

Structures de données pour objets flous dans les Systèmes d'Information Géographiques

Data structures for fuzzy objects in Geographic Information Systems

Enguerran Grandchamp

LAMIA Laboratory, French West Indies University,
Campus de Fouillole, 97157 Pointe-à-Pitre, Guadeloupe, France
egradch@univ-ag.fr

Topics : 1 – Vegetation 2 – Biodiversity, agrobiodiversity 7 – Environment monitoring and modeling 13 – Classification techniques and data mining

Keywords: GIS, classification, fuzzy, data structures

Résumé

Les structures de données vecteur présentes dans les *Systèmes d'Information Géographiques (SIG)* sont bien adaptées à la représentation d'objets bien localisés dans l'espace et le temps. C'est le cas de la plus part des objets (tels que les habitations, les routes, etc.) ou concepts (frontières, limites administratives, etc.) manufacturés. Mais les données naturelles ont souvent des contours flous car leur évolution dans l'espace suit un gradient lié à des phénomènes physiques (changement progressif d'altitude, de température, etc.). C'est le cas par exemple des forêts pour lesquelles les transitions peuvent s'étendre sur des centaines ou milliers de mètres. Les représentations classiques de telles données sont (i) un modèle vectoriel stricte fixant arbitrairement les frontières des objets (ii) un modèles raster flou donnant une indication sur la fonction d'appartenance de chaque éléments spatial (à définir) à chacun des objets. Dans cet article nous proposons différentes manières de représenter des données diffuses à l'aide de structures de données vectorielles avec des illustrations sur la classification et la représentation de forêts. Les structures présentées sont (i) un modèle vectoriel flou complet et simplifié (ii) un modèle basé sur la notion de zone de transition (iii) un modèle basé sur la notion de squelette d'objet.

1. Introduction

Les objets manipulés par un SIG possèdent une information géographique contenant leur description géométrique (forme, coordonnées, etc.). Leur stockage doit donc faire appel à des structures de données adaptées. La représentation de points, de lignes ou des surfaces est depuis longtemps maîtrisée et fait appel à un chaînage classique de sommets (vecteur) localisés par leurs coordonnées dans l'espace (2D ou 3D) conduisant au modèle vectoriel de représentation des données.

Ce type de structure est bien adapté aux objets pour lesquels il est facile de localiser les contours, mais lorsqu'il s'agit de représenter des objets ou des phénomènes dont les frontières sont imprécises (dans leur localisation ou dans leur définition même) ces structures trouvent leurs limites et conduisent à une représentation erronée ([1-5]). Dès 1997, dans les conclusions du numéro spécial *Spatial Data Types for Database Systems* **Error! Reference source not found.** les auteurs soulignent l'importance des structures de données dans les *SIG* et leur manque d'adaptation aux données floues.

Nous présentons dans la suite de cet article plusieurs modèles permettant de palier ces inconvénients, le premier est basé sur une représentation vectorielle floue (section 2) avec un modèle complet et un modèle simplifié, le second est basé sur des zones de transition (section 3) et enfin le troisième sur une squelettisation des objets (section 4).

2. Structures floues

La première représentation utilise une partition du territoire en objets vectoriels élémentaires suivis de leur classification à l'aide d'un arbre de décision flou. Cette première représentation nécessite une carte floue par objet à représenter. Les objets vectoriels sont définis de manière sémantique et doivent tenir compte des paramètres permettant de définir l'objet ou le phénomène à modéliser. Il s'agit, par exemple, dans le cas de la modélisation de forêts de tenir compte de paramètres topographiques (altitude, pente, etc.) et environnementaux (humidité, sols, etc.).

2.1 Représentation complète

Dans cette représentation, chaque objet élémentaire O_i est annoté par un nombre N de caractéristiques ($F_i = \{F_{i1}, F_{i2}, \dots, F_{iN}\}$) provenant de la définition sémantique (ou physique) de l'objet ou du phénomène.

L'emprise spatiale de l'objet flou est ensuite définie par une méthode de classification utilisant des arbres de décision flous de type *FT* (arbres fonctionnels [2]) ou *C4.5* ([7]).

Ces arbres de décision calculent un degré de confiance par régression logistique [8] pour chaque objet élémentaire : $\mu_i = \{\mu_{i1}, \dots, \mu_{in}\}$, μ_{ik} indique le degré de confiance dans l'assertion : l'objet élémentaire i appartient à l'objet (ou le phénomène) k .

Ceci conduit à un modèle flou complet de chaque objet. La figure 1. illustre le modèle obtenu pour trois types de forêts présentes sur la Basse-Terre en Guadeloupe. Plus le niveau de gris est élevé plus la valeur du degré de confiance est élevée.

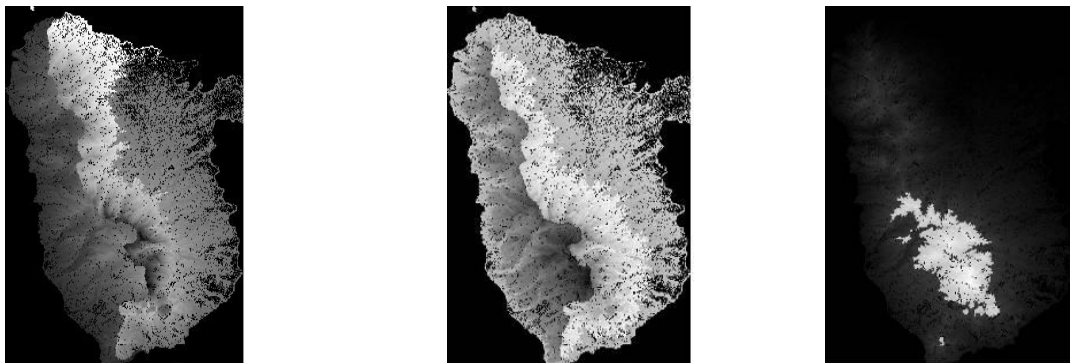


Figure 1 – Modèle flou complet

2.2 Représentation simplifiée

Pour faciliter l'interprétation et la visualisation, nous avons également définis des représentations simplifiées du modèle précédent. La figure 2. donne trois représentations simplifiées du modèle complet situé à droite dans la figure 1.

Ces représentations permettent de mieux faire apparaître le caractère transitoire définissant l'objet et de mettre l'accent sur certaines de ses parties en définissant des zones homogènes.

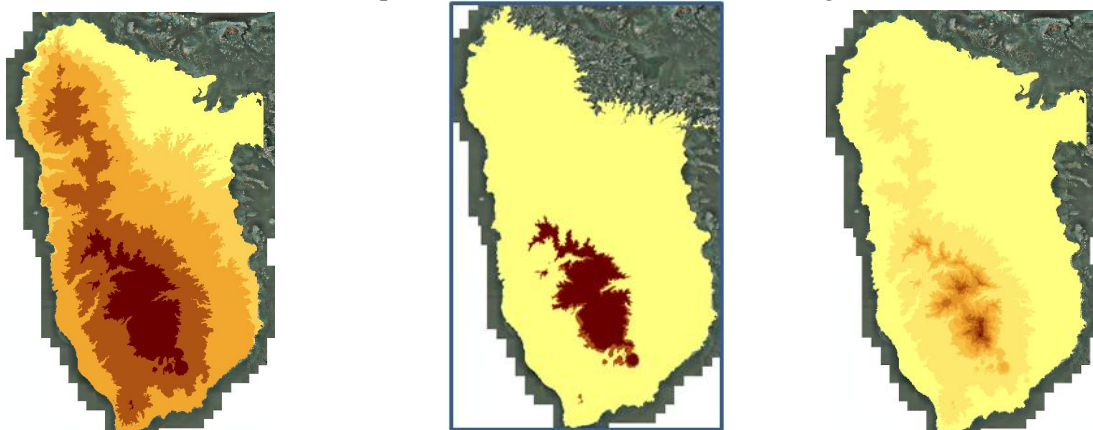


Figure 2 – Modèle flou simplifié

Pour définir ces zones on fixe un nombre arbitraire (dépendant du contexte) d'intervalles sur le degré de confiance (g_v) $I^v = \{I_0^v, I_2^v, \dots, I_{g_v}^v\}$ avec $I_0^v = 0$, $I_{g_v}^v = 1$ et $I_p^v < I_{p+1}^v$. Un groupement est alors définis comme étant $G_{kp}^v = \{O_i \mid I_{p-1}^v < \mu_{ik} \leq I_p^v\}$.

Il existe plusieurs manières de fixer les intervalles en fonction de ce que l'on souhaite obtenir.

1. Pour une partition régulière de $[0,1]$, on fixe $I_p^v = \frac{p}{g_v}$
2. Pour des intervalles centrés sur les valeurs du degré de confiance les plus représentées. I_p^v est fixé de manière à ce que $|G_{kp}^v| \simeq \frac{M}{g_v}$ ($|\cdot|$ renvoyant le cardinal et M étant le nombre d'objets élémentaires), $M = \sum_{p=1}^{g_v} |G_{kp}^v|$). Pour cela on choisit les intervalles de manière à minimiser $\sum_{p=1}^{g_v} (|G_{kp}^v| - \frac{M}{g_v})^2$.
3. En définissant des intervalles de faible amplitude autour des valeurs élevées du degré de confiance, on représente plus en détail le centre des objets. Dans ce cadre, la règle à respecter pour fixer les valeurs des bornes est : $I_{p+1}^v - I_p^v < I_p^v - I_{p-1}^v$. Toute fonction f strictement croissante et concave de $[0,1]$ dans $[0,1]$ (avec $f(0) = 0$ et $f(1) = 1$) permettra d'obtenir une telle partition à partir d'une partition régulière telle qu'exprimée en 1 : $I_p^v = f(\frac{p}{g_v})$

Cette représentation se rapproche de modèles basés sur des buffers concentriques réguliers ([9,10]) mais avec une plus grande fiabilité puisque les zones concentriques sont ici irrégulières et adaptées à l'évolution locale du degré de confiance.

3. Zones de transition

L'inconvénient de la représentation précédente est quelle nécessite une carte floue par objet ce qui peut être contraignant dans le cas où l'on souhaite éditer ou visualiser une unique carte représentant l'ensemble des objets. Afin de préserver la notion d'objet flou et de phénomène de recouvrement, nous définissons un second modèle basé sur la notion de zone de transition. Ces transitions sont localisées dans des zones complexes où aucun objet (ou phénomène) n'est clairement majoritaire.

On définit les zones de transition à partir des vecteur de degré de confiance obtenus précédemment. Pour cela, on fixe un seuil S et pour chaque objet élémentaire O_i on trie les objets auxquels il peut appartenir par ordre décroissant de degrés de confiance : $\{\mu_{ij_1} \geq \mu_{ij_2} \geq \dots \mu_{ij_n}\} \rightarrow \{O_{j_1}^i, O_{j_2}^i, \dots, O_{j_n}^i\}$. On attribue à chaque O_i la liste minimale des objets définie par $E_i(S) = \{O_{j_1}^i, O_{j_2}^i, \dots, O_{j_p}^i \mid \sum_{k=1}^p \mu_{ij_k} > S\}$.

On remarque que $E_i(0) = \{O_{j_1}^i\}$ revient à attribuer chaque objet élémentaire à l'objet pour lequel le degré de confiance est le plus élevé. Dans les autres cas, les objets élémentaires associés à plusieurs objets font partie de zones de transitions dont la complexité dépend du nombre d'objets concernés.

La figure 3. illustre le résultat obtenu pour différentes valeurs de S , les zones de transition sont affichées en noir. Les objets (ici des forêts) sont représentés par des couleurs différentes.

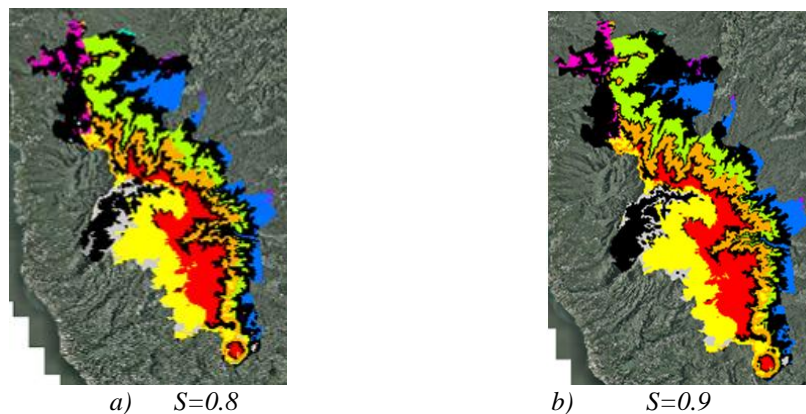


Figure 3 – Zones de transition (noir)

4. Squelettisation

Les deux représentations précédentes sont basées sur des structures de données existantes au sein des SIG. Mais le passage à une véritable représentation vectorielle floue nécessite de redéfinir les objets élémentaires (point, ligne, surface) et les opérations ensemblistes (union, intersection, etc.) présents dans les SIG. Ce passage ne peut se faire qu'en plusieurs étapes en s'appuyant par exemple sur des modèles comme celui présenté dans [11] où les auteurs représentent un objet flou par trois objets stricts : l'intérieur, la frontière et l'extérieur (L'extérieur représentant la zone d'incertitude). C'est une vue restrictive des objets qui a néanmoins le mérite de simplifier les traitements et analyses et que le modèle précédent étend en ne distinguant pas ces trois parties et en proposant différentes zones intermédiaires.

Mais certains aspects des ensembles flous sont encore manquants dans notre modèle comme la notion d'ensemble ouvert (figure 4) ou de gradient local de la fonction d'appartenance. Les polygones ouverts permettraient de conserver la notion de gradient et de ne pas fixer de limite à un objet mais ils nécessiteraient une algèbre particulière et les opérations ensemblistes sont à redéfinir.

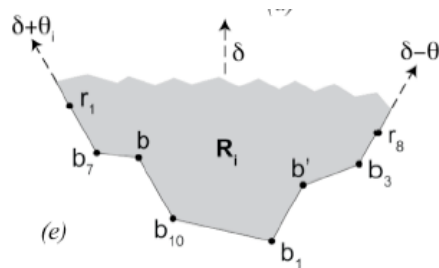


Figure 4 – Polygone ouvert (extrait de [12])

Afin de réaliser un premier pas vers un tel modèle, entièrement développé pour les ensembles flous, nous avons définis la notion de squelettisation vectorielle afin de tenir compte du gradient d'évolution de l'objet et d'autoriser une représentation simultanée de plusieurs objets sur la même carte. Cette squelettisation sera en outre nécessaire lors de l'application des intersections de polygones ouverts.

La première étape consiste, à partir d'une représentation simplifiée d'un objet vectoriel flou (section 2.2), à ajouter à chaque sommet des différentes zones concentriques un vecteur indiquant le gradient local de la fonction d'appartenance. La figure 5. illustre cet ajout d'information au modèle simplifié.

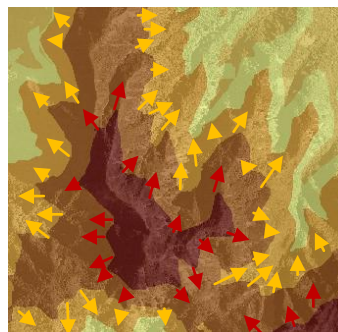


Figure 5 – Gradient local de la fonction d'appartenance

Ce gradient local prend son importance lors de l'union ou l'intersection de deux ensembles, en apportant plus de souplesse puisqu'il donnera lieu à une déformation des contours des zones concentriques

Ce modèle est bien adapté aux phénomènes transitoires (spatialement ou temporellement).

5. Conclusion

Le développement de structures spécifiques pour les SIG adaptées aux données floues est nécessaire car les structures existantes ne sont pas fiables.

L'avantage des deux premières représentations proposées (modèle vectoriel flou et zones de transitions) et qu'elles utilisent des structures de données déjà existantes dans les SIG. Ces représentations sont plus fiables que les représentations vectorielles présentent dans la littérature et utilisant par exemple des buffers concentriques réguliers.

6. Références

- [1]. Altman, D., 1994, Fuzzy set theoretic approaches for handling imprecision in spatial analysis, *Internat. J. Geographical Inform. Systems* 8 (3) 271–289.
- [2]. Cross, V.V., 2001, Fuzzy extensions for relationships in a generalized object model, *International Journal on Intelligent Systems* 16, 843–861.
- [3]. Fisher, P., 2000, Sorites paradox and vague geographies, *Fuzzy Sets and Systems* 113 7–18.
- [4]. Kainz, 2011, *Introduction to Fuzzy Logic and Applications in GIS – Examples*
- [5]. Sawatzky, D. Raines G. L., and Bonham-Carter, G., 2008, *Spatial Data Modeller*, Technical Report,
- [6]. Schneider, M., 1997, *Spatial Data Types for Database Systems Finite Resolution Geometry for Geographic Information Systems Series: Lecture Notes in Computer Science*, ISBN 978-3-540-63454-6, Vol. 1288, 275 pp
- [7]. Quinlan, R., 1993, *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA.
- [8]. Gama, J, 2005, *Functional Trees*. Niels Landwehr, Mark Hall, Eibe Frank. *Logistic Model Trees*.
- [9] Ursula C. Benz and al., "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information," *ISPRS Journal of Photogrammetry & Remote Sensing*, vol. 58, pp. 239-258, 2004.
- [10] M. Karimi, M.B. Menhaj, and M.S. Mesgari, "Preparing Mineral Potential Map Using Fuzzy Logic In GIS Environment," in *ISPRS*, vol. XXXVII, 2008.
- [11] Jan T. Bjorke, "Topological relations between fuzzy regions: derivation of verbal terms," *Fuzzy Sets and Systems*, vol. 141, pp. 449–467, 2004.
- [12] S. Coros, JingBo Ni, and Pascal Matsakis, "Object Localization Based on Directional Information: Case of 2D Vector Data," in *14th Int. Symposium on Advances in Geographic Information Systems (ACM-GIS)*, 2006.