



Règles d'association pour l'analyse de l'évolution d'un territoire

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

► To cite this version:

Asma Gharbi, Cyril de Runz, Herman Akdag, Sami Faiz. Règles d'association pour l'analyse de l'évolution d'un territoire. Conférence Internationale Francophone Spatial Analysis and Geomatics (SAGEO), 2019, Clermont-Ferrand, France. , pp.331-336, 2019, SAGEO2019. <hal-02375525>

HAL Id: hal-02375525

<https://hal.science/hal-02375525v1>

Submitted on 22 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Règles d'association pour l'analyse de l'évolution d'un territoire

Asma GHARBI¹, Cyril DE RUNZ², Herman AKDAG¹, Sami FAIZ³

1. LIASD, Université de Paris 8, Saint-Denis, France

Herman{asma.haj-ali-gharbi}@univ-paris8.fr

2. LIFAT Université de Tours, IUT de Blois, Blois, France

cyril.derunz@univ-tours.fr

3. ISAMM, La Manouba, Tunisie

sami.faiz@isa2m.rnu.tn

RÉSUMÉ. Dans cet article, nous partons de l'hypothèse que les dynamiques spatiales et les usages des objets géographiques peuvent, en partie, être expliqués voire anticipés par leurs historiques de changements de fonctions et de co-localisations. Ainsi, nous proposons une démarche qui, se basant sur la recherche de règles d'association tout en considérant l'asymétrie des données, permet de générer des modèles explicatifs voire prédictifs de l'évolution d'un territoire. Appliquées sur des données issues de Corine Land Cover, les méthodes de fouille proposées ont fait leurs preuves dans la gestion du déséquilibre des données et les règles générées ont mis en évidence des dynamiques urbaines réalistes.

ABSTRACT. This work starts from the hypothesis that spatial dynamics and the change of geographical objects functions could be, partly, explained or anticipated by the history of their functions and co-locations changes. Hence, a process, based on association rules-based approaches and considering data asymetrie, for the extraction of an explicative/predictive model of territorial evolution is proposed. Applied on data extracted from Corine Land Cover, the proposed mining methods proved their worth in the management of data imbalance and the generated rules highlight realistic urban dynamics.

MOTS-CLÉS : Règles d'association, MSApriori, règles d'évolution, dynamique spatiale

KEYWORDS: Association rules, MSApriori, Evolution rules, spatial dynamics

1. Introduction

Un territoire évolue, se bâtit et se transforme au fil du temps. Ainsi, nos sociétés ont besoin de modèles permettant d'expliquer, voire d'anticiper, quelles seront les évolutions futures de nos territoires (urbains, rurales, etc.). Un large panel de travaux, se basant sur la fouille de données spatio-temporelles pour traiter la question des changements d'occupation/usage du sol (e.g. (Qiang, Lam, 2015)). La plupart de ces travaux visent essentiellement à identifier et à caractériser ces changements ainsi qu'à découvrir les relations qu'ils ont avec les différentes variables naturelles et anthropogéniques. Bien qu'efficace, cette démarche spécifique au type de couverture suivi (e.g. végétations, bâti), se concentre sur les caractéristiques internes des zones étudiées et néglige les relations spatiales et temporelles des données (i.e., une telle propriété est spatialement voisine à une telle propriété, une telle propriété apparaît avant une telle propriété sur une même localisation). Notre approche tente de répondre à ces problématiques en explorant les dépendances entre l'historique de l'évolution de l'objet spatio-temporel (relations temporelles de succession de fonctions), l'historique de ses co-localisations (relations spatiales de voisinage) et sa future fonction (future occupation/usage du sol). Pour cela, nous cherchons à extraire des règles d'association spatiotemporelles pour construire un possible modèle explicatif/descriptif de l'évolution d'un territoire. L'approche que nous proposons est constituée d'une première phase de suivi de l'évolution des objets pour la création de la base d'apprentissage, puis d'une seconde phase de recherche de règles d'association sur la base d'apprentissage créée. La base d'apprentissage générée étant dans les faits asymétrique, nous exploitons une méthode adaptée à cette problématique – MSAPriori (Liu *et al.*, 1999) – pour laquelle nous proposons d'utiliser pour valeurs de seuil des estimateurs statistiques de position.

2. Suivi de l'évolution et construction de la base d'apprentissage

L'objectif de cette première phase consiste à partir d'une série temporelle de cartes vectorielles décrivant un territoire à des dates données, pour définir un modèle de données permettant le suivi de l'évolution de celui-ci.

Dans ce contexte, nous considérons :

- L'espace comme étant un support immuable où se situent les objets et les relations exprimant leurs configurations spatiales.
- Le temps comme étant linéaire sous forme d'une succession d'évènements de changement de fonction et/ou de forme ou de topologie.
- L'objet géographique comme le produit de trois composantes à savoir, une dimension spatiale (S) (l'étendu spatial de l'objet), une dimension temporelle (T) (i.e., la date d'observation de l'objet), et une dimension sémantique (F) (la fonction de l'objet) (Perret *et al.*, 2015).

Tenant compte de ces éléments nous, considérons que l'évolution d'un objet est caractérisée par la modification de la valeur d'au moins une de ses composantes ou attributs. Ce changement donne, ainsi, lieu à la génération d'une nouvelle version de l'objet, portant une nouvelle identité. Ainsi, son évolution est décrite par ces différentes versions. Afin de garder le lien entre elles et de pouvoir en suivre l'évolution, comme proposé, dans (Kauppinen, Hyvönen, 2007), nous plaçons le traceur de l'identité sur l'un des attributs de l'objet : l'empreinte spatiale. Ainsi, l'objet est identifié par sa zone spatiale allouée et ses successeurs peuvent être repérés à l'aide des requêtes spatiales de recouvrement et de voisinage pour reconstruire sa trajectoire de vie (cf. figure1.a).

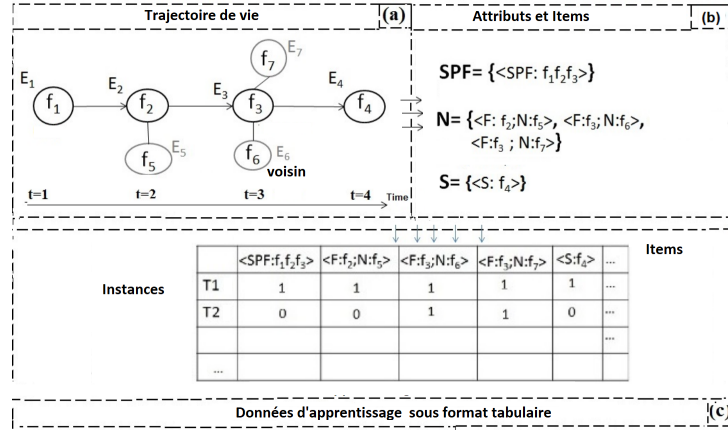


FIGURE 1. Exemple illustrant la génération de données d'apprentissage.

Dans le contexte du présent travail, nous stipulons que l'évolution d'un territoire peut être en partie explicitée voire anticipée par l'historique des fonctions de ses objets et de leurs configurations spatiales. À cet égard, notre trajectoire de vie (cf. figure1.a) peut être décrite sous forme d'une transaction composée de trois types d'information (cf. figure1.b). Le premier type (SPF) correspond à la succession de ses fonctions et dans notre exemple a pour valeur $\langle f_1, f_2, f_3 \rangle$. Le deuxième type (N) donne les valeurs des voisins au cours du temps. Par exemple, la relation de voisinage entre l'objet E_2 de fonction f_2 et l'objet E_5 de fonction f_5 , sera inscrite dans l'ensemble des voisinages comme $\langle F : f_2; N : f_5 \rangle$. Enfin, le dernier type (S) correspond à la fonction de l'objet successeur, dans l'exemple $\langle S : F_4 \rangle$. Nos différents attributs sont transcrits sous forme de table binaire indiquant leurs présence/absence dans la transaction (cf. figure1.c). Chaque attribut présent dans une transaction est appelé item. Faisant le choix que chaque transaction ait un seul attribut S et un seul SPF, une trajectoire de vie pourra donc être transcrite par une ou plusieurs transactions dans la base d'apprentissage suite au possible fusion/scission d'un

objet au cours du temps. Il convient aussi de noter que les voisinages sont identifiés selon la relation topologique «se toucher».

3. Extraction des règles d'association pour la modélisation de l'évolution territoriale

L'une des approches de modélisation, évoquée dans la littérature, notamment dans (Silva, Wu, 2012), est la modélisation à base de règles. L'un des challenges ici, c'est la définition des meilleures combinaisons de règles de transitions lorsqu'il y a beaucoup de variables à gérer. Dans ce contexte, plusieurs méthodes basées sur les algorithmes d'apprentissage ont été proposées dont l'analyse des associations qui permet de découvrir les relations entre différentes variables stockées dans une base de données (Solanki, Patel, 2015). Ces méthodes, notamment Apriori et MSApriori (Liu *et al.*, 1999), fonctionnent en deux étapes. Une première étape cherchant à regrouper les valeurs présentes (appelées items) dans des ensembles d'items, appelés itemsets. Seuls sont regroupés les items dont la fréquence d'apparition est supérieure à un seuil (minSup) ou à plusieurs seuils (nommés *SMI*), dans le cas de MSApriori, pour gérer des possibles déséquilibres dans les données. Ces items sont donc considérés comme fréquents, et les itemsets associés aussi. Sur ces itemsets fréquents, une approche de génération de règles d'association est effectuée afin de récupérer les possibles règles de co-occurrences.

Une première application sur nos données d'Apriori a révélé une asymétrie de données. En effet, seules des règles composées exclusivement par des items de type N sont générées (cf. tableau 1). Ceci s'explique par le fait que les items de voisinage sont très fréquents par rapport aux autres types d'items (S, SPF). Si on définit un minsup assez faible les items de type S et SPF apparaîtraient dans les candidats fréquents, toutefois, ceci mènera à un problème d'explosion combinatoire, i.e., un très grand nombre de candidats. Au travers ses multiples valeurs de support minimum, MSApriori rend possible d'extraire des règles comportant des éléments plus rares que si on avait un seuil global unique tout en réduisant la complexité combinatoire si un unique seuil bas serait défini. Cependant, dans cette méthode d'affectation de *SMI* l'expert doit définir la valeur des seuils en fonction d'une évaluation subjective de sa satisfaction à l'égard des modèles générés. Dans cet article, nous cherchons à minimiser la supervision du processus d'affectation des seuils par l'utilisation d'estimateurs de position dans un échantillon.

Méthodes à base de quartiles (QuartilesBased) : Partant de l'hypothèse qu'ils peuvent servir de frontières entre les valeurs "basses" et "élevées", nous nous intéressons aux estimateurs robustes de position que sont la médiane (Q2), les premier et troisième quartiles (Q1 et Q3). En utilisant ces valeurs, la population est divisée en quatre groupes de tailles équivalentes correspondant à quatre intervalles semi-ouverts. Pour chaque groupe, sa médiane est affectée comme le *SMI* de tous les items qui lui appartiennent. Nous appelons cette première

variante QuartilesBased.

Méthode à base de clustering (ClusterBased) : Afin d'affiner le partitionnement, nous pouvons nous servir d'algorithmes de clustering permettant de considérer les fréquences des items et les comparer mutuellement afin de regrouper les items par similarité de leurs fréquences. Ne souhaitant pas présumer du nombre de groupes de fréquences, nous avons opté, dans la variante ClusterBased, pour l'algorithme Espérance-Maximisation (EM) ne nécessitant pas ce paramétrage.

Vers la prise en compte de la sémantique des items

Pensant que la sémantique (S, SPF, N) des items doit pouvoir jouer un rôle dans la démarche, nous proposons une variante de QuartilesBased appelée QuartilesBasedSem ainsi qu'une variante de ClusterBased appelée ClusterBasedSem.

QuartilesBasedSem : Dans cette variante, nous partitionnons dans un premier temps les items selon leur sémantique afin d'obtenir 3 groupes (groupe d'item SPF, groupe d'item N, groupes d'item S). Pour chaque groupe, la valeur du quartile adéquat ($Q1$, $Q2$, $Q3$) est calculée puis attribuée en tant que *SMI* de tous les items de ce groupe. Pour les attributs les moins fréquents (S et SPF), la médiane ($Q2$) est adoptée comme *SMI*. Pour les attributs les plus fréquents (N), le quartile supérieur ($Q3$) est adopté.

ClusterBasedSem : Dans cette variante, le clustering se fait sur des données ayant pour attributs la fréquence des items et leur sémantique. La médiane de chaque groupe d'items lui sert de *SMI*.

4. Résultats

Nous avons appliqué nos propositions sur des données réels extraites de Corine Land Cover pour Paris. Quatre cartes vectorielles datant de 1990, 2000, 2006 et 2012 sont fournies. Les trois premières sont utilisées pour la génération des modèles et la quatrième pour l'évaluation de ceux-ci. La base de données d'apprentissage générée à partir de ces cartes vectorielles, compte 3913 instances, 190 items (97 N et 93 S et SPF). Dans le tableau 1, la proportion d'items SPF, N sur le nombre total d'items extraits montre que toutes les méthodes proposées ont réussi à faire apparaître les items S et SPF. Nous constatons que les méthodes à base de quartiles surpassent celles à base de clustering, et que le fait de prendre en compte la sémantique des éléments dans l'affectation des minsups améliore ces ratios (ex. 0,55 pour QuartilesBased contre 0,75 pour QuartilesBasedSem). Cependant, nous n'avons pu extraire de règles dans un temps raisonnable pour QuartilesBasedSem. Nous distinguons deux types de règles : les règles de stabilité (pas de changement de fonction entre les instants t et $t-1$); et les règles de transition où la fonction d'un objet change. Nous pourrions constater dans le tableau 1 que les trois méthodes pour lesquelles nous avons pu extraire des règles fournissent le même nombre de règles de transition mais pas le même nombre de règles de stabilité.

TABLE 1. Performances in terms of generation of S and SPF items

	$\frac{ {\{itemsSetSPF\}} }{ {\{items\}} }$	Règles Pertinentes générées	Règles de Stabilité	Règles de transition
US : Apriori (minsup=40%)	0	0	0	0
US : MSApriori ($\beta=0.5$)	0.086	0	0	0
MS:QuartilesBased	0.550	24816	2469	127
MS:QuartilesBasedSem	0.753	-	-	-
MS:ClusterBased	0.462	6264	6137	127
MS:ClusterBasedSem	0.5	9363	9236	127

5. Conclusion

Dans cet article, nous avons extrait des relations spatio-temporelles de voisinage et de succession d'usages des objets géographiques pour produire un modèle prospectif de l'évolution d'un territoire. Pour cela, à partir des différents snapshots, notre approche identifie des trajectoires de vies des objets, transcrivent celles-ci dans une base d'apprentissage binaire. Sur cette base d'apprentissage et afin de gérer l'asymétrie des données, nous avons appliqué MSApriori avec différentes stratégies d'affectation des minSup basées sur des estimateurs de position. Cela nous a permis d'extraire des règles d'évolution et de stabilité.

Bibliographie

- Kauppinen T., Hyvönen E. (2007). Modeling and reasoning about changes in ontology time series. In R. Sharman, R. Kishore, R. Ramesh (Eds.), *Ontologies: A handbook of principles, concepts and applications in information systems*, p. 319–338. Boston, MA, Springer US.
- Liu B., Hsu W., Ma Y. (1999). Mining association rules with multiple minimum supports. In *Proceedings of the fifth acm sigkdd international conference on knowledge discovery and data mining*, p. 337–341.
- Perret J., De Runz C., Rodier X., Varet-Vitu A., Dumenieu B., Saligny L. *et al.* (2015). Études des dynamiques de l'occupation du sol : questionnement, simplification et limites. *Revue Internationale de Géomatique*, vol. 25, n° 3, p. 301–330.
- Qiang Y., Lam N. S. N. (2015). Modeling land use and land cover changes in a vulnerable coastal region using artificial neural networks and cellular automata. *Environmental Monitoring and Assessment*, vol. 187, n° 3, p. 57.
- Silva E., Wu N. (2012). Surveying models in urban land studies. *CPL bibliography*, vol. 27, n° 2, p. 139–152.
- Solanki S. K., Patel J. T. (2015). A survey on association rule mining. In *2015 fifth international conference on advanced computing & communication technologies*, p. 212–216.