

Mise à jour de référentiel géographique par fusion multi- source de données collaboratives

Thèse proposée à l'Institut National de l'Information Géographique et Forestière,
Laboratoire COGIT, en collaboration avec l'université de Tours

Mot clés : référentiel, géomatique, fouille du web, mobilités, qualification, intégration, fusion, crowdsourcing

Problématique générale

Le crowdsourcing, ou saisie collaborative, et le community-sourcing sont en plein développement dans le domaine de l'information géographique (Goodchild, 2007). Actuellement plusieurs sites permettent à des utilisateurs de faire remonter leurs données géographiques (OpenStreetMap, FixMyStreet, etc.). Le plus souvent ces informations de type traces issues de systèmes de géo-positionnement tel le GPS sont imparfaites et incomplètes. Elles deviennent néanmoins extrêmement nombreuses, de qualité croissante, et sont une source d'information riche pour un opérateur public chargé de la création et de la gestion d'un référentiel de données, tel que l'IGN. La convergence de l'action publique et de l'action privée dans ce contexte pose notamment de nombreuses questions sur l'intégration de données hétérogènes et sur la qualification des données. Des travaux de recherche ont d'ores et déjà établi l'efficacité des données collaboratives pour l'acquisition des bases de données géographiques, ou l'amélioration et la mise à jour de bases de données géographiques déjà existantes (Cao et Krumm, 2009 ; Zhang et al., 2010 ; Biagioni et Eriksson, 2012). Ces travaux s'appliquent, pour la plupart, au réseau routier (zones urbaines et voies rapides).

Problématique scientifique de la thèse

Dans ce contexte, le sujet de thèse portera sur l'analyse de déplacements pour la mise à jour des données de référence, ou au moins la levée d'alertes pour ses mises à jour, et concernera plus particulièrement la mise à jour des chemins et sentiers à travers l'analyse de traces obtenues par positionnement GPS. De nombreux sites internet dédiés à la randonnée, au VTT et à la course à pied mettent à la disposition des utilisateurs des traces de déplacements (VTTrack.fr, Randogps.net, TracesGPS.com, etc.). L'usage des traces GPS dans le contexte de la mise à jour de données géographiques engendre de nouveaux verrous méthodologiques et théoriques qu'il convient de dépasser.

Le premier verrou concerne la qualification et l'enrichissement sémantique des traces GPS. En effet, généralement, les traces GPS sont uniquement caractérisées par une suite ordonnée temporellement de coordonnées géographiques, nulle information sémantique et de qualité n'y est associée. De plus, elles présentent des imprécisions, incomplétudes et hétérogénéités liées à la qualité du GPS lui-même et aux conditions d'enregistrement (climatiques, de relief, etc.). Il s'agit donc de proposer un modèle spatio-temporel permettant de décrire ces caractéristiques et d'enrichir les traces sémantiquement. Par exemple, la vitesse de l'objet mobile et la confrontation aux caractéristiques du terrain peut permettre de discerner le type d'objets et donc le type de chemin (cyclable, pédestre, équestre).

Le deuxième verrou concerne l'identification et la qualification des objets à mettre à jour dans un référentiel, à partir de traces GPS. Il s'agit de détecter et qualifier les objets qui ont changé (géométrie ou attribut) ou qui n'existent pas dans le référentiel. Si de nombreux outils permettant de définir des liens d'appariement, c'est-à-dire de mettre en correspondance des données

représentant le même objet du monde réel (Devogele et Mustière, 2008 ; Olteanu 2008 ; Mascret et al. 2006) existent, une attention particulière sera accordée à la qualification des liens d'appariement.

La qualification des liens d'appariement s'avère une tâche difficile due aux hétérogénéités spatiotemporelles des traces GPS. Afin de qualifier les liens d'appariement, une solution serait de proposer une méthode basée sur la fusion de différentes sources de données telles que les capteurs fixes ou mobiles, les cartes topographiques, les images aériennes ou satellites ainsi que des données provenant du Web, non localisées mais dont la localisation est implicite. On peut, par exemple, imaginer confronter une trace GPS susceptible de révéler un chemin manquant dans les données de référence en croisant des informations textuelles associées à la trace, des images ou une analyse du relief de la zone concernée. L'utilisation des sources multiples pour qualifier les données représente un enjeu essentiel pour pouvoir mettre à jour un référentiel à la qualité maîtrisée. Ainsi un troisième verrou scientifique concerne la fusion des sources de données hétérogènes et la gestion des conflits, puisque nous devons proposer des méthodes adaptées pour simplifier, synthétiser et combiner l'information spatiale et non spatiale.

Profil et Compétences

- Titulaire d'un diplôme de Master en géomatique ou en informatique avec une sensibilité pour le domaine de l'information géographique.
- Solides connaissances de programmation
- Bonne capacité relationnelle, motivation pour la recherche et pour le travail en équipe, esprit d'initiative, capacité rédactionnelle.

Accueil et encadrement

La thèse sera réalisée au Laboratoire COGIT de l'IGN, Saint-Mandé (limitrophe Paris), en partenariat avec l'université de Tours. Le contrat doctoral, de trois ans, pourra être réalisé avec ou sans charge de cours, selon le profil du candidat et les besoins de l'Ecole Nationale des Sciences Géographiques.

Co-Direction de thèse : Sébastien Mustière, HDR, Laboratoire COGIT, IGN

Thomas Devogele, Professeur, Université de Tours

Encadrement : Ana-Maria Olteanu-Raimond, Laboratoire COGIT, IGN

Contact : Pour tout renseignement: Ana-Maria Olteanu-Raimond (ana-maria.raimond@ign.fr, 01 43 98 80 00 +75 50)

Bibliographie

Biagioni, J. et Eriksson, J. Inferring road maps from GPS traces: Survey and comparative evaluation. In Transportation Research Board, 91st Annual, 2012.

Cao, L. et Krumm, J. , From GPS Traces to a Routable Road Map, 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL GIS 2009), November 4-6, 2009, Seattle, WA, pp. 3-12.

Goodchild, M.F., Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. In: International Journal of Spatial Data Infrastructures Research, 2007, Vol. 2, pp. 24-32.

Mascret A., Devogele T., Le Berre I., Hénaff A., Coastline Matching Process based on the discrete Fréchet distance - In Proceedings of the 12th International Symposium on Spatial Data Handling (SDH), W. Kainz, A. Reidl, and G. Elmes (eds.), Vienna, Austria, Springer-Verlag, pages 383-400, 2006.

Mustière, S. et Devogele, T., Matching networks with different levels of detail. Geoinformatica, 2008.

Olteanu A-M., Fusion de connaissances imparfaites pour l'appariement de données géographiques, Mémoire de thèse, Université Marne la Vallée, 2008.

Zhang, L., Thiemann, F., Sester M., Integration of GPS Traces with Road Map, ACM SIGSPATIAL'10, November 2-5, 2010, San Jose, CA, USA.