

Navigation et appariement d'objets géographiques dans une ontologie

Rémy Brisson*, Omar Boussaïd*
Pierre Gançarski**, Anne Puissant***, Nicolas Durand**

*ERIC EA 3083, Université Lumière Lyon 2, 5 avenue Pierre Mendès-France,
F-69676 Bron Cedex

remybrisson@yahoo.fr, omar.boussaid@univ-lyon2.fr

**LSIIT-AFD UMR 7005 CNRS, Université Louis Pasteur, Bd Sébastien Brant,
F-67400 Illkirch

pierre.gancarski@lsiit.u-strasbg.fr, durand@lsiit.u-strasbg.fr

***IDEES-GEOSYSCOM UMR 2795 CNRS, Université de Caen, Esplanade de la paix,
F-14032 Caen Cedex
anne.puissant@unicaen.fr

Résumé. L'ACI FoDoMuSt se propose d'élaborer un processus de fouille de données multi-stratégies pour la reconnaissance automatique d'objets géographiques sur des images satellitaires ou aériennes. Ces dernières sont segmentées afin d'isoler des polygones définis par un ensemble de descripteurs de bas niveaux. Afin de leur affecter une sémantique, on applique dans un premier temps une classification. Si aucun objet géographique n'est identifié, on tente alors un appariement du polygone avec les concepts d'une ontologie d'objets géographiques. Un algorithme de navigation dans l'ontologie et une mesure de comparaison sémantique ont ainsi été développés, paramétrables selon le contexte d'appariement. Cette mesure évalue la pertinence d'un appariement et comprend une composante locale (comparaison au niveau du concept) et une composante globale (combinaison linéaire de mesures locales). La méthode proposée a été développée en JAVA et intégrée à la plate-forme FoDoMuSt. Les premières expérimentations et évaluations humaines sont très encourageantes.

1 Introduction

Dans le cadre de l'ACI Masse de Données, le projet FoDoMuSt¹ élabore un processus de Fouille de Données Multi-Stratégies pour l'extraction et la qualification de la végétation urbaine à partir de bases de données d'images de télédétection. A cette fin, ce projet tire ses ressources notamment de sa banque de données images, d'une base de données géographiques et d'une ontologie d'objets géographiques (voir figure 1). Dans un premier temps, les images aériennes ou satellitaires sont segmentées par des procédés de fouilles supervisées et non supervisées. Des polygones sont ainsi définis et caractérisés par un ensemble de descripteurs de

¹<http://lsiit.u-strasbg.fr/afd/sites/fodomust/fr-accueil.php>

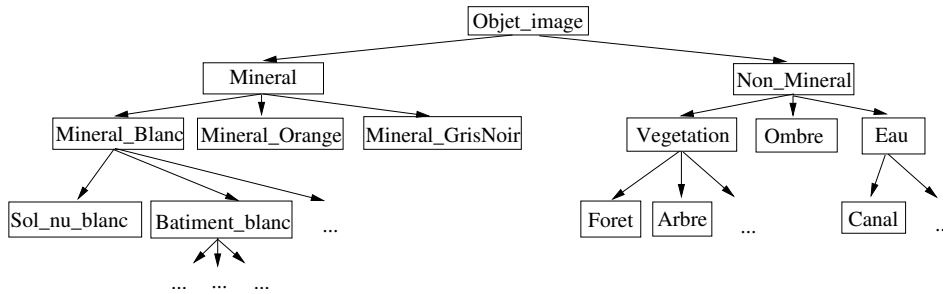


FIG. 1 – Extrait de l'ontologie d'objets géographiques.

bas niveau. On tente alors d'identifier et valider ces polygones en tant qu'objets géographiques par une classification. Lorsque cette dernière échoue, les concepts de l'ontologie d'objets géographiques sont étudiés pour affecter une sémantique aux polygones. Cette étape nécessite un processus de navigation efficace dans l'ontologie, ainsi que l'élaboration d'une métrique adaptée à la comparaison des polygones aux concepts. L'objectif de cet article est donc de proposer un mécanisme de navigation, de comparaison et d'appariement de polygones avec des concepts d'une ontologie issue du domaine de la géographie et dédiée à l'analyse d'une banque de données images (aériennes et/ou satellites).

Dans la première section, nous abordons les approches existantes au niveau de la comparaison sémantique et des mesures de similarité. La deuxième section expose la méthode de navigation et de comparaison mise en place pour répondre à notre objectif. Ensuite, nous présentons une utilisation de notre approche avec la plate-forme FoDoMuSt, avant de conclure et évoquer des perspectives à notre travail.

2 État de l'art

L'identification d'un polygone en utilisant l'ontologie implique une comparaison des propriétés du polygone avec celles des concepts de l'ontologie. Les propriétés (attributs et contraintes sur les attributs) d'un concept sont des conditions individuellement nécessaires et collectivement suffisantes pour établir la relation d'appartenance. Autrement dit, un objet doit toutes les satisfaire pour être membre de la catégorie (dans notre cas un concept) et tout objet les satisfaisant en est un membre. En terme fonctionnel, une catégorie peut être définie par un prédicat d'appartenance sous la forme d'une fonction qui retourne une valeur booléenne selon l'appartenance ou non de l'objet à la catégorie (Mariño Drews, 1993). Ce raisonnement s'appuie néanmoins sur deux hypothèses fortes. La première concerne les propriétés utilisées qui doivent être suffisamment discriminantes. La deuxième est l'hypothèse du monde clos qui présuppose que tout objet peut être parfaitement représenté dans un mode donné de représentation et qu'il est possible de déterminer s'il appartient ou non à une catégorie donnée. Les différentes approches présentes dans la littérature articulent la comparaison sémantique entre une stratégie et une fonction de comparaison (ou mesure de similarité).

D'une manière générale, les stratégies de comparaison admettent, explicitement ou non, une distinction entre structure *locale* (interne, intrinsèque) et *globale* (externe, extrinsèque) des concepts. La première correspond aux propriétés définies au niveau du concept lui-même, la deuxième correspond à la place du concept dans la structure globale et aux relations auxquelles il participe (héritage, relation de composition, etc.). Pour plus d'informations sur les approches existantes, nous reportons le lecteur à (Schvaiko et Euzenat, 2005).

En ce qui concerne l'évaluation de la similarité, le passage à une représentation "attribut-valeur" permet le calcul de distances euclidiennes dans un espace multidimensionnel. Schwering et Raubal (2005) raisonnent sur des régions géospatiales conceptuelles représentées par des vecteurs à valeurs numériques. Selon l'approche de Tversky (1977), la comparaison de deux ensembles (par exemple les attributs du polygone et ceux d'un concept) ne dépend que de leurs ressemblances et de leurs différences. Il définit ainsi le *contrast model* et le *ratio model*. Rodriguez et Egenhofer (2003) ont étendu ce dernier en proposant la *Matching Distance Similarity Measure* (MDSM) qui est une mesure asymétrique et contextuelle. Brodeur (2004) adopte une approche similaire dans le cadre d'une proximité géosémantique.

Ces approches abordent donc simultanément les structures internes et externes des concepts. Le souci de définir des mesures asymétriques où la relation de généralisation/spécialisation influence sur la mesure, traduit l'idée d'appréhender les concepts dans toute leur complexité. Si de nombreuses mesures évaluent la similarité selon la profondeur des concepts, le plus court chemin entre eux, etc., certaines utilisent ces caractéristiques (explicitement ou non) comme facteur d'asymétrie ou de spécialisation.

3 Méthode d'appariement proposée

Un objet géographique potentiel est dans un premier temps un polygone sur lequel un certain nombre de caractéristiques (dimensions, indices divers tels que l'indice de Miller, etc.) sont calculées dans une phase de vectorisation. Un polygone est alors donné sous la forme d'attributs-valeurs. Notre approche d'appariement d'un polygone aux concepts de l'ontologie, est orientée attributs ("feature-based"). Elle consiste à vérifier la validité des valeurs selon les propriétés et les contraintes définies au niveau des concepts. Cependant, un polygone n'ayant pas de structure sémantique, nous ne pouvons pas nous baser directement sur les formules évoquées en section 2, comme par exemple MDSM. Un polygone peut être a priori apparié à n'importe lequel des concepts. Les attributs d'un polygone permettant l'appariement à un concept ne sont donc pas identiques selon le concept étudié. Par exemple, le concept "Bâtiment" est défini par de nombreux indices (élongation, Miller, ...) et des informations radiométriques, alors que le concept "Ombre" est défini uniquement à partir d'attributs radiométriques. Sans connaissance a priori, il est alors nécessaire pour chaque polygone, de calculer tous les attributs possibles, même si la majorité s'avéreront inutiles pour le traitement d'un concept.

Afin de tenir compte de toutes ces spécificités nous avons élaboré notre propre méthode d'appariement. Nous inspirant de la littérature pour affecter une sémantique aux polygones, nous avons placé au cœur du mécanisme d'appariement une mesure différenciant les composantes locales et globales des concepts. Nous avons défini une nouvelle mesure de similarité (locale) et un score d'appariement (global) pour évaluer la pertinence dans une hiérarchie de concepts.

3.1 Score d'appariement

La mesure de similarité (locale) compare les attributs d'un polygone P avec les attributs spécifiques à un concept C et est définie par la formule suivante :

$$Sim(P, C) = \frac{\sum_{i=1}^n \alpha_i Valid(A_i, v_i)}{\sum_{i=1}^n \alpha_i}$$

$$Valid(A_i, v_i) = 1 \text{ si } v_i \text{ satisfait les contraintes sur } A_i, 0 \text{ sinon.}$$

où v_i sont les valeurs dans P pour les attributs A_i , et α_i les pondérations rattachées aux A_i du concept C (traduisant des rôles plus ou moins discriminants). Notons que la mesure locale ne compare que les attributs communs entre ceux de P et ceux qui sont spécifiques à C (i.e. attributs surchargés ou définis dans ce concept).

Le score d'appariement (global) évalue la pertinence de l'appariement dans la hiérarchie de concepts. Il est défini par l'équation ci-dessous, où les C_j sont les concepts constituant le chemin de la racine à C_m .

$$S(P, C_m) = \frac{\sum_{j=1}^m \beta_j Sim(P, C_j)}{\sum_{j=1}^m \beta_j}$$

Ce score d'appariement est une combinaison linéaire des similarités locales obtenues avec les C_j . Les similarités locales sont "propagées" par héritage aux concepts plus spécifiques. Nous intégrons dans ce calcul, un facteur de spécialisation β basé sur la profondeur des concepts. Nous privilégions ainsi a priori la spécialisation, en considérant que toute nouvelle information apporte une nouvelle sémantique.

Notons qu'un polygone n'est pas comparé aux différents concepts avec les mêmes propriétés (car les concepts sont décrits différemment). On ne peut donc pas assurer que les valeurs respectent les caractéristiques des distances (symétrie, séparation et inégalité triangulaire). Il semble alors logique que des mesures spécifiques aux différents concepts puissent être intégrées par la suite au module d'appariement. On pourrait ainsi traiter spécifiquement les relations entre concepts (autre que l'héritage) dans une approche plus spécialement adaptée aux relations de composition par exemple.

3.2 Navigation dans l'ontologie

Le score entre un polygone et un concept étant défini, il faut à présent parcourir l'ontologie pour déterminer le ou les concepts les plus proches. Pour cela nous opérons un parcours en largeur en utilisant deux heuristiques pour diminuer l'espace de recherche et ainsi accélérer les traitements. Un seuil minimum pour le score d'appariement permet d'élaguer les branches dont le concept de départ ne donnerait pas un score suffisant. Cette stratégie est basée sur le fait qu'un concept ayant peu de propriétés vérifiées par le polygone ne sera pas pertinent au niveau des concepts spécialisés. Remarquons que dans le cas où la similarité locale ne serait pas calculable (aucun attribut à comparer), alors le seuil n'est pas pris en compte et le parcours se poursuit. La deuxième possibilité offerte à l'utilisateur consiste à fixer une profondeur d'exploration. Cela permet ainsi de choisir le degré de détail que l'on souhaite utiliser. Par exemple, sur la figure 1, si on fixe la profondeur à 3, seules les catégories usuelles seront visitées (végétation, eau, etc.). La navigation est bien sûr paramétrable de façon à utiliser ou non ces stratégies.

3.3 Implantation

L'ontologie de FoDoMuSt a été créée avec Protégé2000². Nous avons développé le module d'appariement sous JAVA en utilisant l'API de développement Protégé2000 qui permet d'aborder la base de connaissances sous forme de frames. Ces dernières sont liées au paradigme objet et sont particulièrement adaptées au traitement informatique et à la manipulation des concepts et de leurs relations. Nous travaillons avec notre propre mécanisme de raisonnement, sans passer par un raisonneur qui nécessiterait un formatage de la base de connaissance.

L'accent a été mis sur une implantation la plus générique possible, aboutissant à un environnement de développement pour l'appariement. Le module a été ensuite intégré à la plate-forme FoDoMuSt.

4 Utilisation dans la plate-forme FoDoMuSt

La plate-forme FoDoMuSt permet de manipuler et d'effectuer des traitements sur des images aériennes ou satellitaires. Il est entre autre possible d'effectuer une classification et/ou une segmentation de l'image afin de construire des polygones en fonction des informations radiométriques des pixels. A l'issue de cette étape, l'utilisateur a la possibilité de lancer l'identification d'un polygone avec notre méthode. Il est aussi possible de lancer le processus sur tous les polygones de l'image. Tout ceci se réalise par l'intermédiaire d'une interface graphique (affichage de l'image segmentée, sélection du polygone, choix des paramètres, restitution des résultats, etc.). Rappelons que les différents paramètres possibles sont principalement le seuil minimum du score d'appariement et le niveau de profondeur souhaité. Dans le contexte d'expérimentation, il faut aussi évoqué les valeurs de pondération des attributs de chaque concept. Cela a été réalisé en concertation avec les géographes. Le ou les concepts ayant obtenus le meilleur score sont proposés à l'utilisateur pour attribuer un label au polygone étudié. Dans le cas d'un traitement sur un seul polygone, une arborescence représentant un sous-ensemble d'intérêt de l'ontologie est aussi affichée avec pour chaque concept les informations concernant les calculs de l'appariement.

Des expérimentations ont été réalisées sur une image satellite (Quickbird MS) montrant une partie de la ville de Strasbourg. Dans un premier temps, le seuil minimum du score n'a pas été utilisé, et le niveau de profondeur a été fixé à 3. Nous avons alors pu identifier de manière très significative la végétation, l'eau, l'ombre et le minéral. De mauvaises identifications ont concerné principalement l'ombre. Dans une seconde expérience, nous avons laissé la navigation aller jusqu'aux concepts feuille de l'ontologie. L'évaluation est ici difficile étant donné le nombre de concepts. Nous allons prochainement travailler sur ce point. Cependant, les premières observations sont très encourageantes.

5 Conclusion

Nous avons présenté une méthode d'identification d'objets géographiques, utilisant des connaissances sous forme d'une ontologie. Notre approche est basée sur le parcours et l'appariement d'objets avec des concepts de cette ontologie. Nous avons défini pour cela une mesure

²<http://protege.stanford.edu/>

de comparaison, orientée attributs, intégrant le calcul d'une similarité locale et une évaluation globale prenant en compte la hiérarchie des concepts. Nos travaux s'intègrent dans le processus de fouille de données multi-stratégies du projet FoDoMuSt pour assister la validation d'objets issus de la segmentation automatique d'images. Les premières expérimentations et évaluations humaines sont très encourageantes.

Nous prévoyons d'ajouter à notre méthode de nouvelles mesures, spécifiques à la sémantique interne de certains concepts de l'ontologie. De plus, les recherches à venir permettront d'améliorer et compléter notre approche, en tirant parti d'autres ressources du projet FoDoMuSt, comme sa base de données géographiques. Une perspective intéressante est d'intégrer un mécanisme de fouille de données permettant, par exemple, l'extraction d'attributs discriminants. Ce mécanisme permettrait de remplir le double objectif de piloter la navigation en affinant la stratégie adoptée et d'enrichir l'ontologie avec de nouvelles connaissances.

Références

- Brodeur, J. (2004). *Interopérabilité des Données Géospatiales : Elaboration du Concept de Proximité Géosémantique*. Thèse de doctorat, Université Laval, Canada.
- Mariño Drews, O. (1993). *Raisonnement Classificatoire dans une Représentation à Objets Multi-Points de Vue*. Thèse de doctorat, Université Joseph Fourier - Grenoble 1.
- Rodriguez, M. A. et M. J. Egenhofer (2003). Determining Semantic Similarity Among Entity Classes from Different Ontologies. *IEEE Transactions on Knowledge and Data Engineering* 15(2), 442–456.
- Schvaiko, P. et J. Euzenat (2005). A Survey of Schema-Based Matching Approaches. *Journal on Data Semantics IV*, 146–171.
- Schwering, A. et M. Raubal (2005). Measuring Semantic Similarity between Geospatial Conceptual Regions. In *The 1st Int. Conf. on GeoSpatial Semantics (GeoS)*, pp. 90–106.
- Tversky, A. (1977). Features of Similarity. *Psychological Review* 84, 327–352.

Summary

The FoDoMuSt ACI aims to create a multi-strategy data-mining process for an automatic recognition of geographic objects on satellite or air pictures. These pictures are segmented in order to distinguish polygons defined by a set of low level descriptors. In order to give polygons a semantic, a classification is first applied. If no geographic object is identified, we then try to match the polygon with the concepts of an ontology of geographic objects. We developed an algorithm to traverse the ontology and a semantic comparison measure, of which the parameters can be accorded to a specific matching context. This measure computes the relevance of a matching and is composed of a local component (dealing with the inner properties of the concept) and a global component (a linear combination of local measures). The matching algorithm and the measures have been developed in JAVA and integrated to the FoDoMuSt platform. The first experimentations and human evaluations are very encouraging.