

Vers une approche efficace d'extraction de motifs spatio-séquentiels

Hugo Alatrasta Salas ^{*,***}, Sandra Bringay ^{**}, Frédéric Flouvat ^{***},
Nazha Selmaoui-Folcher ^{***}, Maguelonne Teisseire ^{*,**}

^{*} Irstea, UMR TETIS, 500 rue Jean-François Breton, 34093 Montpellier - France
{prénom.nom}@teledetection.fr

^{**} LIRMM, UMR 5506, 161 rue Ada, 34392 Montpellier - France
{prénom.nom}@lirmm.fr

^{***} PPME - Université de la Nouvelle-Calédonie, BP R4, Nouméa, Nouvelle-Calédonie
{prénom.nom}@univ-nc.nc

Résumé. Ces dernières années, l'augmentation de la quantité d'informations spatio-temporelles stockées dans les bases de données a fait naître de nouveaux besoins, notamment en matière de gestion des risques naturels, sanitaires ou anthropiques (p. ex. compréhension de la dynamique d'une épidémie de Dengue). Dans cet article, nous définissons un cadre théorique pour l'extraction de motifs spatio-séquentiels, séquences de motifs spatiaux représentant l'évolution dans le temps d'une localisation et de son voisinage. Nous proposons un algorithme d'extraction efficace qui effectue un parcours en profondeur en s'appuyant sur des projections successives de la base de données. Nous introduisons également une mesure d'intérêt adaptée aux aspects spatio-temporels de ces motifs. Les expérimentations réalisées sur des jeux de données réels soulignent la pertinence de l'approche proposée par rapport aux méthodes de la littérature.

1 Introduction

Dans la vie quotidienne, nous pouvons observer de nombreux phénomènes se déroulant dans l'espace et dans le temps simultanément. Par exemple, les informations décrivant des trajets associent des informations spatiales (les coordonnées de la ville de départ et d'arrivée) et temporelles (l'heure de départ et d'arrivée). D'autres applications, avec des dynamiques plus complexes, sont beaucoup plus difficiles à analyser. Par exemple, dans le cas des épidémies de dengue, les experts en santé publique savent que l'évolution de la maladie dépend de facteurs environnementaux (p. ex. climat, proximité des zones infectées en points d'eau, mangroves...) et des interactions entre les humains et le vecteur de transmission (p. ex. le moustique qui transporte la maladie). Toutefois, l'impact des facteurs environnementaux et des interactions reste mal connu. Dans cet article, nous nous intéressons aux méthodes de découverte de connaissances permettant de mieux comprendre l'évolution des systèmes complexes pour la surveillance épidémiologique. Dans ce contexte, les méthodes de fouille de données spatio-temporelles apportent des solutions très pertinentes via l'identification sans hypothèse *a*