

CLASSIFICATION ET ANALYSE DE CONTIGUÏTE

Ludovic Lebart

CNRS - Ecole Nationale Supérieure des Télécommunications
46 rue Barrault, 75013, Paris, France.
lebart@enst.fr

Lorsque des observations statistiques (multivariées) sont associées à un graphe (séries temporelles, données géographiques), les variances et covariances locales permettent de prendre en compte la dépendance des observations vis-à-vis de la structure de graphe. L'analyse de contiguïté permet alors de confronter structures locales et globales. Si le graphe est associé à une partition, on retrouve l'analyse linéaire discriminante de Fisher. On étudie ci-dessous le cas où le graphe n'est pas une donnée externe, mais est construit à partir des données elles-mêmes, à partir des plus proches voisins de chaque observation. Cette étude permet de mettre en évidence des zones d'inégales densité, et des structures intermédiaires entre celles que détectent les méthodes factorielles, et celles mises en évidence par les méthodes de classification. L'idée de trouver une métrique susceptible de mettre en évidence des classes remonte aux travaux de Art *et al.* [ART82] et Gnanadesikan *et al.* [GNA82]. Nous présentons ici la contribution de l'analyse de contiguïté à de telles approches.

Considérons n sommets d'un graphe symétrique G dont la matrice associée est M ($m_{ii} = 1$ si les sommets i et i' sont joints par une arête $m_{ii'} = 0$ sinon). Ces sommets sont simultanément décrits par p variables, (x_{ij}) est la valeur de la variable j pour le sommet i). Une telle situation se présente quand les sommets représentent des instants, des zones géographiques. L'analyse de contiguïté utilise simultanément une matrice des covariances locales C et une matrice des covariances globale V .