Le problème de la sectorisation à partir de centres en cartographie : résolution par la Recherche Opérationnelle

TANG Xin, SOUKHAL Ameur, T'KINDT Vincent

Université François-Rabelais de Tours, Laboratoire d'Informatique, Polytech'Tours ; 64 avenue Jean Portalis, 37200
tang@poste.isima.fr, {ameur.soukhal,vincent.tkindt}@univ-tours.fr.fr

Mots-Clés : *Cartographie statistique, Partitionnement, Recherche opérationnelle, Complexité, Modélisation mathématique.*

1 Introduction

La cartographie statistique est un domaine d'étude en plein essor. Partant d'une carte géographique et de données de nature statistique, la cartographie statistique consiste à réaliser des traitements sur les données et à visualiser les résultats directement sur la carte. Le problème de la sectorisation est un problème de partitionnement particulier. Étant donnée une zone géographique (exemple : département, région, pays, continent,...) découpée en éléments (exemple : communes, communautés de communes, départements, ...) sur chacun desquels une ou plusieurs informations statistiques sont attachées (exemple : surface, polulation, nombre de clients,...), l'objectif est de déterminer un ensemble de J secteurs (un secteur est défini comme un ensemble d'éléments). Ces secteurs doivent être "homogènes", c'est-à-dire dont la quantité d'information statistique varie peu d'un secteur à l'autre ou est proche d'une valeur de référence. Ce problème est largement étudié en théorie des graphes sous le nom de partitionnement de graphe. On suppose par ailleurs dans le problème étudié qu'à chaque secteur on attache un centre qui est défini comme étant un élément géographique donné par avance (par exemple, une ville de départ/arrivée pour un commercial). On connait autant de centres qu'il y a de secteurs à construire.

2 Modélisation du problème

Le problème de k sectorisation à centres fixés peut être modélisé à l'aide d'un graphe comme suit. Soit $G = (V, E)$ un graphe non orienté où V est l'ensemble des sommets et E est l'ensemble des arêtes. Un sommet représente un élément géographique de la carte associée au graphe, et une arête du graphe lie deux sommets voisins, c'est-à-dire deux éléments géographiques contigus sur la carte. On note $S = \{v_1, v_2, ..., v_s\}$ l'ensemble des centres qui est une donnée du problème. $P = \{p_1, ..., p_J\}$ est l'ensemble des partitions (secteurs) à calculer. A chaque sommet v_i on associe un vecteur de Q valuations $c^i = [c^i_1, ...c^i_q, ...c^i_Q]$ (les informations statistiques attachées à chaque élément de la carte). Pour chacune de ces informations statistiques on construit un critère d'évaluation de l'homogénéité entre deux secteurs, soit on construit un critère de la mesure de la distance moyenne, mesurée sur chaque secteur, à une valeur de référence e_q^j du secteur p_j. Une contrainte forte du problème est que chaque secteur soit connexe, i.e. chaque partition construite sur le graphe G correspond à un ensemble connexe de sommets.

Ce problème de sectorisation est NP-difficile [2], ce qui justifie l'utilisation de méthodes heuristiques pour résoudre ce type de problèmes . Dans cette communication nous nous intéressons à des premières pistes de modélisation et de résolution que nous abordons très brièvement dans la section suivante.

3 Approches de résolution

D'un point de vue théorique, ce problème de sectorisation peut être vu comme un problème de classification supervisé [1]. Néanmoins, cette approche dispose de nombreux inconvénients : impossibilité de prendre en compte plusieurs critères de classification (ce qui se traduit par une impossibilité de considérer plusieurs informations statistiques pour chaque élément), inadéquation de la notion d'homogénéité entre la méthode de classification et le problème de sectorisation. Notre objectif est de reprendre ce problème sous l'angle de la Recherche Opérationnelle en utilisant ses outils pour proposer des algorithmes efficaces de résolution.

La première question que nous nous posons est la suivante : le problème de sectorisation à centres fixés est-il modélisable par la programmation mathématique linéaire mixte ? Nous proposons un premier modèle qui reprend l'ensemble des contraintes et critères du problème. Ce modèle sera ensuite utilisé à des fins de comparaison avec une heuristique rapide, par exemple par construction progressive. En effet, cette heuristique devra pouvoir être validée sur des instances de très grande taille (plusieurs milliers de sommets dans le graphe).

Références

[1] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Wiley-Intersciences, 2nde édition. 2001.

[2] Michael R. Garey and David S. Johnson, Computers and Intractability : A Guide to the Theory of NP-Completeness, *W.H. Freeman and Company*, 1979.