

SAFFIET : un système d'extraction de règles d'associations spatiales et fonctionnelles dans les séries de données géographiques

Asma Gharbi*, Cyril de Runz**
Sami Faiz***, Herman Akdag*

*Laboratoire d'informatique Avancé de Saint Denis EA 4383 2 rue de la Liberté
93526 SaintDenis Cedex France

akdag,gharbi,derunz@ai.univ-paris8.fr

**CRESTIC, IUT de Reims Châlons Charleville Chemin des Rouliers CS30012 51687
REIMS CEDEX 2 France

cyril.de-runz@univ-reims.fr

***ISAMM Campus Universitaire 2010 La Manouba Tunisie
sami.faiz@insat.rnu.tn

Résumé. Nous partons de l'hypothèse que les dynamiques spatiales et l'évolution des usages des objets géographiques peuvent en partie être explicitées (voire anticipées) par leurs différentes évolutions précédentes et les configurations spatiales dans lesquelles ils se situent. Aussi afin d'analyser et comprendre les changements de fonction des objets géographiques au cours du temps, et en déduire un modèle prospectif et puis prédictif, nous proposons l'outil SAFFIET qui exploite la recherche des motifs fréquents et des règles d'associations, pour extraire des règles d'évolution régissant les dynamiques spatiales.

1 Introduction

Un territoire évolue, se bâtit et se transforme au fil du temps. Aussi, nos sociétés ont besoin d'outils permettant d'estimer, voire de prédire, quelles seront les évolutions futures de nos territoires (urbains, rurales, *etc.*). Les modèles sous-jacents à ces outils peuvent aussi permettre de tester des hypothèses et envisager leur possible répercussion en terme d'aménagement.

Nous partons de l'hypothèse que l'évolution des fonctions des objets géographiques (usage des sols ou bâti) peut en partie être explicitée (voire anticipées) par l'historique des mutations précédentes et les configurations spatiales dans lesquelles elles se situent. Nous considérons des jeux de données constitués de séries temporelles de données géographiques. C'est dans ce contexte que nous proposons une approche en trois phases : une phase d'étude qui consiste à étudier les historiques des voisinages et des changements d'usage des objets géographique formant un territoire (trajectoires de vie) et les représenter sous forme de séquences ; une phase d'apprentissage, fouillant dans la base des séquences produites, pour en extraire des règles régissant l'évolution du territoire sur les niveaux fonctionnel (usage des objets) et spatial (to-

pologie de voisinage) ; et une dernière phase de prédiction appliquant ces règles sur la carte de la date de référence (représentant la donnée actuelle) pour produire sa carte dans le futur.

Fait à mentionner, ces phases ont déjà fait l'objet d'un développement (Gharbi et al., 2014) mais celui-ci s'est limité aux historiques des fonctions ou usages des objets géographiques pour extraire des règles d'évolution dites fonctionnelles (régissant les changements de fonction des objets), et les employer pour anticiper les futurs usages de ces objets et les visualiser sous format cartographique. Dans le présent article, nous proposons un système opérationnel, appelé SAFFIET (*Spatial And Functional Frequent Itemset Extraction Tool*), considérant aussi les configurations spatiales des objets (voisinage) lors de l'extraction des règles d'évolution. Correspondant aux deux premières phases, ce système, exploite les fonctionnalités des solutions existantes telles que PostGIS pour le stockage et la manipulation de l'information spatiale, Weka pour l'apprentissage, et Qgis pour la visualisation des résultats d'un prochain module de prédiction.

Dans cet article, nous présentons en section 2 les différentes caractéristiques techniques du système. La présentation du contexte d'utilisation (données utilisées et algorithme d'apprentissage) fait l'objet de la 3^{ème} section de l'article. Enfin, nous exposerons la conclusion de cet article ainsi que quelques perspectives.

2 Présentation de l'outil SAFFIET

2.1 Contexte : évolution d'un territoire (espace support)

Dans ce travail, nous percevons l'évolution des objets selon le paradigme identitaire (Cheylan et Lardon, 1993) qui se focalise sur l'entité géographique et considère qu'elle est composée des atomes qui représentent ses attributs sémantiques et spatiaux. Selon ce paradigme, la vie d'un objet géographique est reconstituée par la projection, sur l'axe temporel, de la suite des atomes appartenant à cet objet. En effet, dans les séries de données géographiques décrivant un même territoire (usage des objets) à des dates spécifiques (séries de cartes), chacune des cartes disponibles fournit une vue statique du territoire étudié car elle forme une (couche) indépendante. Dans ce contexte chaque objet dans ces cartes possède un identifiant qui change, d'une carte à l'autre, en cas d'une évolution (changement de fonction ou de forme) ce qui ne permet pas de maintenir le lien de composition en entre l'objet et ses atomes en d'autres termes, ne permet pas de suivre l'évolution de l'objet au fil de temps. Pour répondre à cette problématique nous avons adopté la solution de Kauppinen (2007) qui consiste à placer le traceur de l'identité sur l'un des attributs de l'objet. Par conséquent, nous avons proposé de considérer l'attribut de l'empreinte spatiale, c'est-à-dire identifier un objet par sa zone spatiale allouée et par la suite repérer ses successeurs à l'aide de quelques requêtes spatiales comme expliqué précédemment. Cette solution exige que le support spatial ne varie pas dans le temps.

2.2 Notre système

Le système proposé est une extension complexe à la version 2.8 du Système d'Information Géographique libre QuantumGIS (*QGIS*) développée en *Python* (v. 2.7.9). Cette extension (voir figure 1) fait collaborer les interpréteurs de ligne de commande *Bash* (v. 4.3.30), *SimpleCLI* (v. 3.6) pour *Weka* et *psql* pour *PostgreSQL* (v. 9.4) pour exécuter des commandes

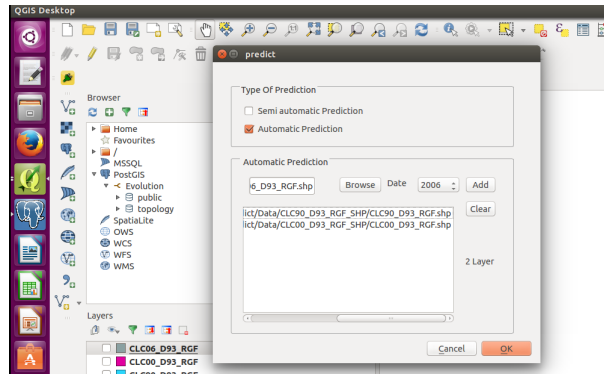


FIG. 1 – Interface graphique de SAFFIET

systèmes, configurer et appeler les algorithmes d'apprentissage et gérer le chargement des données géographiques dans la base de données. Le système de gestion de base de données utilisé est *Postgresql* (v. 9.4) muni de son extension spatiale *PostGIS*. La lecture des données se fait par une connexion à la base via le module *psycogp* qui adapte automatiquement les types *Python* aux types *PostgreSQL*.

Le pipeline de fonctionnement de SAFFIET, illustré dans la figure 2 est le suivant.

1. Le système prend en entrée un ensemble de cartes géographiques de type vectoriel (shapfile) décrivant le territoire (occupation du sol, bâti, etc.) à des dates différentes et antérieures à la date de référence (représentant le présent). Ces cartes sont converties et stockées, temporairement, sous la forme d'une table *PostGIS* dans *PostgreSQL* à l'aide de la commande *shp2pgsql*.
2. L'outil récupère, à l'aide des requêtes *SQL*, les différents objets présents dans les dites cartes, dans une table *geoentity*. Afin de définir le contexte spatial de chaque objet, nous recherchons d'abord les différents voisinages (ici adjacence) entre les objets, à l'aide de la requête *ST_Touches*, puis, en utilisant les requêtes *ST_Intersects*, *ST_Overlaps*, *ST_Covers* nous cherchons les continuités spatiales (couple d'objets dont l'intersection est non nulle) entre les objets stockés aux différentes dates. Par exemple, un objet ID2 de la carte 2000 dont l'intersection avec l'objet (ID1) de la carte 1990 est non nulle, représente le successeur de l'objet ID1 dans un changement possible d'usage. l'ensemble des couples extraits forme la table *Transitions*, à partir de laquelle les séquences (trajectoires) seront construites.
3. SAFFIET construit, à l'aide des requêtes *SQL*, la base d'apprentissage (fichier *Arff*) en correspondant chaque instance (séquence d'évolution) aux valeurs d'attributs (voisinage, évolutions précédentes, et successeur) qui décrivent l'évolution représentée par cette instance.
4. L'algorithme Apriori (Agrawal et Ramakrishnan, 1995) est exécuté, sous *Weka*, via une ligne de commande spécifiant que le précédent fichier *Arff* est la base d'apprentissage ainsi que les valeurs du support et de la confiance. Cet algorithme permet d'extraire

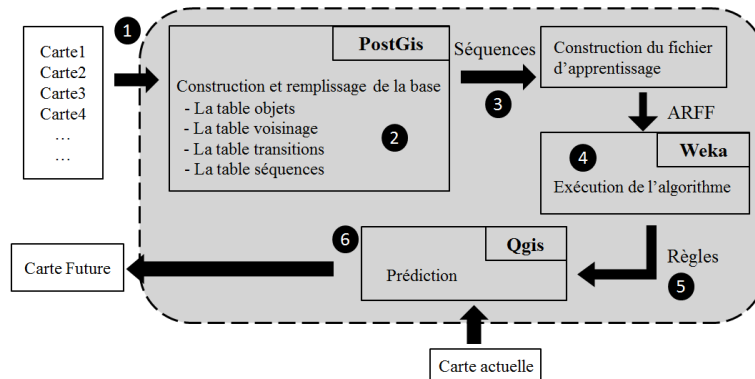


FIG. 2 – Schéma du fonctionnement de SAFFIET

des règles associant les différentes valeurs des attributs, qui ont fréquemment apparue ensemble dans la base d'apprentissage.

5. Les règles d'évolution seront, automatiquement, sélectionnées, parmi les règles produites, selon un ensemble de formes jugées porteuses d'informations utiles pour la prédiction et préalablement définies par l'utilisateur.
6. Une dernière étape sera, dans nos travaux futurs, dédiée à l'application des règles d'évolution trouvées sur la carte actuelle (ou de référence) du territoire étudié, pour estimer les éventuelles évolutions de ses objets constituants.

Notons que l'utilisation d'algorithmes plus récents qu'Apriori pourrait agir sur la qualité des règles produite et répondre aux problématiques de la scalabilité de notre système. Cependant, à ce stage, l'objectif porte plus sur la preuve de concept et de faisabilité d'un outil générique quelque soit le territoire étudié que sur le choix a proprement parler de la méthode d'extraction de règles. D'ailleurs, de par l'utilisation de Weka, nous pensons à proposer un ensemble de méthodes d'extraction de règles d'association à l'utilisateur.

La démonstration du prototype de SAFFIET proposée dans ce papier concerne les quatres premières étapes présentées ci-dessus. L'intervention de l'utilisateur dans cette démonstration consiste uniquement à charger les cartes géographiques à étudier, et à spécifier la date pour chacune.

SAFFIET représente un système d'extraction de règles d'évolution dont la partie centrale consiste à implémenter une méthode de suivi de l'évolution des objets géographiques.

3 Cadre de la démonstration

3.1 Cas d'application et données

Nous proposons d'utiliser comme cas d'application les données Corine Land Cover (CLC, Union européenne – SOeS, CORINE Land Cover, 2012) concernant le département 93 (Seine Saint-Denis) en France. CLC est une base de données d'occupation biophysique du sol sur le

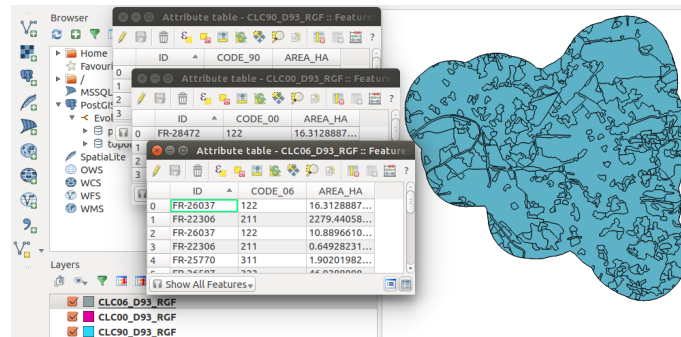


FIG. 3 – Les données d'étude Corine Land Cover

territoire européen. Diffusée par l'Agence européenne de l'environnement (AEE), les données CLC sont disponibles pour quatre dates différentes formant ainsi quatre cartes géographiques (1990, 2000, 2006 et 2012 – figure 3) au format vectoriel.

3.2 Extraction des règles d'association

Le résultat à ce stade de développement, est un ensemble de règles décrivant les associations qui peuvent y avoir entre les variables de l'évolution telle que nous la percevons : les fonctions des voisins d'un objet, son historique d'évolution et la fonction probable de son élément successeur.

La figure 4 présente quelques exemples de règles d'association produites. La première règle indique que les 29 instances (support de l'antécédent) qui possède l'attribut $\langle F=122; V=211 \rangle$ possèdent aussi l'attribut $\langle F=133; N=211 \rangle$. Cette règle concerne les 29 instances (support de la règle) et elle est donc vraie pour toutes les instances (confiance de 1). Comme on peut le constater dans la figure 4, les règles les plus « fortes » (ayant un grand support et une grande confiance) décrivent le plus souvent les associations de fonction entre objets voisins : par exemple, l'association $\langle F=133; N=211 \rangle$ où F indique que la fonction de l'objet 133 et N celle de son voisin, l'objet 211. Aussi, afin d'avoir une analyse spatiotemporelle plus fine, nous essaierons par la suite de rechercher des règles qui tiennent à la fois compte des associations entre voisins d'un objet (*i.e.* $\langle F=133; N=211 \rangle$, $\langle F=133; N=121 \rangle$) et de son historique d'évolution (*i.e.* $\langle SP : 133-122-111 \rangle$) pour indiquer la fonction probable de son élément successeur ($\langle S = 120 \rangle$). Ce genre de règle pourrait avoir la forme suivante :

$$\langle F=133; N=211 \rangle \langle SP : 133-122-111 \rangle \langle F=133; N=121 \rangle \implies \langle S = 120 \rangle$$

3.3 Conclusions et perspectives

Dans cet article, nous présentons un prototype d'un système d'extraction de règles d'évolution dont la partie centrale consiste à implémenter une méthode de suivi de l'évolution des objets géographiques à partir d'une série de cartes vectorielle décrivant un même territoire. En identifiant un objet par sa zone géographique allouée, SAFFIET définit, à l'aide de requêtes

SAFFIET : un outil d'extraction de règles spatiales et fonctionnelles

Best rules found:

1. $\langle F:122;N:211 \rangle = 1 \ 29 \implies \langle F:133;N:211 \rangle = 1 \ 29$ conf(1)
2. $\langle F:133;N:211 \rangle = 1 \ 29 \implies \langle F:122;N:211 \rangle = 1 \ 29$ conf(1)
3. $\langle F:211;N:211 \rangle = 1 \ 29 \implies \langle F:133;N:211 \rangle = 1 \ 29$ conf(1)
4. $\langle F:133;N:211 \rangle = 1 \ 29 \implies \langle F:211;N:211 \rangle = 1 \ 29$ conf(1)
5. $\langle F:124;N:211 \rangle = 1 \ 29 \implies \langle F:133;N:211 \rangle = 1 \ 29$ conf(1)
6. $\langle F:133;N:211 \rangle = 1 \ 29 \implies \langle F:124;N:211 \rangle = 1 \ 29$ conf(1)
7. $\langle F:121;N:211 \rangle = 1 \ 29 \implies \langle F:133;N:211 \rangle = 1 \ 29$ conf(1)
8. $\langle F:133;N:211 \rangle = 1 \ 29 \implies \langle F:121;N:211 \rangle = 1 \ 29$ conf(1)
9. $\langle F:112;N:211 \rangle = 1 \ 29 \implies \langle F:133;N:211 \rangle = 1 \ 29$ conf(1)
10. $\langle F:133;N:211 \rangle = 1 \ 29 \implies \langle F:112;N:211 \rangle = 1 \ 29$ conf(1)

FIG. 4 – Quelques exemples de règles d'association produites

PostGIS, un modèle relationnel permettant de caractériser les trajectoires des objets à l'aide de séquences. Ces séquences forment un fichier d'apprentissage duquel est extrait, en appliquant APriori, des règles d'évolution.

Afin d'améliorer SAFFIET, nous envisageons d'améliorer la pertinence des règles produites et de permettre le passage à l'échelle en agissant sur le choix de l'algorithme d'extraction de règles d'association, sur celui d'autres métriques de sélection (*e.g.* lift) et sur la définition du concept de voisinage.

Références

- Agrawal, R. et S. Ramakrishnan (1995). Mining sequential patterns. In *Proceedings of the Eleventh International Conference on Data Engineering, March 6-10, 1995, Taipei, Taiwan*, pp. 3–14.
- Cheylan, J.-P. et S. Lardon (1993). Towards a conceptual data model for the analysis of spatio-temporal processes : the example of the search for optimal grazing strategies. In A. Frank et I. Campari (Eds.), *Spatial Information Theory A Theoretical Basis for GIS*, Volume 716 of *Lecture Notes in Computer Science*, pp. 158–176. Springer Berlin Heidelberg.
- Gharbi, A., C. de Runz, S. Faiz, et H. Akdag (2014). An association rules based approach to predict semantic land use evolution in the french city of saint-denis. *International Journal of Data Warehousing and Mining* 10, 1–17.
- Kauppinen, T et Hyvönen, E. (2007). Modeling and reasoning about changes in ontology time series. In R. Kishore, R. Ramesh, et R. Sharman (Eds.), *Ontologies : A handbook of principles, concepts and applications in information systems*, Integrated Series in Information Systems, pp. 319–338. Springer-Verlag, New York (NY).

Summary

We start from the hypothesis that spatial dynamics and geographical object usage evolution may be in part caused by the different previous evolution and the spatial configuration in which they are situated. Therefore, in order to analyze and understand changes in the geographical object functions over time, and to build a prospective then predictif model, we introduce the SAFFIET system. It exploits data mining (through the search of frequent itemsets and association rules) for extracting evolution rules that are guiding spatial dynamics.