

Une méthode de classification sous contrainte de contiguïté en géo-marketing

Mihai Calciu

Université des Sciences et Technologies de Lille

Une méthode de classification sous contrainte de contiguïté en géo-marketing

Résumé

Bon nombre d'opérateurs marketing agissant sur un territoire possèdent des données sur leurs clients. A partir des adresses, ils arrivent, par des techniques de géo-codage, à positionner leur clientèle sur des cartes et à agréger cette information au niveau d'une rue, d'un îlot ou d'un quartier. Pour leur permettre de mieux connaître leur territoire, nous allons analyser quelques modalités afin d'établir des typologies de ces unités territoriales, tout en prenant en compte les contraintes spatiales. Une méthode de classification sous contrainte de contiguïté et des programmes informatiques nécessaires à sa mise en place sont proposés.

A clustering method under contiguity constraint in geo-marketing

Abstract

Many marketing operators acting in geographical areas have information about their customers. Using their address and geo-coding techniques, they are able to place the customers on maps and to aggregate this information at the level of a street, a group of buildings, a district. In order to give them a better knowledge of their territory, we analyse some clustering techniques taking into account spatial constraints. A clustering method under contiguity constraint and computer programs for its implementation are being suggested.

Une méthode de classification sous contrainte de contiguïté en géo-marketing

1. Introduction

Pour les agents économiques, le territoire est en même temps source d'informations et d'opportunités. Beaucoup d'opérateurs marketing, qui agissent sur le territoire comme les distributeurs, les banques, les assurances, etc., possèdent des données sur leurs clients (leur comportement d'achat, leur adresse). Ils ont aussi des possibilités de géo-codage qui leur permettent l'ajout d'informations géographiques à une base de données de telle sorte que les objets auxquels se réfèrent les données puissent être affichés sur une carte. Ils arrivent donc à positionner placer leur clientèle sur des cartes et à agréger cette information au niveau d'une adresse, d'une rue, d'un îlot ou d'un quartier. Souvent, les opérateurs peuvent obtenir des données secondaires en provenance d'organismes spécialisés comme l'INSEE à un niveau d'agrégation aussi fin que l'îlot. En s'appuyant sur ces données, obtenues directement ou indirectement, et en les agrégeant au niveau de l'unité géographique choisie (îlot, quartier, région), les sociétés peuvent établir des typologies (classifications) de ces unités pour mieux connaître leur territoire et mieux adapter leur politique.

En marketing, Stewart et Punj (1983) distinguent trois grands domaines d'applications de l'analyse typologique: 1) la segmentation des marchés, 2) l'identification des types de clients et la compréhension de leurs comportements, 3) la politique de développement de produits où sont regroupés des marques ou des produits mettant en évidence des groupes en compétition et permettant d'adapter l'offre par rapport à la concurrence.

L'intérêt de la classification, en général, provient d'un nombre de contraintes qui s'imposent quand une activité humaine essaye d'approcher, voire de contrôler, une réalité complexe. La première *contrainte* est de nature *cognitive*: "le besoin de simplifier une réalité complexe par la constitution de groupes d'objets ou d'individus similaires"(Chandon et Pinson, 80). En marketing, on classe les clients pour identifier des groupes qui se ressemblent du point de vue des variables qui influent sur l'acte d'achat. Dans ce cas, il y a ici non seulement une *contrainte cognitive* mais aussi une *contrainte économique*; on ne peut pas toujours adapter l'offre et la politique marketing aux attentes spécifiques de chaque client mais on doit assurer une certaine économie d'échelle.

Quand les opérateurs marketing agissent sur le territoire, s'ajoutent également des *contraintes spatiales*. Leurs politiques et leurs actions marketing doivent s'adapter aux attentes de la clientèle de leur territoire. Certains auteurs appellent géo-marketing cette démarche dont le but est d'adapter la politique marketing à l'espace géographique de la firme. Dans une telle approche, la localisation d'un opérateur (distributeur, agence, etc.) doit prendre en compte la typologie des clients potentiels de l'espace d'opération (îlots, régions, pays).

Même si l'analyse typologique a fait l'objet de nombreuses recherches en marketing, peu d'études ont pris en compte les contraintes spatiales. C'est la raison pour laquelle cette démarche propose une méthode qui permet de classer de manière automatique la clientèle localisée dans des

petites unités géographiques (les îlots dans les villes) tout en tenant compte des contraintes spatiales (la contiguïté et, éventuellement, la distance géographique)¹.

Deux approches possibles sont envisagées.. L'une consiste à introduire une contrainte de contiguïté et/ou de distance géographique dans un algorithme de classification.

L'autre vise à proposer des mécanismes de regroupement des types géographiques obtenus, par une méthode de classification sans contrainte, en zones contiguës convenables, pour une politique marketing adaptée au territoire. On considère que les types de consommateurs sont éclatés dans le territoire de manière similaire aux groupes en conflit en Bosnie avant le lancement des hostilités. L'algorithme à développer devrait constituer des zones contiguës regroupant plusieurs entités géographiques du même type, en utilisant des critères de type frontière commune, couloir de communication sur territoire étranger entre deux zones du même type, etc.

Cette démarche explore l'introduction des contraintes spatiales dans un algorithme de classification par deux modalités. La première modalité introduit les coordonnées géographiques parmi les variables qui caractérisent les observations et fait varier leur poids en fonction de l'importance qu'on veut donner à la distance géographique dans la classification². La deuxième modalité introduit une contrainte de contiguïté à chaque étape d'un algorithme de classification.

Un algorithme de classification hiérarchique sous contrainte de contiguïté est alors proposé, testé et évalué à partir des données sur l'utilisation de services bancaires, concernant 651 îlots de la ville de Roubaix.

Un programme informatique, intitulé TYPOGEO, a été mis au point pour exécuter l'algorithme. Il est réalisé dans une approche orientée objets en langage C++.

Pour compléter le programme et le faire évoluer vers un système d'aide à la décision géo-marketing, plusieurs programmes utilitaires ont été créés:

- MCARTE, programme d'enregistrement des coordonnées de carte (sous Windows), qui permet de construire des bases de données sur des observations territoriales.

- MIFVOISI, utilitaire de reconnaissance des contiguïtés à partir de fichiers MIF³ du système de cartographie MAPINFO.

Un système d'analyse typologique des cartes en langage MAPBASIC de MAPINFO a été également mis en place. Il exploite les résultats obtenus par les algorithmes et programmes énumérés ci-dessus pour servir d'outil d'aide aux décisions dans la politique territoriale marketing.

¹ L'idée d'introduire la contrainte de contiguïté dans la classification des îlots est issue d'un besoin exprimé dans le domaine de la distribution de services bancaires et a été suggérée par notre collègue C. Benavent, maître de conférences à l'IAE de Lille. Cette démarche est un élément d'un projet de recherche plus vaste, en géo-marketing, conduit avec le soutien de la Caisse d'Epargne de Flandre

² Cette manière de faire nous a été suggérée aussi par J-L. Chandon lors d'une présentation faite aux journées AFM de Lille, en septembre 1994.

³ Les fichiers MIF sont des fichiers texte ASCII ayant l'extension ".mif"

2. Une brève revue des typologies territoriales

Un parcours sommaire de la littérature marketing et de la littérature en géographie concernant les méthodes de classification montre que les techniques de classification sont encore peu adaptées à la délimitation des régions géographiques.

Quelques algorithmes de classification sous contrainte de contiguïté spatiale sont présentés par Gordon (1981), qui montre que la contiguïté entre les objets, peut être représenté convenablement sous forme d'un graphe connexe. Chaque objet est représenté par un sommet du graphe. Deux objets sont contigus, lorsqu'ils sont reliés par une arête et le graphe est connexe quand il existe une arête entre chacune des paires d'objets.

La classification sous contrainte cherche à enlever des arêtes pour décomposer le graphe en sous graphes contenant des objets similaires.

La situation la plus simple est quand le graphe est linéaire, comme dans la figure 1, ou on impose une contrainte d'ordre sur les objets, signifiant qu'un objet peut être contigu avec au maximum deux autres. La classification sous une telle contrainte de contiguïté est appliquée en stratigraphie dans l'analyse du pollen préservé dans les strates sédimentaires. Dans l'évolution linéaire du temps on cherche des périodes compactes (strates contiguës) caractérisées par des types similaires de pollen, ce qui permet d'identifier les périodes du passé où la végétation a eu une composition stable. La contrainte de contiguïté est ici spatiale et temporelle en même temps, mais elle est surtout linéaire, ce qui réduit beaucoup le nombre de partitions possibles et donc simplifie la tâche de classification.

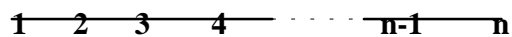


Fig.1 Représentation sous forme d'un graphe linéaire des objets soumis à une contrainte de contiguïté

L'ensemble de n objets peut être partitionné en g groupes d'objets contigus en enlèvent (g-1) arêtes ou en plaçant des "marqueurs" entre quelques uns des paires d'objets.

Des classifications hiérarchiques de tels objets ordonnés, qui donnent, pour chaque nombre de groupes, la partition qui minimise un critère de variabilité⁴, sont obtenus par des algorithmes divisives (Gill, 1970; Gordon et Birks, 1972) et agglomératives (Mehring et al., 1977) qui sont présentés par Gordon (1981). L'algorithme divisive introduit à chaque étape (itération) un marqueur dans un intervalle entre les objets à classer, tout en gardant fixes les marqueurs placés précédemment et en positionnant le nouveau marqueur de manière à obtenir la plus grande réduction de la mesure de variabilité (la somme des carrés). Basé sur la même contrainte l'algorithme de Fisher (1958) décrit par Chandon et Pinson (1981) trouve par une relation de récurrence la partition optimale en g groupes.

Gordon (1973) propose la méthode des "barrières variables" pour distinguer entre les nucléés des groupes et les objets transitionnels. Il introduit des barrières de hauteur variable (entre 0

⁴ En générale la somme des carrés des écarts des membres des groupes par rapport aux centres de groupes auxquels ils appartient.

et 1) dans tous les intervalles entre les objets. Les hauteurs des barrières sont trouvées en minimisant, sous certaines contraintes, un critère de moindres carrés pondérés (par rapport à des centroïdes de groupe). Les objets dis similaires sont séparés par des hautes barrières. Les objets de transition apparaîtront entre des barrières hautes adjacentes. Dans la figure 2(a) 5 et 6 sont des objets de transition entre les groupes homogènes (1,2,3,4) et (7,8,9,10). Un arbre de classification peut être construit à partir de ces barrières, comme le montre la figure 2(b).

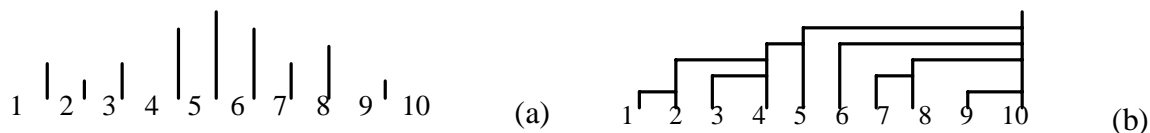


Fig.2 - Construction d'un arbre de classification par la méthode des barrières variables

Dans les algorithmes analysés jusqu'ici la contiguïté résulte d'une contrainte d'ordre imposé aux objets, et même si on peut les utiliser avec des contraintes de contiguïté qui ont des représentations sous forme de graphes un peu plus complexes, les conditions⁵ que doivent respecter les structures de ces graphes sont trop limitatives, ce qui fait que ces méthodes de classification ne sont applicables qu'à des configurations de régions, très particulières.

Quelques essais de résolution des problèmes de groupement de régions nous ont inspirés dans notre démarche.

Berry (1961) utilise un calcul de distance généralisée pour grouper neuf grandes circonscriptions du recensement des Etats Unis. Dans cette étude, il s'intéresse à six facteurs à partir d'une série de mesures et calcule les "distances généralisées entre points" (D^2) dans un espace de six dimensions. En mettant ensemble les deux circonscriptions les plus semblables, il a ainsi ramené les neuf régions à huit. Puis il a calculé les distances généralisées entre les huit régions restantes et, ainsi de suite, par répétition de la procédure, il a réduit progressivement le nombre de régions jusqu'à ce que toutes les régions soient réunies en une seule. Il obtient donc neuf niveaux de divisions régionales qui sont enregistrés dans un arbre de classification. Chacune de ces divisions est valable à son niveau particulier; toutefois, laquelle est globalement la plus valable? Berry (1969) montre, qu'au cours de ce processus, on gagne progressivement en généralité et on perd en précision. On peut évaluer la perte du détail, à chaque étape, à l'aide de la distance généralisée calculée auparavant. Cette méthode ressemble beaucoup à l'algorithme classique de classification hiérarchique ascendante.

Cette approche ne prend pas en compte l'exigence de contiguïté qui est spécifique à la géographie et, par extension, au géo-marketing: "la Californie ne peut pas être mise dans la même région que New Jersey parce qu'elle est séparée géographiquement" (Hagood et Price, 1952, p.542).

⁵ la structure du graphe doit être telle que l'enlèvement de n'importe quelle arête, décompose le graphe en deux sous graphes connexes

Quand on cherche à regrouper deux aires contiguës, on devrait comparer progressivement chaque unité spatiale avec ses voisins. Les approches développées dans la littérature géographique reposent sur l'analyse de variance et l'analyse de corrélation.

Zobler (1958) a utilisé l'analyse de variance pour rattacher un état des Etats-Unis (la Virginie) à un des trois groupes d'états limitrophes. Il part du raisonnement que, lorsqu'on construit des régions à partir d'unités plus petites (dans ce cas les états), il existe deux facteurs de variation: variation entre les Etats à l'intérieur d'une région (variation intra-régionale) et variation entre les régions (variation inter régionale). Pour mesurer les variances inter et intra-régionales, Zobler emploie l'analyse de variance en considérant trois cas: il rattache la Virginie, tour à tour, à chacune des trois régions. Pour chaque cas, il divise la "variance inter régionale", qui mesure les variations des régions par rapport à la moyenne de toutes les régions, par la "variance intra-régionale" qui mesure les variations des Etats par rapport à leurs moyennes régionales respectives, de façon à obtenir le rapport des variances ou rapport de Fisher-Snedecor (rapport F). Ainsi, le rapport des variances définit la mesure dans laquelle l'opération de groupement a réussi à maintenir ensemble les Etats semblables et à maintenir séparés les Etats dissemblables. La solution plus générale suggérée est de rattacher une aire à celui des groupements d'aires limitrophes pour lequel l'union donne une valeur de F maximum.

Dans cette formulation, la solution proposée est difficilement applicable à un grand nombre de régions, comportant beaucoup de niveaux de groupements.

Une autre méthode, proposée par M. J. Hagood (1943), repose sur la corrélation et les liens régionaux. Le problème étudié par Hagood consistait à partager les Etats-Unis en six ou douze groupements d'états contigus. Il fallait rendre chaque groupement d'états aussi homogène que possible, compte tenu de 104 variables tirées du recensement démographique et agricole de 1940. Les variables ont été partagées en deux groupes principaux (agriculture et population) pour établir "le profil démographique et agricole" de chaque Etat. Les 104 variables ont été réduites de façon à ce que la moyenne de chacune d'entre elles, pour les 48 Etats, soit égale à 50, que l'écart type soit égal à 10, et on a calculé les coefficients de corrélation (r) entre les profils d'états limitrophes. Les Etats qui marquaient une forte corrélation avec leurs voisins étaient considérés comme "noyaux" et formaient le centre d'une région homogène; les autres Etats étaient considérés comme "marginaux" et devaient être affectés à une des régions définies par les noyaux. Pour l'affectation M.J.Hagood se sert d'un indice comparatif unique "indice combiné agricole-démographique" qu'elle a calculé par analyse factorielle.

Les liens établis par calcul de corrélation ne résolvent pas les problèmes de régionalisation, en ce sens qu'ils ne créent pas l'homogénéité là où elle n'existe pas. Cette approche, comme celle de Zobler, n'envisage pas une solution algorithmique applicable à un grand nombre d'aires, mais elle peut donner des idées pour une telle solution.

Le problème des contraintes spatiales est aussi signalé par Chandon et Pinson (1981) qui suggèrent que, "dans le but d'adapter la classification aux particularités des données collectées, il est

souhaitable d'introduire des contraintes d'ordre pour les phénomènes se déroulant dans le temps ou des contraintes spatiales de contiguïté pour le classement d'entités géographiques"; un article concernant la classification sous contrainte spatiale par Monestiez (1977) est également signalé.

Gordon (1981) présente deux solutions algorithmiques qui vont dans le sens de notre démarche: l'un est proposé par Webster et Burrough (1972) et l'autre par Openshaw (1977).

L'algorithme de Webster et Burrough utilise une méthode de classification hiérarchique ascendante. A partir de n groupes à un seul membre on regroupe successivement les paires de groupes qui donnent la moindre croissance dans le critère de variabilité, tout en restant contigus.

L'algorithme de Openshaw (1977) utilise une méthode de réallocation qui tente d'améliorer une partition initiale en g groupes d'objets contigus. Successivement un objet peut être déplacé d'un groupe à un autre groupe contigu à condition que ce déplacement ne cause pas la décomposition du groupe de départ en plusieurs sous graphes non connectés. La réallocation continue jusqu'à ce que le critère de variabilité ne puisse plus être réduit par aucun des déplacements d'objets possibles.

3. La typologie sous contrainte spatiale

Pour explorer les effets de l'introduction de la contrainte spatiale dans un algorithme de classification hiérarchique, on a utilisé des données concernant les comportements d'achat de produits/services bancaires, concernant 651 îlots de la ville de Roubaix.

Le territoire d'opération d'une société qui distribue des produits ou des services peut être représenté par un espace euclidien bidimensionnel où les unités territoriales (région, quartier, îlot) apparaissent comme des polygones contigus ayant des positions et contours bien définis. Dans des cas bien précis, l'analyse typologique doit prendre en compte la position géographique de telles unités territoriales sous forme de contrainte spatiale.

Nous pouvons introduire la contrainte spatiale soit de manière *implicite* en ajoutant des variables caractérisant la position géographique aux variables du domaine de recherche⁶, soit de manière *explicite* sous forme de moindre distance entre deux unités territoriales, permettant de les regrouper, ou sous forme de contiguïté. Au niveau de la contiguïté, on peut aller plus loin et nuancer la contrainte en terme de longueur de la frontière commune entre deux territoires à regrouper.

Dans cette démarche, on explore deux manières d'introduire la contrainte spatiale dans l'analyse typologique, une implicite et l'autre explicite. La première est d'introduire les données géographiques (centroïdes des polygones voir îlots) parmi les variables qui caractérisent les observations et faire varier leur poids en fonction de l'importance qu'on veut donner à la distance géographique dans la classification; c'est une méthode implicite. La deuxième modalité est explicite et introduit une contrainte de contiguïté à chaque étape d'un algorithme de classification.

⁶ Nous appelons variables du domaine de recherche les variables qui caractérisent les observations autres que les variables représentant la position géographique.

3.1. Les coordonnées géographiques comme variable de classification (contrainte spatiale implicite)

Utiliser uniquement les coordonnées géographiques comme variables de classification donne lieu à une hiérarchie dans laquelle les unités territoriales se regroupent pour former des zones contiguës. Nous avons vérifié empiriquement cette hypothèse en appliquant plusieurs algorithmes de classification hiérarchique sur les coordonnées des centroïdes des 651 îlots de la ville de Roubaix. Les deux premières cartes de la page suivante montrent des partitions des îlots en 10 et 30 groupes. Les zones obtenues à différents niveaux d'agrégation (10 à 50 groupes) étaient contiguës et de surfaces relativement égales. Cela montre qu'il est possible d'utiliser les coordonnées géographiques pour introduire une contrainte spatiale et que l'on peut même interpréter comme une contrainte relaxée de contiguïté. L'intensité de cette contrainte spatiale peut être dosée par le poids accordé aux coordonnées géographiques dans la combinaison avec les autres variables qui caractérisent les données.

Concrètement, nous avons remplacé une quarantaine de variables spécifiques au domaine de recherche avec les composantes principales qui concentraient la plus grande partie de l'information totale et on les a combinées avec les coordonnées géographiques des observations pour obtenir le tableau suivant.

NO_ILOT	FACT1	FACT2	FACT3	CENTROIDX	CENTROIDY
1	-0,47499	-0,25794	-0,61751	-1,19354	1,67796
2	-0,92167	-0,22533	0,08028	-1,27180	1,80005
3	-0,88465	1,02480	0,01906	-1,20606	1,88353
4	-0,74383	-0,57268	-0,53490	-1,09753	1,89083
.....
650	-0,41742	-0,16408	0,00337	-1,22589	1,99832
651	-0,55015	-0,06364	-0,67750	-1,31354	1,94301

Table I. - Combinaisons de données résumant le comportement d'achat et la position géographique pour 651 îlots
FACT1, FACT2, FACT3 sont des variables résumant les comportements d'achat de produits bancaires; CENTROIDX et CENTROIDY sont les coordonnées sur la carte des centres de chaque îlot.

Les composantes étant extraites à partir de valeurs centrées et réduites, nous avons calculé les valeurs centrées et réduites des coordonnées géographiques. Pour éviter de déformer les distances géographiques, la réduction des coordonnées géographiques a été opérée en divisant chaque valeur au moindre des écarts types sur X et Y.

Le tableau combinant les variables du domaine de recherche avec les variables géographiques a été traité par un algorithme de classification hiérarchique. Les regroupements obtenus marquent une homogénéité géographique nettement supérieure à ceux qui tiennent compte uniquement des variables du domaine. Ces résultats sont illustrés à la page suivante pour le cas des partitions en 10 et 30 groupes d'îlots. Chaque intensité de gris sur la carte marque un groupe d'îlots qui se ressemblent.

Le fait de mélanger deux espaces qualitativement différents pose certainement un problème qui peut faire l'objet d'une analyse ultérieure. Les groupes (clusters) obtenus par l'analyse typologique sont définis le plus souvent par des points relativement contigus dans l'espace (Stewart

1981). On a suggéré qu'ils doivent se caractériser par cohésion interne et isolation externe (Cormack, 1971 et Everitt, 1974). La cohésion interne signifie que les membres d'un groupe doivent être similaires. L'isolation externe exige que les membres d'un groupe soient isolés par rapport aux membres d'un autre groupe par des espaces presque vides. Ces caractéristiques doivent être satisfaites pour les groupes obtenus dans chacun des deux espaces analysés. Le problème qui se pose est que les groupes formés dans l'espace des variables du domaine ne correspondent pas aux groupes formés dans l'espace géographique.

Dans l'essai de juxtaposer la hiérarchie dans l'espace des variables avec une hiérarchie dans l'espace géographique, une piste pourrait être explorée: le croisement de classifications⁷.

3.2. Algorithme de classification hiérarchique sous contrainte de contiguïté (contrainte spatiale explicite)

La tentative d'introduire une contrainte dans un algorithme de classification hiérarchique n'est pas nouvelle. Nous avons présenté les démarches de Webster et Burrough (1972) et Openshaw (1977). La méthode TAXMAP de Carmichael et Sneath (1969), présentée par Chandon et Pinson (1981), modifie la méthode du voisin le plus proche en fixant un seuil au-delà duquel les objets ne sont plus agrégés à un groupe existant, pour éviter la formation du chaînage spécifique à celle-ci. Nous proposons ici une méthode de façon à introduire la contrainte de contiguïté dans un algorithme de classification hiérarchique indépendamment de la méthode utilisée pour agréger les groupes (méthode du voisin le plus proche, du voisin plus éloigné, etc.) et un programme informatique qu'on a appelé TYPOGEO.

Dans un algorithme classique, les distances entre tous les objets deux à deux (d_{ij}) sont conservées dans un tableau à un seul indice, contenant la partie située en dessous de la diagonale principale de la matrice des distances (Roux, 1985). Le tableau contient les éléments de la matrice dans l'ordre des lignes comme illustré dans la figure 4.

On procède de manière itérative. A chaque itération, on réunit les deux objets les plus proches formant un groupe et on calcule les distances entre le groupe nouvellement créé et les autres objets. Le processus continue jusqu'à ce que tous les objets aient été réunis dans un seul groupe. Lorsque cela est achevé, on dresse un arbre hiérarchique dont les noeuds représentent les fusions successives.

En géo-marketing les observations ont une position géographique et occupent une place qui est délimitée par des lignes de frontière, ce qui signifie qu'entre deux observations (quartiers) il peut y avoir contiguïté ou non-contiguïté. Le tableau des données brutes, qui entre dans les calculs de classification, à part les valeurs des variables, doit indiquer pour chaque observation les observations contiguës. Les aspects techniques, que nous avons dû résoudre pour la reconnaissance des contiguïtés à partir de données cartographiques, sont présentés en annexe.

⁷ Pour plus de détails à ce sujet voir Diday E., 1983

Néanmoins, le simple tableau des distances, utilisé dans les méthodes classiques, n'est pas suffisant pour prendre en compte la contiguïté dans la construction de la hiérarchie. C'est la raison pour laquelle nous avons ajouté un tableau des contiguïtés qui ressemble au tableau des distances, mais, à la place des distances, il contient les valeurs zéro et un (un pour indiquer la contiguïté et zéro pour l'absence de contiguïté).

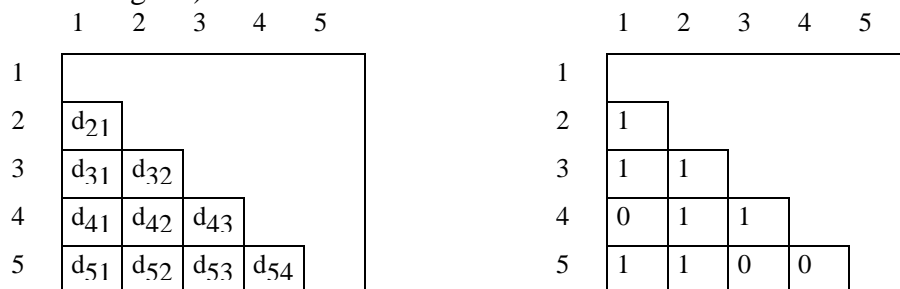


Fig.4 Organisation de la matrice des distances dans un tableau unidimensionnel et la matrice des contiguïtés

Dans le processus itératif de regroupement (fusion) des observations, on introduit la contrainte de contiguïté. A chaque itération, on réunit les deux objets *contigus* les plus proches formant un groupe. On calcule les distances et on enregistre les contiguïtés entre le groupe nouvellement créé et les autres objets. Comme dans l'algorithme classique, le processus continue jusqu'à ce que tous les objets aient été réunis dans un seul groupe. Lorsque cela est achevé, on dresse un arbre hiérarchique dont les noeuds représentent les fusions successives et leur niveau est donné par la moindre des distances recalculées à chaque itération.

Pour éviter un éventuel phénomène "d'inversion" comme celui illustré dans la figure 5, on doit s'assurer que les distances recalculées sont supérieures au niveau du noeud que l'on vient de former.



Fig.5 - Le phénomène d'inversion dans un arbre de classification hiérarchique

(a) est un arbre normal où le niveau (ou distance) du dernier noeud qui réunit l'objet 3 au groupe (1,2) est supérieur au niveau du noeud précédent ; (b) est un arbre où l'inversion c'est produit par ce que le niveau (distance) du dernier noeud est inférieur à celui du noeud précédent.

Si la moindre des distances recalculées est inférieure au niveau calculé du dernier noeud, il faut, pour pouvoir construire un dendrogramme intelligible, augmenter le niveau du noeud par rapport au niveau précédant (avec un nombre inversement proportionnel au nombre d'observations). Quand on impose une contrainte de contiguïté, ce phénomène de "rupture de monotonie" apparaît souvent. C'est la raison pour laquelle, durant le déroulement de l'algorithme, on enregistre le numéro des noeuds (regroupements) où une telle rupture intervient et, à la fin, on calcule aussi un indicateur qui indique la proportion des regroupements où il y a eu rupture de monotonie dans l'ensemble des regroupements.

Cet indicateur, en combinaison avec des mesures d'homogénéité (variance intra-groupe), peut servir pour fixer le nombre de groupes à retenir.

3.3. Programme informatique de classification sous contrainte de contiguïté

Le programme de classification sous contrainte de contiguïté proposé traite des fichiers dBase (.DBF) ayant sur la première colonne les noms des objets à classer, sur plusieurs colonnes les valeurs ou codes des variables, et, sur la dernière colonne, une chaîne de caractères contenant pour chaque objet les numéros des objets voisins (du point de vue géographique) séparés par des points virgules. Un exemple réduit d'un tel fichier et du déroulement des calculs est présenté dans la figure 6. Le fichier est représenté par un tableau contenant des données sur huit îlots. Chaque îlot est caractérisé par les variables Epar. (nombre de produits d'épargne possédés) et Prev. (nombre de produits de prévoyance possédés). A partir des valeurs de ces variables on calcule les distances entre les îlots (distances euclidiennes) et on obtient la matrice des distances, sur la quelle on applique en parallèle notre algorithme de regroupement sous contrainte de contiguïté et un algorithme témoin sans contrainte. La colonne Voisins, du fichier, est utilisée pour établir la matrice des contiguïtés. A chaque étape on regroupe les deux îlots ou groupes îlots les plus proches. La première étape montre que les deux îlots, qui se ressemblent le plus, sont 1 et 4. L'algorithme sans contrainte va les regrouper, en créant un premier groupe dans la hiérarchie. L'algorithme sous contrainte de contiguïté ne peut pas les regrouper, car illustration cartographique et la matrice des contiguïtés montre qu'ils ne sont pas contigus. Ce dernier algorithme va regrouper donc les deux îlots contigus les plus proches: 2 et 7. Le groupe nouvellement créé par exemple (2,7) prendra les positions de celui de ses membres qui est le plus haut placé dans la matrice des distances. Ici ce sont les positions de 2 (hachurés en gris clair). Les positions de l'autre membre du groupe, ici 7, disparaîtront. Ses positions (hachurées en gris foncé) seront prises par la dernière ligne de la matrice des distances. La matrice sera ainsi réduite d'une dimension. Dans cet exemple, les distances du groupe nouvellement créé par rapport aux autres observations ou groupes sont calculés par la méthode du voisin le plus éloigné⁸, ce qui correspond aux plus grandes distances séparant ce groupe par rapport aux autres. Concrètement la distance du groupe nouvellement créé (2,7) par rapport à un autre objet par exemple 1, est égale au maximum des distances de chacun des deux membres (2 et 7) par rapport à l'objet 1, donc $d_{(2,7),1} = \max(d_{2,1}; d_{7,1}) = 7,07$.

Au bout de six étapes tous les objets sont classés, et la hiérarchie est construite. A la fin du processus nous retenons trois groupes de régions représentées par des intensités de gris différentes. Les régions obtenues par notre algorithme sous contrainte sont contiguës, et celles obtenues par l'algorithme témoin ne le sont pas.

4. Une application dans le domaine du marketing bancaire

Nous avons appliqué l'algorithme de classification sous contrainte de contiguïté à des données concernant le marché bancaire. Une observation récente des évolutions dans le domaine

⁸ le programme intègre aussi les critères du plus proche voisin et du lien moyen

bancaire⁹ sur le marché américain montre que la structure des filiales d'une banque est devenue un élément prioritaire dans la politique de contrôle des coûts. La consolidation sur le marché est devenue essentielle pour les banques qui ont tendance à concentrer leurs efforts sur les parts de marché existants ou à pénétrer dans une zone géographique contiguë en assurant une présence immédiate et significative. Les banques sont de plus en plus orientées clients et font des analyses démographiques et causales pour déterminer le profil du client ayant la plus grande valeur. Cette analyse peut servir à orienter l'expansion dans le territoire, à la constitution ou à l'acquisition de certaines filiales et à l'abandon de certaines.

4.1. Les données

Les données qu'on a utilisées pour tester notre algorithme concernent la détention de certains produits d'épargne par les clients d'une banque dans la ville de Roubaix. Les données individuelles ont été agrégées au niveau des îlots et ainsi on a obtenu un tableau avec 651 observations (individus) représentant des îlots. Chaque îlot est caractérisé par une quarantaine de variables exprimant essentiellement des taux de détention des différents produits proposés par la banque. Pour éliminer les corrélations entre les variables et réduire les dimensions de l'analyse, on a effectué une analyse factorielle en composantes principales et on a retenu trois dimensions sur lesquelles on a essayé d'appliquer à la fois l'algorithme classification sans contrainte et l'algorithme équivalent sous contrainte de contiguïté pour comparer les résultats et évaluer la perte d'homogénéité des groupes obtenus sous contrainte de contiguïté.

Dans l'analyse sous contrainte de contiguïté, on utilise une colonne finale appelée "Voisins"; elle donne au programme de classification l'information nécessaire pour établir la matrice des contiguïtés.

ILOT	FACT1	FACT2	FACT3	VOISINS
1	-0,47499	-0,25794	-0,61751	2;5;637;640;641;643;
2	-0,92167	-0,22533	0,08028	3;4;651;
3	-0,88465	1,02480	0,01906	2;4;650;651;
...
.....				
651	-0,55015	-0,06364	-0,67750	2;3;4;649;650;

Table II. - Données résumant le comportement d'achat et la contiguïté pour 651 îlots

Les mêmes données ont été traitées à la fois par un algorithme classique qui ne tient pas compte des contiguïtés et par l'algorithme que nous avons proposé pour le regroupement régional sous contrainte de contiguïté.

Lors de l'avancement dans la construction de la hiérarchie, le programme classification sous contrainte de contiguïté signale les noeuds où cette contrainte provoque une rupture de monotonie.

⁹ "Slight increase in branch sales", United Sales Banker (USI), vol.103, jan.1993, p.13-18

4.2. Résultats et Evaluation

L'application parallèle sur les mêmes données de l'algorithme de classification sous contrainte de contiguïté et d'un algorithme témoin sans contrainte, donne lieu à deux arbres de classification (dendogrammes) différents et à des partitions différentes pour le même nombre de groupes (clusters).

L'objectif fixé étant d'explorer des modalités pour prendre en compte la contrainte spatiale et proposer un algorithme de classification sous contrainte de contiguïté, la comparaison des résultats obtenus par notre algorithme par rapport à l'algorithme témoin nous a semblé nécessaire.

Le jugement concernant la qualité des résultats doit porter sur l'homogénéité des groupes obtenus, uniquement dans l'espace des variables du domaine de recherche, et non pas sur les variables géographiques. Evidemment, les classes obtenues en imposant une contrainte de contiguïté sont moins homogènes que celles qui résultent de l'algorithme témoin sans contrainte. La contrainte de contiguïté géographique, imposée pour des raisons managériales, entraîne des ruptures en terme d'homogénéité dans l'espace des variables caractérisant le comportement d'achat.

Nous avons pris comme mesure de l'homogénéité des partitions la trace de la matrice des covariances intra-groupe¹⁰ utilisant la procédure d'Analyse Discriminante du programme SPSS. La figure ci-dessus présente la mesure de la qualité du regroupement en classes sous contrainte de contiguïté et sans contrainte pour des nombres différents de groupes (clusters) retenus.

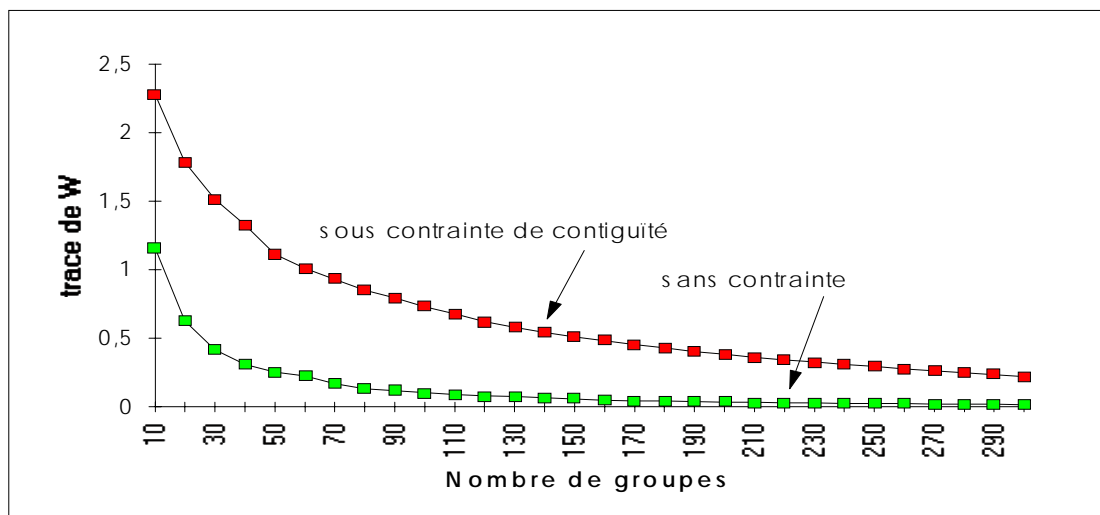


Fig.7 Mesures comparatives de l'homogénéité des groupes (trace de la matrice des covariances intragroupes W)

A partir de cette comparaison, un critère de choix possible du nombre de classes à retenir, quand on impose une contrainte géographique, serait d'atteindre un niveau d'homogénéité proche de celui qu'on obtient dans le cas d'un regroupement sans contrainte. L'analyse des résultats comparatifs illustrés dans le graphique ci-dessus montre que les écarts dans la mesure d'homogénéité déterminés

¹⁰ La trace de cette matrice appelé **W** dans la littérature, donne la somme des variances intra-groupes. Arnold (1979) montre qu'elle est utilisée comme mesure d'hétérogénéité des groupes qui doit être minimisée dans l'algorithme de Howard-Harris (utilisé par Green et Rao 1972; Hustad et Pessemier 1973; et d'autres).

par la contrainte de contiguïté sont variables mais restent importants lorsqu'on fait varier le nombre de groupes (clusters).

Une comparaison relative sous forme de rapport des deux mesures d'homogénéité montre que plus le niveau d'agrégation est important (moins le nombre de groupes) plus les indices d'homogénéité deviennent comparables. On constate aussi une certaine corrélation entre la fréquence des ruptures de monotonie et le rapport des mesures d'homogénéité sous contrainte et sans contrainte.

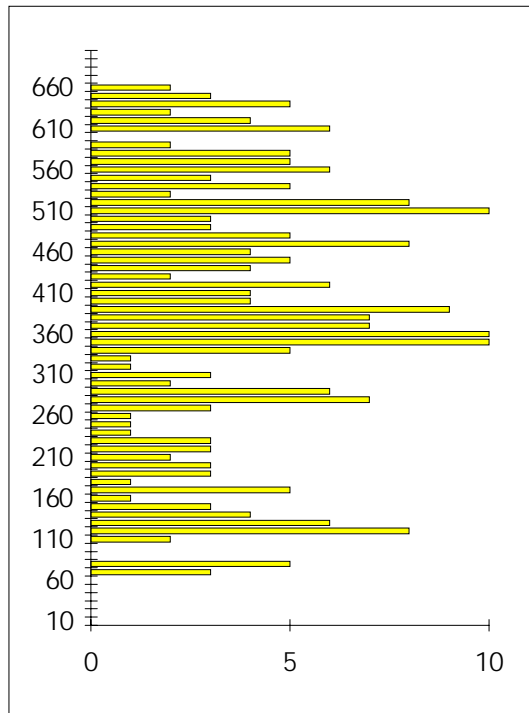
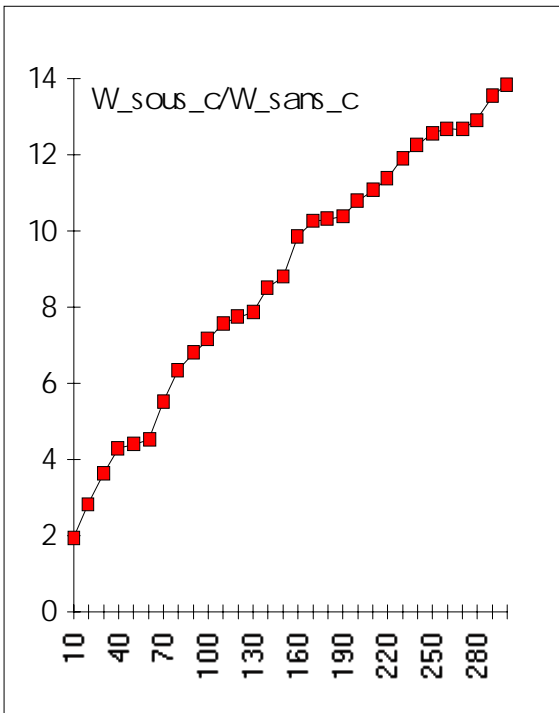


Fig.8 Rapport entre la variance intragroupe sous contrainte de contiguïté et sans contrainte

Fig.9 Fréquence des ruptures de monotonie

Le nombre de groupes à retenir est une décision qui dépend à la fois des contraintes managériales et de la qualité statistique des groupes obtenus. Une discussion détaillée des problèmes liés à la validité d'une classification et des critères de choix du nombre de groupes se trouve dans Chandon et Pinson (1980), Caumon et Chandon (1989) et Devaussuzenet (1994).

L'utilité pratique des algorithmes et solutions proposés est illustrée par les résultats cartographiques présentés à la page suivante. Sur la carte de la ville de Roubaix, composée de 651 îlots, on constate que les groupes obtenus sous contraintes spatiales se reflètent par des zones contiguës et nettement plus compactes que les groupes obtenus par l'algorithme témoin. L'algorithme de classification sous contrainte de contiguïté semble bien avoir isolé les quartiers résidentiels, les zones administratives, commerciales, et les quartiers défavorisés de la ville.

5. Discussion et voies de recherche

L'objectif de cette recherche était d'élaborer un algorithme et de mettre en place une méthode informatisée qui permette d'effectuer des typologies des régions géographiques en marketing. A ce jour, l'algorithme a été testé dans un nombre réduit de situations réelles ou simulées sur ordinateur. Le test, dans plusieurs situations différentes, pourrait apporter plus d'information sur son fonctionnement.

L'intérêt de cette approche est de pouvoir dresser des cartes typologiques qui regroupent des entités géographiques avec des niveaux de détail variables, en fonction des besoins des analystes en marketing.

Le fait d'introduire une contrainte de contiguïté dans le processus d'agrégation hiérarchique entraîne une certaine perte d'homogénéité au niveau des groupements, mais cette contrainte est souvent imposée par la spécificité des zones géographiques et par les intérêts managériaux.

La légitimité d'une telle approche devra être confirmée par des applications futures. Cependant, les caractéristiques suivantes de l'analyse typologique nous laissent croire à cette légitimité:

- l'analyse typologique est une technique purement empirique et donc inductive (Gerard 57), qui, par rapport à d'autres techniques de classification comme l'analyse discriminante et l'AID, ne fait aucune présomption concernant les différences importantes dans la population (Punj et Stewart, 1983);

- il n'y a pas de solution analytique permettant de décider combien de groupes doivent être identifiés, quoique l'on puisse déterminer à quelle étape une nouvelle réduction du nombre de groupes augmente de façon minimale la somme des carrés des distances à l'intérieur d'un groupe (Roux, 1985);

- il n'y a pas des normes uniques concernant la délimitation des groupes ou pour décider si une observation doit être attribuée à un groupe ou à un autre;

- Cattell (1978) suggère que les groupes (clusters) sont des construits "vagues" (fuzzy).

Toutes ces éléments et bien d'autres, comme les raisons managériales énoncées dans la première partie de cet article, nous font croire à l'utilité des démarches d'analyse typologique sous contrainte spatiale. Tester les méthodes existantes dans des situations différentes, trouver d'autres méthodes nous semble nécessaire. Cela suppose aussi des efforts pour affiner et adapter les mesures de distance et même éventuellement nuancer la notion de contiguïté par des paramètres comme la longueur de la frontière commune.

Une autre direction de recherche complémentaire à cette démarche serait d'explorer l'importance sectorielle de la contrainte spatiale: examiner les différences en terme d'intensité de la contrainte spatiale pour des produits/services de proximité par rapport d'autres produits. L'observation longitudinale des implantations d'agences, de la diffusion territoriale des réseaux de distributions des services nous permettra de calculer des coefficients d'intensité de la contrainte spatiale.

Références

- *** , Analyse de Système en Géographie, textes réunis par Yves Guermond, Presses Universitaires de Lyon, 1984
- Arnold J. S. (1979), A Test for Clusters, Journal of Marketing Research , Vol. XVI , 545-51.
- Berry B. J. L. (1961), A Method for Deriving Uniform Regions, Przegląd Geograiczny, 33, 263-282.
- Berry, B. J. L. (1971), Géographie des marchés et du commerce en détail, Paris, Armand Colin
- Berry, B. J. L. (1967), Grouping and Regionalizing: an approach to the problem using multivariate analysis, dans GARRISON, W. L. et MARBLE, D. F., edit., Quantitative Geography, part I. Evanston, Studies in Geography n° 13, Northwestern University , p.219-251.
- Bon, J et Gregory P. (1986), Techniques marketing, Paris, Vuibert.
- Cassetti E. (1964), Classificatory and Regional Analysis by Discriminant Iterations, Computer Application in the Earth Science Project, Tchn. Rept., n° 12, Dept. Geogr., Northwestern University.
- Caumont D. et Chandon J.-L. (1989), Quelques problèmes liés à la validité d'une classification, Recherche et Applications Marketing, 3, 78-92
- Celex G., Diday E., Govaert G., Lechevalier Y., Ralambondrainy H. (1989), Classification automatique des données. Environnement statistique et informatique, Paris, Dunod
- Diday E. (1983), Croisement, ordres et ultramétriques, Mathématiques et Sciences Humaines, 83, 31-54.
- Devaussuzenet E. (1994), Méthodes Typologiques - La détermination du nombre de groupes, Mémoire mineur de DEA, IAE de Lille
- Gordon (1973), Classification in the presence of constraints, Biometrics, 29, 821-827.
- Gordon (1981), Classification, London, Chapman and Hall.
- Hagget P. et Chorley R. (1972), Network Analysis in Geography, London, Edward Arnold.
- Hagget P. (1973), l'Analyse Spatiale en géographie humaine, Paris , Armand Colin.
- Hagood M. J. (1943), Statistical Methods for Delineation of Regions Applied to Data on Agriculture and Population, Social Forces, 21, 288-297.
- Hagood M. J. et Price D. O. (1952), Statistics for Sociologists, Nex York.
- Johnson S. C. (1967), Hierarchical Clustering Schemes, Psychometrika, 32, 241-254
- McQuitty L.L. (1967), A Mutual Développement of Some Typological Theories and Pattern-Analitic Methods, Educational and Psychological Measurement, 17, 21-46
- Monestiez P. (1977), Méthode de classification automatique sous contrainte spatiale, Statistique et Analyse des Données, 3, 75-84.
- Nordbeck S. (1962), Location of Areal Data for Computer Processing, Lund Studies in Geography, Series C, General and Mathematical, 2.
- Lance G.N., et W.T. Williams (1967), A General Theory of Classificatory Sorting Strategies. II. Clustering Systems, The Computer Journal, 10, 271-276.

- Punj G., et Stewart D.W. (1983), Cluster Analysis in Marketing Research: Review and Suggestions for Application, Journal of Marketing Research, Vol.XX, 134-148
- Oppenshaw S. (1977), A geographical solution to scale and aggregation problems in region building, partitioning and spatial modelling, Trans. Inst. Brit. Geog., NS2, 459-472.
- Racine J.-B. et Reymond H. (1973), L'analyse quantitative en géographie, Paris, Presse Universitaires de France
- Rogers D. F. (1988), Algorithmes pour l'infographie, Paris, McGraw-Hill.
- Roux M. (1985), Algorithmes de Classification, Paris, Masson.
- Sneath, P.H. et Sokal R.R. (1973), Numerical Taxonomy, San Francisco: W.H. Freeman.
- Webster R. et Burrough P.A. (1972 a), Computer based soil mapping of small areas from sample data. I. Multivariate classification and ordination, Journal of Soil Science, 23, 210-221.
- Webster R. et Burrough P.A. (1972 a), Computer based soil mapping of small areas from sample data. II. Classification Smoothing, Journal of Soil Science, 23, 222-234.
- Zobler L. (1958), Decision Making in Regional Construction, Annals of the Association of American Geographers.

ANNEXE - Aspects techniques de la reconnaissance de la contiguïté à partir de données cartographiques

Les données géographiques concernant des aires (région, quartier, etc.) sont stockées dans des fichiers contenant les coordonnées (x, y) des sommets qui forment des polygones¹¹.

Concrètement, si on regarde la carte imaginaire présentée ci-dessous représentant vingt régions, le contenu du fichier des coordonnées géographiques obtenues par l'intermédiaire de notre programme (MCARTE) est représenté par le tableau à gauche de la carte. Pour aboutir à un tableau indiquant les contiguïtés de chaque aire (voir tableau au-dessus de la carte), on doit trouver un algorithme de reconnaissance des contiguïtés à partir des coordonnées des sommets des polygones.

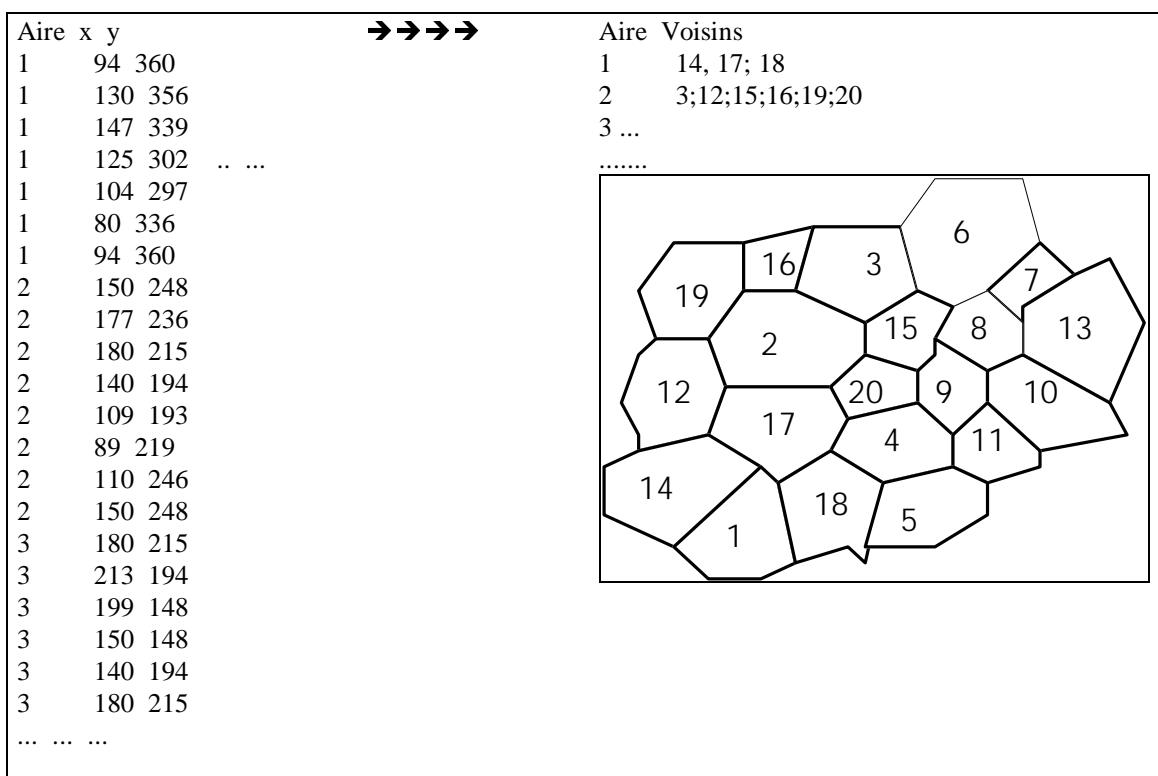


Fig.7 Le problème de la détection des contiguïtés

6.1. L'organisation des données cartographiques

Le programme de reconnaissance des contiguïtés entre régions que nous avons développé traite les fichiers ASCII du type MIF de MAPINFO¹² pour identifier les contiguïtés (intersections) entre les régions d'une carte. Un exemple illustre la structure d'un fichier MIF:

¹¹ Selon une convention largement utilisé en cartographie (Nordbeck, 1962), les aires sont amenées, par approximation à des polygones, dont les sommets sont numérotés de P₁, P₂, ... P_n, dans un sens inverse de celui des aiguilles d'une montre, P_n étant égal à P₁. P₁ a la plus faible valeur de la coordonnée y, et si cette valeur n'est pas unique, P₁ a la plus faible valeur de la coordonnée x. L'aire correspondant à chaque polygone est alors définie comme étant la surface située à gauche de toutes les lignes qui relient deux sommets consécutifs.

¹²MapInfo © est une marque déposée de Mapping Information Systems Corporation, Troy, New York, c'est le système de traitement des cartes qu'on a utilisé pour enregistrer les îlots de la ville de Roubaix.

**Exemple de fichier ASCII permettant de lire des données cartographiques
enregistrées par le système MAPINFO:**

<pre>Version 2 Delimiter "," Index 1 CoordSys Earth Projection 1, 62 Columns 4 CODE Char(2) NAME Char(17) POP_89 Decimal(10, 2) SUP_KM2 Decimal(10, 2) Data</pre>	<pre>Entête d'un fichier indiquant les séparateurs utilisés le nombre de fichiers index le système de coordonnées géographiques utilisées le nombre et la structure des champs de données Début des coordonnées des régions (une région peut contenir un ou plusieurs polygones)</pre>
<pre>Region 1 93 19.395638 41.910659 19.439318 42.034053 19.450602 42.105763 19.634621 41.807775 19.395638 41.910659 Pen (1,2,0) Brush (2,16777215,16777215) Center 20.208275 41.177455</pre>	<pre>Première Région avec 1 Polygone le polygone à 93 sommets (points) coordonnées x, y du 1-er point coordonnées x, y du 2-ième point coordonnées x, y du 93-ième point Type de contour ou bordure du polygone (la plume..) Type de peinture à l'intérieur de polygone (le pinceau) Coordonnées x, y du centroïde du polygone</pre>
<pre>Region 1 16 1.449677 42.634486 1.460548 42.644417 1.45612 42.589596 1.449677 42.634486</pre>	<pre>Deuxième Région avec 1 Polygone le polygone à 16 sommets (points) coordonnées x, y du 1-er point coordonnées x, y du 2-ième point coordonnées x, y du 16-ième point</pre>

Fig.8 La structure d'un fichier cartographique (.MIF)

6.2. Lecture et organisation interne des coordonnées cartographiques

Le programme commence par une routine pilot (driver) qui lit le fichier ASCII et le traduit dans un format interne facilement exploitable.

Les étapes de la routine de conversion consistent à:

- 1) détecter chaque apparition du mot "Region" qui annonce le début des coordonnées de carte d'une nouvelle région par l'intermédiaire de la fonction qu'on a appelée FindNextPoly();
- 2) lire le nombre entier qui est séparé par un caractère blanc du mot Région. Ce nombre indique le nombre de polygones qui forment la Région;
- 3) alimenter dans une boucle TANT QUE une structure interne de mémoire de type Région, avec les conversions numériques des coordonnées des polygones lues dans le fichier ASCII.

La structure TRegion (type région) assure un format interne numérique adapté aux calculs nécessaires pour l'identification des contigüités. Elle est composée d'une valeur entière indiquant le nombre de Points du polygone, suivie par une structure de type rectangle (TRectangle) dans laquelle

on enregistre les coordonnées qui encadrent le polygone, puis par un pointeur sur la série des points (sommets) qui forment le polygone.

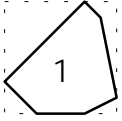
	<pre>type TRegion {int NoPoints; TRectangle Rectangle; TPoint *Points;}</pre>	<pre>type TRectangle {double gauche, haut, droite, bas}</pre>	<pre>type TPoint {double x; double y;}</pre>
---	---	--	---

Fig.9 Organisation interne des données cartographiques sur une Région

Dans l'exemple ci-dessus la structure TRegion de la Région 1 contient:

NoPoints = 7

Rectangle = {80, 297, 147, 360} les coordonnées extrêmes (gauche, haut, droite, bas).

*Points = pointeur sur la série de 7 sommets --> {94,360} {130,356} {147,339} {125,302} {104,297} {80,336} {94,360}.

6.3. Détection des contiguïtés

Les voisins de chaque région sont détectés dans une boucle qui analyse chaque couple de polygones et qui vérifie s'ils s'intersectent ou pas par l'intermédiaire d'une fonction Intersect qu'on a construit.

La fonction Intersect prend comme arguments deux polygones et retourne zéro (pas contigu) si les rectangles qui encadrent les polygones ne se touchent pas. Si les rectangles se touchent la fonction vérifie si il y a des segments qui se touchent et retourne un (contigu) si c'est vrai et zéro (pas contigu) si non.

La condition nécessaire pour tester le fait que deux rectangles ou segments ne se touchent pas est qu'il y ait point minimum supérieur à un point maximum de l'autre sur l'axe horizontal ou vertical.

L'existence de la structure Rectangle dans la structure Region permet d'éviter des longues opérations de comparaison de chaque segment d'un polygone avec tout les segments d'un autre polygone pour voir si ces polygones s'intersectent. Dans l'exemple ci-dessous on s'aperçoit que, dans le cas du polygone 8, le fait que le rectangle qui l'encadre ne touche pas les rectangles qui encadrent les polygones 1, 17 etc. permet de reconnaître que 8 n'est pas contigu avec 1 et 17; par contre, quand les deux rectangles sont intersectés pour vérifier que les polygones sont contigus, on doit parcourir dans une boucle toutes les paires de segments jusqu'au moment ou l'intersection de deux segments est détectée.

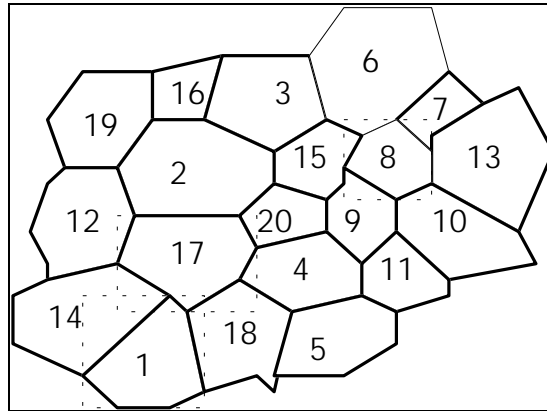


Fig.10 Accélération de la détection des contiguïtés par l'intermédiaire des rectangles

Il est possible d'améliorer encore l'algorithme de reconnaissance de la contiguïté de deux régions. En infographie, il y a beaucoup de développements dans ce sens qui méritent d'être explorés¹³.

¹³ Pour plus d'informations à ce sujet voir D. F. Rogers (1988)