

CONTINUUM : UN MODÈLE SÉMANTIQUE SPATIO-TEMPOREL POUR L'ANALYSE DE PHÉNOMÈNES DYNAMIQUES

par Benjamin Harbelot et Christophe Cruz

Laboratoire Le2i, UMR CNRS 6306,

Université de Bourgogne, 21000 Dijon

benjamin.harbelot@check.sem.fr, christophe.cruz@u-bourgogne.fr

Les gestionnaires de territoires souhaitent avoir un aperçu des connaissances en cours et de l'évolution de certaines caractéristiques du territoire. Pour ce faire, les outils de télédétection enregistrent une grande quantité d'informations liées à la couverture territoriale permettant l'étude de processus dynamiques. La littérature du domaine de la modélisation spatio-temporelle est vaste et a donné lieu à de nombreux modèles. Chacun révèle une certaine aptitude à capturer l'évolution des différentes caractéristiques d'un environnement géospatial. Toutefois, les données nécessitent de nouveaux outils pour identifier des motifs et extraire de la connaissance. Nous proposons un modèle capable de découvrir la connaissance sur des données parcellaires et permettant l'analyse des phénomènes dynamiques à l'aide de données temporelles, spatiales et thématiques. Le modèle est appelé Continuum et se base sur les technologies du Web sémantique pour permettre une représentation accrue du contexte de l'environnement géospatial et fournir des résultats d'analyses proches de ceux des experts du domaine via des opérations de raisonnement automatique. En définitive, ce modèle permet d'améliorer notre compréhension de la dynamique des territoires.

Introduction

La dynamique d'un environnement se caractérise par la succession de deux éléments : les états et les transitions. Un état fait référence à la représentation d'une entité à un instant T tandis qu'une transition décrit le processus induisant le passage d'un état vers un autre état. Dans la littérature des ontologies, les états peuvent être assimilés à des objets dits endurants, également connus sous le nom de continnants, et les transitions peuvent être assimilées à des objets dits perdurants, aussi connus sous le nom d'occurents. Les endurants sont définis comme des entités pouvant être perçues comme des concepts complets, quel que soit l'instant de temps. Par exemple, une zone urbaine, une voiture, un humain sont des endurants. Les perdurants sont des entités pour lesquelles seule une partie existe quel que soit l'instant de temps d'observation. Les perdurants sont souvent assimilés à des processus tels que l'expansion urbaine, ou plus simplement le fait de se déplacer. Dès lors, il est intéressant de noter que ce qui est décrit lors de la définition d'un objet perdurant est toujours décrit en référence à un endurant. Et chaque perdurant est similaire ou différent d'un autre en fonction du continuant concerné.

Afin de répondre à des problématiques non encore explorées des SIG spatio-temporels, le modèle Continuum a été proposé et implémenté pour répondre aux limites des solutions existantes. Celui-ci possède la particularité d'être basé sur une ontologie formelle

définie à l'aide des technologies du Web sémantique. De ce fait, la structure du modèle est, par nature, un graphe. Les graphes ont été identifiés comme des structures particulièrement adaptées aux besoins de la modélisation. En effet, l'étude de la dynamique d'un environnement nécessite de s'intéresser aux diverses interactions spatiales, temporelles et thématiques qu'entretiennent entre elles les entités d'un territoire. La grande majorité des modèles spatio-temporels fournis pour la gestion d'entités géographiques spatio-temporelles sont implémentés dans des bases de données de type relationnel, relationnel étendu, objet, objet-relationnel. Tous les modèles étudiés utilisent des supports de stockage dont l'expressivité est limitée. Toutefois, les modèles les plus récents et les plus aboutis ont opté pour une modélisation sous forme de graphe révélant ainsi la nécessité d'étudier les relations qu'entretiennent les entités au sein de l'environnement géospatial dans le cadre de la modélisation spatio-temporelle [Harbelot, 2015].

Les approches par graphe semblent s'être imposées comme des supports adaptés à la modélisation spatio-temporelle et permettent de représenter la dynamique d'entités spatio-temporelles. Cependant, les motifs et processus permettant de décrire la dynamique ne s'intéressent qu'à la dynamique spatiale et ne peuvent représenter explicitement des connaissances liées aux entités. C'est pourquoi il est nécessaire de fournir des outils de raisonnement afin de permettre de donner

une interprétation des motifs spatiaux plus proche d'un phénomène du monde réel en renforçant les capacités sémantiques du modèle. Les SIG de nouvelle génération requièrent une modélisation et une analyse toujours plus proches de la réalité afin de rapprocher l'humain et la machine. En effet, afin de permettre d'assister l'humain dans sa gestion de l'information, il est intéressant de permettre à la machine d'accéder à la compréhension de l'humain sur un domaine particulier. Dès lors, le manque d'expressivité des supports actuels est une forte limitation aux futurs défis de la modélisation spatio-temporelle. Dans ce contexte, il est fondamental d'envisager le développement de systèmes capables de prendre en compte des structures sémantiques de plus haut niveau ainsi que de fournir des capacités d'analyses qualitatives pour l'étude de dynamiques géospatiales. La principale limite des modèles proposés dans la littérature concerne le manque de capacité à représenter explicitement la connaissance liée à l'environnement géospatial ainsi que sa dynamique. Les technologies du Web sémantique, et plus particulièrement leur application au sein des ontologies, comblent cette limite. Toutefois, la littérature du Web sémantique ne propose pas d'outils pour la modélisation spatio-temporelle. Dès lors, le premier verrou consiste à proposer un modèle de représentation de connaissances spatio-temporelles à l'aide des technologies du Web sémantique.

Le modèle « Continuum »

Le Web sémantique et les technologies associées proposent des outils puissants permettant la représentation des connaissances ainsi que des mécanismes de raisonnement. En outre, les technologies du Web sémantique dissocient le schéma et les données du système d'information offrant ainsi une structure plus flexible et plus facile à appréhender pour un humain. Enfin, ces outils sont soumis à l'hypothèse du monde ouvert. L'hypothèse du monde ouvert s'oppose à l'hypothèse du monde fermé, valable dans les bases de données traditionnelles, en autorisant la représentation de connaissances incomplètes. Sous l'hypothèse du monde fermé, toute connaissance non explicitement déduite est considérée comme fautive, alors que dans le monde ouvert ce qui est omis reste inconnu. Le manque de connaissance est une caractéristique récurrente du domaine spatio-temporel, c'est pourquoi l'hypothèse du monde ouvert semble la plus adaptée. Toutefois, la représentation de données incomplètes peut également nuire à l'exploitation des données. En effet, l'hypothèse du monde ouvert n'impose pas que toutes les propriétés d'une instance d'un concept soient représentées. Par exemple, dans une application spatio-temporelle, les parcelles d'un territoire possèdent un attribut ou une propriété permettant de stocker et d'associer

une géométrie. Dans le cadre du monde ouvert, il est possible de représenter des parcelles sans géométrie. Dès lors, toute analyse spatiale réalisée sur les données fournira des résultats partiels difficilement exploitables. Ainsi, bien que le monde ouvert montre des aptitudes évidentes pour la découverte de connaissances implicites, il est, en revanche, nécessaire d'établir des limites sur l'incomplétude des données pour définir les nécessités techniques du modèle afin de garantir la consistance du système.

Un « Continuum » [Harbelot et al., 2015b] décrit un ensemble d'éléments homogènes tels que l'on puisse passer de l'un à l'autre de façon continue. Qu'il s'agisse d'un changement ou d'un mouvement, une évolution appliquée à un objet enclenche un processus spatio-temporel spécifique à cet objet. La notion d'évolution pose alors la question de l'identité d'un objet. Autrement dit, jusqu'à quel point un objet peut-il changer avant d'être considéré comme un objet différent ? Par conséquent, à partir de quel instant considère-t-on que le processus spatio-temporel de l'objet en question doit s'arrêter et qu'un nouveau processus doit être enclenché sur le nouvel objet ? Le continuum permet de fixer cette limite. Le continuum délimite le début et la fin d'un processus spatio-temporel. Dans [Claramunt et al., 1996], les auteurs soulignent que les objets dynamiques doivent distinguer les aspects thématiques, spatiaux et temporels. Afin de représenter une entité dynamique dans le modèle Continuum, nous étendons cette proposition en créant un ensemble de tranches de temps d'objets (ou représentation d'une entité appelée *timeslice*), chacun constitué de quatre composantes :

- **Sémantique** : elle décrit les connaissances associées à l'entité.
- **Spatiale** : elle est la représentation spatiale de l'objet.
- **Temporelle** : elle représente les instants ou intervalles de temps qui décrivent la durée d'existence.
- **Identité** : elle se définit comme le caractère unique d'un objet, indépendamment de ses attributs ou de ses valeurs.

L'identité est un point central de l'évolution spatio-temporelle et doit être prise en compte dans le modèle [Hornsby et al., 2000]. Par exemple, une parcelle de terre peut passer de l'état de forêt à l'état de zone urbaine. Dans cet exemple, la géométrie n'a pas changé, cependant on constate un changement sémantique (fig. 1). Il est également possible que la sémantique puisse ne pas changer tandis que la géométrie évolue. Par exemple, une parcelle de terrain à l'état de zone urbaine

pourrait s'étendre par l'achat des parcelles voisines. Nous parlerons alors de relation de dérivation ou de continuation.

La célèbre approche 4D-fluent [Welty et al., 2006], traditionnellement utilisée pour faire évoluer des objets, permet de représenter des entités à différents instants de temps. Cependant, chacune des représentations d'une entité n'est pas liée ni à son prédécesseur ni à son successeur, par conséquent, dans le cas d'un changement d'identité de l'entité entre deux représentations, il est alors impossible de retrouver et comprendre le lien de succession entre ces deux représentations : l'approche 4D-fluent ne permet pas de gérer le suivi d'objet dans le temps ni le changement d'identité. Le but du modèle Continuum est notamment de suivre l'évolution des entités à travers le temps et d'explicitier le phénomène spatio-temporel révélé par ces évolutions. Pour atteindre cet objectif, le modèle établit une relation de filiation entre les différentes représentations de chaque entité et tente d'améliorer la connaissance portée par cette relation.

De la filiation à la qualification d'un phénomène lors du suivi d'un objet dans le temps

Les phénomènes du monde réel à modéliser sont soumis à des contraintes. Plus un phénomène présente de contraintes devant être représentées ou détectées, plus ce phénomène se distingue des autres pour révéler sa nature spécifique. Inversement, la réduction du nombre de contraintes permet de regrouper des phénomènes selon des caractéristiques communes. S'il est évident qu'extraire la nature unique d'un phénomène s'avère l'option la plus intéressante, cette analyse est généralement limitée par le nombre de contraintes exploitables à disposition dans le système d'information. Toutefois, une analyse lancée sur la base de nombreuses contraintes se heurte généralement à des temps de calcul relativement longs. En réponse aux problématiques exposées précédemment, notre approche se base sur un nombre de contraintes suffisamment générique pour regrouper un nombre conséquent de résultats, et suffisamment spécifique pour servir de support à des analyses plus poussées. Le résultat de l'analyse est stocké sous la forme de relations organisées en couche. En somme, notre approche agit comme un prétraitement des données permettant d'accélérer les temps de calcul d'analyse de plus haut niveau ainsi que de fournir de nouveaux supports d'analyse exploitables. Dans notre modèle, nous proposons de spécialiser la relation de filiation au sein d'une hiérarchie de subsomption. La hiérarchie se compose de trois couches génériques :

- La couche filiation : seule la relation de filiation est établie afin de suivre le cycle de vie d'une entité au travers de ses évolutions. Aucune connaissance spécifique n'est portée par cette relation.
- La couche identité : la relation de filiation est spécialisée soit en une relation de dérivation si l'entité change de nature, soit en une relation de continuation si cette entité conserve sa nature au travers de l'évolution représentée.
- La couche topologie et identité : cette couche regroupe des contraintes spatiales ainsi que des contraintes d'identité pour fournir huit motifs d'évolution d'une entité spatiale.

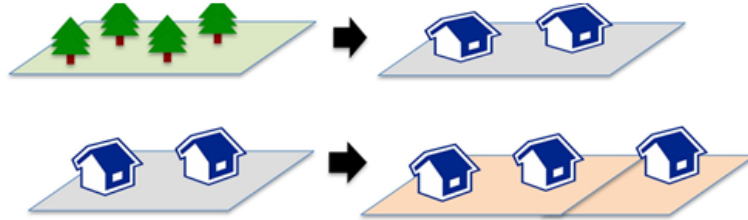
Une quatrième couche a été définie (fig. 2), mais elle ne peut pas être définie de manière générique puisqu'elle dépend des informations stockées dans l'ontologie de domaine. Cette dernière couche est la plus spécifique et la plus expressive. Elle révèle des phénomènes spatio-temporels en se basant simultanément sur les motifs définis dans la 3^e couche ainsi que sur les informations stockées dans l'ontologie qui donne un sens à ces motifs. Dès lors, les motifs trouvent une correspondance dans le monde réel et révèlent un phénomène particulier permettant d'accroître la connaissance et la compréhension de l'environnement géospatial.

Conclusion

Le modèle s'intéresse essentiellement à la relation de filiation et définit quatre couches permettant chacune d'offrir une vision de la dynamique d'une entité. Dans ces travaux, une méthodologie a été proposée afin de distinguer les relations de filiations spatiales des relations de filiation valides. Les filiations spatiales font référence aux relations établies entre des parcelles dont les géométries se chevauchent dans le temps, ces relations peuvent être valides s'il est admis qu'il existe une filiation entre les parcelles dans le monde réel, ou bien non valides, dans le cas contraire. Toutefois, la distinction entre les deux types d'entités peut parfois s'avérer arbitraire et se rapporte au problème de l'incertitude. Dans le cadre de ces travaux, un seuil est défini afin d'établir clairement la limite entre ces deux relations et sert de support au calcul des différentes couches du modèle.

Afin de détecter des phénomènes, au plus proche de la réalité modélisée, le modèle Continuum s'appuie sur une hiérarchie de concepts permettant de regrouper les entités spatio-temporelles au sein de classes plus

Changement d'identité = Dérivation



Pas de changement d'identité = Continuation

Figure 1 : L'évolution et l'identité

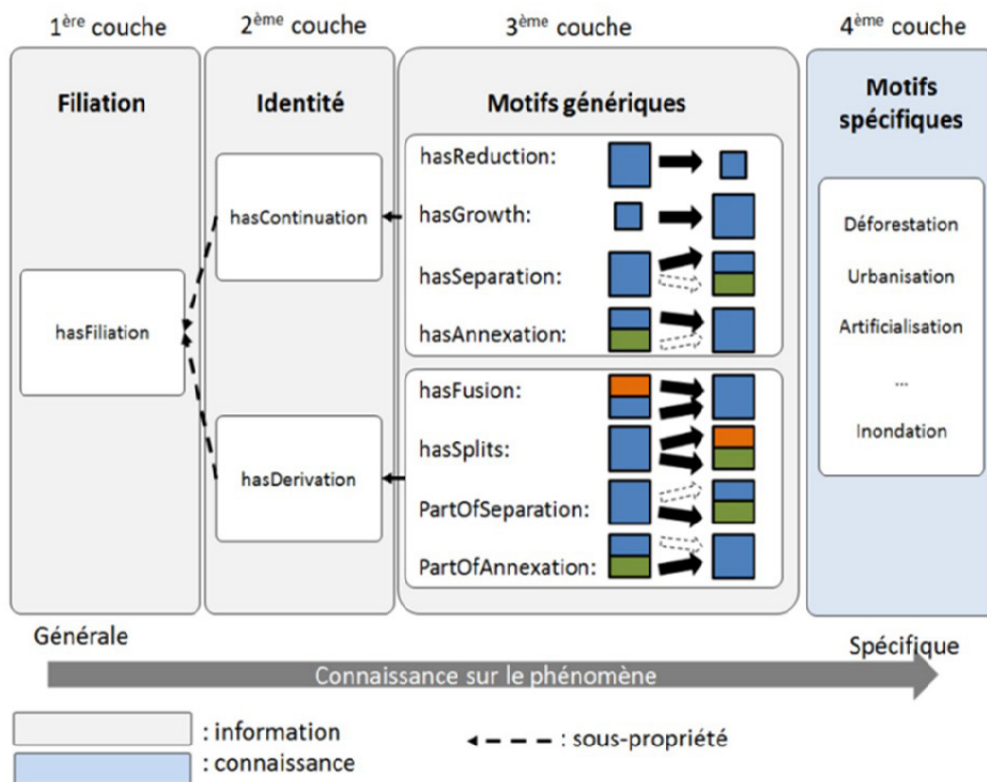


Figure 2 : Les couches de connaissances liées à la relation de filiation

ou moins génériques. Dans ces travaux, la hiérarchie est exploitée afin de permettre d'étudier la profondeur du changement. Dès lors, un changement impliquant l'identité d'une entité et associé à la relation de dérivation dans le modèle est vu comme un changement depuis une classe parent de la hiérarchie vers une classe enfant de la hiérarchie. Le modèle Continuum permet de hiérarchiser ce type de changement en étudiant le niveau maximum de la hiérarchie sur lequel il est possible de visualiser le changement. Ainsi, plus le changement intervient sur des concepts génériques et plus celui-ci est considéré comme fort, car plus des concepts sont génériques et plus leur sémantique est éloignée. A l'inverse, un changement intervenant entre des concepts très spécifiques suggère une modification faible de la couverture terrestre.

De plus, l'évolution d'une entité, qu'elle soit spontanée ou inscrite dans la durée, peut également être influencée par l'évolution des entités voisines. Le territoire modélisé peut être vu comme un espace fixe et délimité, composé d'un ensemble de parcelles évoluant au cours du temps. Dès lors, on comprend que l'évolution spatiale d'une entité implique une évolution conjointe des parcelles environnantes. L'ajout de sémantique, ainsi que les outils de fouille de graphes proposés par le modèle Continuum, permettent d'assister les experts dans la découverte des phénomènes connexes et complexes au sein du territoire.

Bibliographie

Claramunt, C., et Theriault, M. 1996. *Toward semantics for modelling spatio-temporal processes within gis*. Advances in GIS Research I, pages 27–43.

Del Mondo, G., Rodriguez, M., Claramunt, C., Bravo, L., et Thibaud, R. (2013). *Modeling consistency of spatio-temporal graphs*. Data & Knowledge Engineering, 84 :59–80.

Harbelot, B. et al. 2015. *LC3 : A spatio-temporal and semantic model for a knowledge discovery from geospatial datasets*. Journal of Web Semantics, Special Issue on Geospatial Semantics, Elsevier, 2015.

Harbelot, B 2015. *Continuum : un modèle spatio-temporel et sémantique pour la découverte de phénomènes dynamiques au sein d'environnements géospatiaux*, thèse de doctorat soutenue à l'Université de Bourgogne le 17 décembre 2015.

Hornsby, K., et Egenhofer, M. J. 2000. *Identity-based change : a foundation for spatio-temporal knowledge representation*. International Journal of Geographical Information Science, 14(3), 207–224.

Welty, C., Fikes, R., et Makarios, S. 2006. *A reusable ontology for fluents in owl*. Dans FOIS, volume 150, pages 226–236.