

Fouille de données spatiales

Approche basée sur la programmation logique inductive

Nadjim Chelghoum*, Karine Zeitouni**, Thierry Laugier*, Annie Fiandrino*,
Lionel Loubersac*

* Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER)
Laboratoire Environnement- Ressources du Languedoc Roussillon (LER/LR)
BP 171, Boulevard Jean Monnet,
34203 Sète Cedex
Tel/Fax (0)4 99 57 32 83/ (0)4 99 57 32 96
prénom.nom@ifremer.fr

** Laboratoire PRiSM, Université de Versailles
45, avenue des Etats-Unis, 78035 Versailles Cedex, France
Tel / Fax: (0)1 39 25 40 46 / (0)1 39 25 40 57
prénom.nom@prism.uvsq.fr

Résumé. Ce qui caractérise la fouille de données spatiales est la nécessité de prendre en compte les interactions des objets dans l'espace. Les méthodes classiques de fouille de données sont mal adaptées pour ce type d'analyse. Nous proposons dans cet article une approche basée sur la programmation logique inductive. Elle se base sur deux idées. La première consiste à matérialiser ces interactions spatiales dans des tables de distances, ramenant ainsi la fouille de données spatiales à la fouille de données multi-tables. La seconde transforme les données en logique du premier ordre et applique ensuite la programmation logique inductive. Cet article présentera cette approche. Il décrira son application à la classification supervisée par arbre de décision spatial. Il présentera aussi les expérimentations réalisées et les résultats obtenus sur l'analyse de la contamination des coquillages dans la lagune de Thau.

1 Positionnement du problème

La fouille de données spatiales (FDS) consiste à appliquer la fouille de données aux données spatiales (Zeitouni, 2000), (Han et Kamber, 2001), (Shashi et Sanjay, 2003). Contrairement aux données classiques, ces données sont de nature dépendantes (Anselin, 1989), (Longley et Goodchild, 1999) car tout phénomène spatial est influencé par son voisinage. Par exemple, la contamination des coquillages dans une lagune est influencée par les champs d'agriculture autour. Cette notion de dépendance entre les données est justifiée par une loi en géographie qui stipule que « *ce qui se passe dans une localité particulière dépend de ce qui se passe dans d'autres localités et ces interactions sont d'autant plus fortes que les localités concernées sont plus proches* » (Tobler, 1979). Du point de vue analyse, ceci revient à dire que l'analyse spatiale nécessite l'analyse des entités spatiales en fonction de leurs caractéristiques, des caractéristiques de leurs voisins et des relations spatiales avec ces voisins. Analyser ces données sans considérer cette spécificité génère des résultats incorrects (Anse-