

Outil de datamining spatial appliqué à l'analyse des risques liés au territoire

Schahrazed Zeghache*, Farida Admane*, Kamel Elaraba Ziane*

* Division Systèmes d'Information, CERIST
Rue des trois frères Aissiou Ben Aknoun, Alger, Algérie
zeghache@mail.cerist.dz, admane@wissal.dz, elaraba@mail.cerist.dz

1 Problématique

Tout géographe s'accorde à dire que tout phénomène à un endroit est lié à l'influence du voisinage (première loi en géographie). Ceci revient à dire que les données spatiales ne sont pas indépendantes et que leurs analyses nécessitent, en plus des caractéristiques des objets à analyser, la prise en compte des caractéristiques des objets du voisinage et des relations spatiales qui les relient.

2 Approche proposée

Dans notre état de l'art, nous avons recensé des insuffisances dans les outils d'analyses spatiales ; et afin d'y remédier, nous avons proposé une méthodologie pragmatique fondée sur des bases théoriques en tenant compte : de l'inexistence des entrepôts de données dans la majorité des organismes ; de la nature complexe des données à référence spatiale ; des limites des fonctionnalités analytiques des outils existants entre autre Systèmes d'Informations Géographiques (SIG) et datamining ; etc. Afin de résoudre cette problématique, nous proposons la combinaison d'un SIG avec un ensemble de techniques de datamining.

Dans un premier temps, nous avons proposé le cadre conceptuel permettant de définir la manière selon laquelle la combinaison devrait s'opérer. Ce cadre a été illustré par la présentation d'un enchaînement de phases devant constituer le processus décisionnel incluant un SIG et un ensemble de techniques d'extraction. Ce processus se présente comme suit :

Préparation des données : Consiste à préparer la base de données géographiques .

Consultation des données : Consiste à analyser les données préparées pour vérifier si les critères du problème posé ont été respectés.

Création d'un index de jointure : Cette étape est définie comme la spécificité du datamining spatial par rapport au datamining classique. Elle permet de pré-calculer la relation spatiale exacte entre les objets spatiaux de deux collections puis de les stocker dans une table, pour y appliquer les techniques de datamining pour une meilleure exploitation. La

méthode que nous avons développée pour la création des index de jointures est une méthode proposée par Zeitouni (2000). Cette méthode est une extension de l'index de jointure qui stocke la valeur de distance entre les objets. Contrairement aux anciennes méthodes proposées pour la jointure basée sur la distance, seuls les objets ayant une distance raisonnable (définie par le concepteur de l'index) sont stockés dans cet index. Ce qui optimise à la fois la construction et l'utilisation de l'index. Dans le cadre de cette étude, une attention particulière a été portée au paramétrage du critère de jointure.

Choix de l'algorithme de calcul : Notre étude porte sur une seule technique de datamining spatial qui est la classification supervisée par les arbres de décisions. Un arbre de décision a pour but de trouver les attributs explicatifs et les critères précis donnant le meilleur classement. L'arbre est construit par l'application successive de critères de subdivision sur une population d'apprentissage afin d'obtenir des sous populations plus homogènes. Dans le cadre de notre étude, nous proposons une extension de la méthode **CART** Zeitouni (2000) ainsi qu'une extension de la méthode **ID3** Zeitouni (2000). Ces deux dernières reposent sur le calcul d'un gain informationnel pour apprécier la subdivision.

Exécution de l'algorithme de calcul : Une fois que l'index de jointure est créé, nous pouvons lancer l'exécution de l'algorithme de classification par arbre de décision.

3 Etude de cas : « Analyse des risques liés au territoire »

L'objectif de cette étude de cas est d'expliquer et de prédire la dangerosité des routes en tenant compte de leurs positionnements ainsi que leurs voisinages. Cet objectif consiste à classer les données sur les accidents par rapport aux objets voisins (administration, école, cité, etc.) afin d'anticiper les aménagements correspondants.

Références

- Boulmakoul, A., N. Chelghoum, et K. Zeitouni (2001). Arbres de décisions spatiales multi thèmes. 8ème rencontres de la Société Francophone de Classification, 17-21 Décembre. Laboratoire PRISM (Versailles).
- Chelghoum, N., K. Zeitouni, and A. Boulmakoul (2002). A Decision tree for multi-layered spatial data. Joint International Symposium on Geospatial Theory, Processing and Applications, Ottawa, Canada, 8-12 juillet.
- Zeitouni, K., M. A. Aufaure, and L. Yeh (2000). Join indices as a tool for spatial datamining. International Workshop on Temporal, Spatial and Spatio-Temporal Data Mining, *Lecture Notes in Artificial Intelligence* n° 2007, Springer, pp 102-114, Lyon, France.

Summary

Our objective is to provide a tool allowing to explore the spatial databases in order to find there knowledge hidden and to create and manage the spatial relations between the data and to effectively assist the analyst in his resolution of the problems .