

STATISTIQUE ET ANALYSE DES DONNÉES

J. M. HUBAC

R. TOMASSONE

**Colloque de taxonomie numérique, 5-6 septembre 1977,
université Paris-sud, 91405 Orsay**

Statistique et analyse des données, tome 2, n° 3 (1977), p. 10-14

http://www.numdam.org/item?id=SAD_1977__2_3_10_0

© Association pour la statistique et ses utilisations, 1977, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

3 - 1977

colloque de taxonomie numérique**5-6 septembre 1977****université paris-sud , 91405 orsay**o
o o

Prévu pour être la onzième conférence internationale de classification, le colloque d'Orsay n'a pu attirer nos collègues d'outre-atlantique ; les raisons évidentes sont bien entendu d'ordre financier. Aussi, ce colloque n'a pas pu prendre la suite de ceux de St Andrews, de Lisbonne ou de Montréal pour ne citer que les derniers. Fort heureusement, à la même époque, l'I.R.I.A., organisait à Versailles les premières journées d'Analyse des données et d'informatique, dont une part importante était consacrée aux problèmes de classification automatique. Il a donc été facile de regrouper les participants à ces deux réunions lors des discussions particulièrement animées ; dans ce numéro spécial du bulletin de l'Association des Statisticiens Universitaires, on trouvera la plupart des textes du colloque d'Orsay ainsi que la liste des communications traitant de classification lors de journées I.R.I.A.

o
o o

Il est difficile de faire un bilan rapide et exact des interventions dans un domaine aux limites quelquefois floues ; ainsi, les méthodes d'analyse factorielle sont pour certains des outils de classification alors que ceux qui donnent à la classification son sens strict les excluent. De même, les techniques d'identification sont beaucoup plus proches des méthodes de classement ; néanmoins elles sont rarement étudiées par les théoriciens, de ce fait elles risquent fort d'être totalement délaissées. Il nous semble donc intéressant de porter à la connaissance de tous ceux qui s'occupent de classification qu'il existe un domaine largement ouvert aux idées nouvelles.

S'il est difficile de faire un bilan, il est par contre facile de dégager quelques impressions ; elles sont de plusieurs types :

. La préoccupation des anglo-saxons paraît actuellement se concentrer sur les problèmes relatifs aux tests de signification (cf. SNEATH P.H.A. & BOCK H.H.) ; on assiste donc à des retrouvailles entre la statistique inférentielle et le domaine qui en était le plus éloigné en analyse des données.

. L'importance aux yeux des français est celle du traitement de grands corpus de données ; ce problème préoccupe aussi bien les théoriciens que ceux qui ne font qu'utiliser des algorithmes mis au point par d'autres.

. La volonté de développer la théorie en fonction des demandes, voire des exigences des utilisateurs ; des notions de contiguïté ou de recouvrement, qui correspondent à une analyse fine des faits, sont souvent prises en compte. Dans le cas de la contiguïté c'est une information a priori qui est introduite ; ici, l'analyse des données ne se réduit pas au rôle purement descriptif dans lequel on la cantonne bien souvent.

. La diversité des champs d'application : elle va de l'agronomie à la linguistique en passant par l'architecture, la géographie et l'écologie. L'informatique elle-même fait un usage important des méthodes de classification pour traiter les propres observations qu'elle génère, par exemple pour l'évaluation d'un système informatique. Cette richesse des applications est le plus sûr garant de la vitalité d'un domaine où les théories n'ont souvent qu'un intérêt secondaire par rapport aux applications réalisées ou réalisables. De plus, on peut penser que la diversité des applications est une source constante d'enrichissement collectif, puisque chaque domaine utilisant la classification apporte ses réflexions et ses exigences propres.

o

o o

Naturellement, ces impressions sont celles des organisateurs du colloque ; elles traduisent donc leurs convictions profondes de chercheurs aux limites de la théorie et de l'application. De plus, elles représentent une vision instantanée d'un domaine en perpétuelle évolution, dans ce sens elles sont donc partiales. Mais alors, ne faudrait-il pas un bilan clair

et précis, pouvant être lu par tous, théoriciens et praticiens ? Un bilan dont le besoin est évident et nécessaire, mais qui serait une suite à celui précieux que CORMACK (1971) nous a déjà fourni ? C'est sans nul doute le souhait que nous pourrions faire et la proposition que nous pourrions lancer aux membres de la Société Française de Classification "qui vient de voir le jour".

Enfin, nous ne saurions terminer cette introduction sans remercier le comité de rédaction du bulletin de l'A.S.U. qui a accepté de publier l'ensemble des communications de ce colloque.

J.M. HUBAC

Université de Paris Sud-Centre d'Orsay
Laboratoire de Taxonomie végétale
expérimentale et numérique - Bât 362
F91405 ORSAY

R. TOMASSONE

Laboratoire de Biométrie
C.N.R.Z.
F78350 JOUY EN JOSAS

Référence

CORMACK R.M. (1971) A review of Classification. J.R. Statist. Soc. A, 134, 321-359.

Titres et Auteurs des Communications Présentées aux Premières
Journées Internationales d'Analyse des Données et d'Informatique
et classées sous le thème CLASSIFICATION

Cluster significance tests and their relation to measures of overlap	
P.H.A. SNEATH	
Automatic classification and piecewise approximation of data	
A. DOROFYUK (<i>text not yet received</i>)	
La méthode des pôles d'attraction	
I.C. LERMAN, H. LEREDDE	
Analyse des préférences du jury PAN (Programme Architecture Nouvelle)	
J.C. BERGES, E. JACQUET-LAGREZE	
Agrégation typologique de quasi-ordres : un nouvel algorithme	
J.L. CHANDON, J. LEMAIRE	
Sur la scalabilité d'une correspondance	
B. MONJARDET	
Classification automatique d'un grand ensemble de données par la méthode des «graphes réductibles» - application à l'analyse de données géographiques	
M. BRUYNOOGHE	
Construction de sous-tableaux homogènes dans un tableau de données binaires	
B. LEFEBVRE	
Classification d'homogénéité maximum	
M. DELATTRE, P. HANSEN	
Une méthode de classification simultanée des lignes et colonnes d'un tableau	
J. TOLEDANO, J. BROUSSE	
Reconnaissance de la structure et adaptation des programmes au milieu paginé à l'aide d'algorithmes de classification hiérarchique	
M.S. ACHARD, J.Y. BABONNEAU, G. LECHEVALIER, G. MORRISSET, B. MOUNAJJED	
Classification et reconnaissance	
M. TERRENOIRE, D. TOUNISSOUX	

Caractérisation et élaboration de la charge représentative d'un système d'exploitation par des méthodes d'analyse de données	
L. BOI, P. CROS, R. MARTIN, D. ROCHET, P. BOURRET	
Méthode de restructuration automatique du moniteur non résident de SIRIS 8	
H. LOCU, P. MICHEL	
Classification avec recouvrement : morphologie verbale en ancien français	
C. JACOB, S. MONSONEGO, R.C. TOMASSONE	
Une méthode d'organisation des données fondée sur la classification automatique	
H. BRIAND	
Application de la théorie des catégories en statistique descriptive	
G. DROUET D'AUBIGNY	
Reconnaissance des formes et mémorisation condensée de courbes et contours	
C. CHARLES	
Clusters as modes	
J.A. HARTIGAN	
On tests concerning the existence of a classification	
H.H. BOCK	
Data structures and their mathematical representation	
A. BELLACICCO	
Algorithme de classification d'un tableau de contingence	
G. GOVAERT	
Typologies en architecture	
J.P. MAROY, J.P. PENEAU	
Etude de traces de références de programmes : analyse de processus «à régimes»	
A. SCHROËDER	
Description de la charge soumise à un système informatique et évaluation des performances obtenues	
J. BIONDI	
L'utilisation de la classification automatique pour la conception logique d'une base de données	
A. FLORY, B. MOULIN	

N.B. Le texte de l'ensemble des communications peut-être obtenu auprès
 de l'I.T.R.I.A., Domaine de Voluceau, Nogencycourt BP 105,
 78150 LE CHESNAY