analyste doit mettre en oeuvre des jointures utilisant des critères spatiaux. Le coût d'exécution des jointures peut être important et pose le problème de performance. En effet, il est bien connu que certaines jointures spatiales, sans techniques spécifiques ont un coût algorithmique en $O(n^2)$.

L'importance du problème des performances a fait l'objet de nombreux travaux. Une part importante de ces travaux s'appliquent essentiellement pour le critère de recouvrement

(ou map overlay) [Chrétien 94, Brinkhoff 94]. Des travaux plus récents concernent la jointure basée sur un critère de distance [Lu 92, Yeh99]. Mais d'autres critères sont également importants, tels que la contiguïté ou l'inclusion stricte [Rotem 91, Gunther 93].

5.2 Utilisation des matrices de voisinages comme accélérateur

Dans le prototype, l'approche proposée est originale dans ce contexte car elle combine l'utilisation directe des matrices de voisinages nécessaires pour l'analyse et leurs utilisations comme accélérateur de la jointure [Zeitouni 00]. L'idée est de constituer un index, extension des indices de jointure proposés par Valduriez [Valduriez 87]. L'extension consiste à rajouter les distances pré-calculées des distances entre les objets géographiques. Cette structure est similaire au matrice de voisinage (« graphe de voisinage ») bien connu en analyse. L'optimisation pour la jointure consiste à s'en servir comme pré-filtre lors des requêtes de jointures spatiales. En effet, toutes ces requêtes peuvent être résolues à l'aval par simple sélection dans l'index constitué par les indices de jointures. D'où l'optimisation des performances par rapport à une utilisation des algorithmes directement existant dans les SIG bureautiques. Ainsi, on peut calculer efficacement les relations topologiques de contiguïté ou d'inclusion ou métriques sur critère de distance. Bien évidemment, il n'est pas utile de stocker les distances entre les objets très éloignés. Seuls les objets ayant une distance raisonnable en fonction du sujet de l'analyse (définie par le concepteur de l'index) sont stockés dans l'index. Ce qui optimise à la fois la construction, le stockage et la recherche dans l'index.

En conséquence, cette structure présente plusieurs intérêts, particulièrement dans l'application au data mining spatial. D'un côté, elle s'applique à différents types de jointures (et pas seulement au map overlay comme le font d'autres travaux). D'un autre côté, elle correspond aux structures exploitées dans l'approche d'analyse de données (« graphe de voisinage »).

5.3 Implémentation

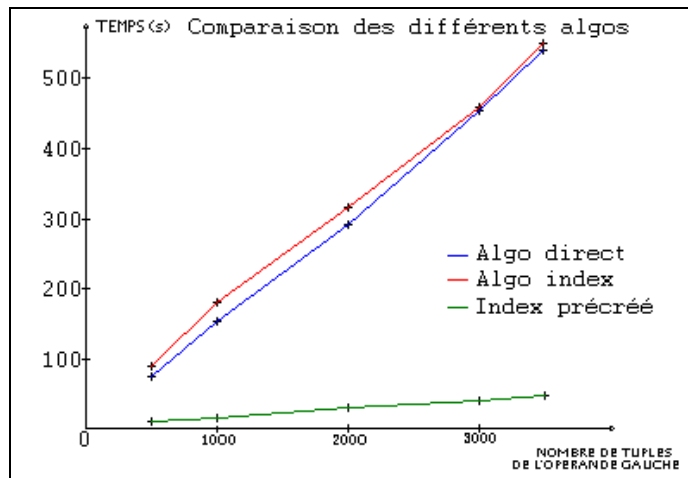
Une attention particulière a été portée, dans l'implémentation, au paramétrage du critère de jointure. Ainsi, pour le calcul de distance on peut choisir entre la distance centre à centre, bord à bord, ou par le plus court chemin. On peut se restreindre à un rayon (à moins d'une distance donnée) ou dans un anneau (entre deux distances données). Il est possible aussi de rechercher les k plus proches voisins au sens de ces distances (permettant la segmentation en « groupe d'objets proches »). Quant aux relations de voisinage, elles peuvent être considérées à différents degrés : voisins directs, voisins de voisins et ainsi de suite).

Plusieurs algorithmes ont été implémentés pour les tests de performances : un algorithme naïf par produit cartésien, un algorithme direct qui remplace le parcours d'une relation par l'accès par index spatial, un algorithme intégrant la construction de l'index de jointure et enfin l'algorithme de jointure avec index pré-construit. Les mesures ont porté sur des volumes de données allant d'une centaine de tuples jusqu'à 30000 tuples pour différents critères. L'algorithme naïf a été écarté en raison de son coût exorbitant (70 minutes au lieu de 3 minutes pour l'algorithme direct).

Les autres mesures de performances montrent le gain obtenu par la jointure avec index par rapport à l'algorithme direct. En outre, l'index est généré une fois et est exploitable plusieurs fois. L'essentiel est donc de réduire le coût de l'opérateur lorsque l'index est

préexistant. Néanmoins, nous avons cherché à éviter des coûts exorbitants de construction de l'index en utilisant une variante de l'algorithme direct. Par conséquent, comme le montre la figure 5, cette étape n'introduit presque pas de surcoût (différence entre algorithme direct et algorithme index qui intègre la création de l'index).

Figure 5. Mesure des performances des jointures spatiales avec et sans index



6. Conclusion

Dans le cadre de l'analyse spatiale du risque routier, la prise en compte des relations géographiques est fondamentale pour comprendre les interactions spatiales pouvant exister entre les objets et extraire de nouvelles connaissances. Ainsi, la construction de matrice de voisinage trouve de nombreuses applications.

De nombreuses méthodes en analyse spatiale reposent sur la formalisation des proximités spatiales. Ainsi les mesures de corrélations spatiales, bien connues en analyse spatiale, permettent de caractériser la forme de la répartition géographique d'une variable sur un territoire [Anselin, 1995, Cliff, 1973, Pumain 1997]. Ces calculs, pourtant à la base de toute étude spatiale, sont très difficiles à mettre en œuvre sur des systèmes d'information géographique bureautique du fait de la complexité de générer les matrices de voisinages. Les développements actuels facilitent ainsi grandement la tâche.

Ce papier décrit la réalisation faite sur le SIG ArcView. Dans ce prototype a été intégré des opérateurs de généralisation. En outre, pour l'analyse, un opérateur a été ajouté pour pouvoir calculer les matrices de contiguïté. Ces matrices sont également utilisées comme accélérateur pour le traitement de la jointure spatiale. Ces différentes fonctionnalités contribuent à offrir un outil plus puissant pour le data mining et l'analyse spatiale.

Références

Anselin L., 1995 : "Local Indicators of Spatial Association - LISA", Geographical Analysis, Ohio State University press, Vol. 27, n° 2, pp. 93-115

- Brinkhoff T., Kriegel H.-P., Schneider R., Seeger B., "Multi-Step Processing of Spatial Joins", Proc. ACM SIGMOD Int. Conf. on Management of Data, Minneapolis, MN, 1994, pp. 197-208.
- Chrétien D., Quilio I., "La jointure spatiale par partitionnement", 5th European Conference and Exhibition on Geographical Information Systems (EGIS'94), Paris, France, Mars 1994, pp. 284-293.
- Cliff A.D., Ord J.K., 1973 : "Spatial autocorrelation", Pion, London.
- Ester M., Frommelt A., Kriegel H.-P., Sander J., "Algorithms for Characterization and Trend Detection in Spatial Databases", Proc. 4th Int. Conf. on Knowledge Discovery and Data Mining, New York, NY, 1998.
- Gunther O., "Efficient Computation of Spatial Joins", Proc of Data Engineering, Vienna, Austria, April 1990, pp. 50-59.
- Han J., Cai Y. & Cerone N., "Knowledge Discovery in Databases; An Attribute-Oriented Approach." Proceedings of the 18th VLDB Conference. Vancouver, B.C., August 1992, pp. 547-559.
- Lu W., and Han J., "Distance-Associated join indices for spatial range search", Proceeding of Eighth International Conference on Data Engineering, Tempe, Arizona, February 1992, pp. 284-292.
- Pumain D., St-Julien, 1997 : "L'analyse spatiale - 1. Localisation dans l'espace", Armand Colin, Paris, 167 p.
- Rotem D., "Spatial join indices", Proc. of 7th Conf. on Data Engineering, Kobe, Japan, 1991, 500-509
- Valduriez P., "Join Indices", ACM Transactions on Database Systems, 12 (2), June 1987, pp. 218-246.
- Yeh T-S., "Spot: Distance based join indices for spatial data", ACM GIS 99, Kansas City, USA, 5-6 Nov 1999, pp 103-110.
- Zeitouni K., Yeh L., "Le data mining et les bases de données spatiales", Revue internationale de géomatique, Vol. 9, N° 4 (99).
- Zeitouni K., "Index de jointures spatiales et application au data mining", Rapport Interne du Laboratoire PRiSM, Université de Versailles-Saint-Quentin, Référence 2000/12, Mai 2000.