

STATISTIQUE ET ANALYSE DES DONNÉES

SOCIÉTÉ FRANÇAISE DE CLASSIFICATION

Résumés - Journées de Statistique, Paris, mai 1979

Statistique et analyse des données, tome 4, n° 3 (1979), p. 79-84

http://www.numdam.org/item?id=SAD_1979__4_3_79_0

© Association pour la statistique et ses utilisations, 1979, 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/>

SOCIETE FRANÇAISE
DE CLASSIFICATION

(Résumés - Journées de Statistique, Paris, MAI 1979)

ETUDE DISTRIBUTIONNELLE DE QUELQUES INDICES DE
SIMILARITE POUR DES DONNEES BINAIRES

UN MODELE DE MO

DE

Auteur : H. BACELAR NICOLAU
Centre de Statistique et Applications (C.E.A.- I.N.I.C.)
Faculté des Sciences de Lisbonne
58, rua da Escola Politécnica
Lisboa - PORTUGAL

Auteur : B. CRAIGNOU
IRIA-Laboria
Rocquencourt
B.P. 105
78150 LE CHESNA
Tél 9549020 pos

Dans ce travail nous faisons une étude comparative du point de vue distributionnel des coefficients le plus couramment utilisés pour les données binaires.

Nous démontrons que, sous une certaine hypothèse que l'on précise, ils se partagent en deux classes : l'une contenant ceux qui ont la même loi exacte (hypergéométrique), l'autre rassemblant les coefficients qui ne sont équivalents qu'en loi limite (normale).

Nous proposons alors comme coefficient général de comparaison pour ce type de données un coefficient probabiliste basé sur la valeur de la fonction de répartition normale limite commune à tous les coefficients. Ce coefficient généralise la notion de similarité selon la "vraisemblance du Lien" introduite par I.C. Lerman.

Modélisation d'
de morphogénèse décrit l
dès son initialisation.
phologiques du modèle.

La première par
croissance de chacune de
tion, grossissement, div

La seconde part
(rangement, lissage).

Le mouvement de
ce de leur grossissement
roulement de l'algorithme
tinu de l'algorithme des
tion de ce mouvement et
forme des cellules.

UNE NOUVELLE CLASSE DE METHODES D'AGREGATION EN CLASSIFICATION
AUTOMATIQUE BASEE SUR LA VRAISSEMBLANCE DU LIEN.

CLASSIFICATION

Auteur : F. COSTA NICOLAU
Centre de Statistique et Applications (C.E.A.-I.N.I.C.)
Faculté des Sciences de Lisbonne
58, Rua da Escola Politécnica
Lisboa - Portugal

Auteur : M. GONDRAN M
EDF, Directi
1, Avenue du
92141 CLAMAR

Nous proposons l'exploitation systématique de la notion de Vraisemblance du Lien (I.C.Lerman - méthode A.V.L., basée sur le lien maximal entre deux classes), et dont des résultats récents montrent l'aspect global, unificateur et sans biais pour le cas des données binaires (H. Bacelar Nicolau).

Nous y traitons surtout la méthode AVD (agrégation par la Vraisemblance du Lien minimal entre deux classes), les méthodes AMVLD (agrégation par les moyennes - arithmétique, géométrique, harmonique, et cas limites- des Vraisemblances des Liens maximal et minimal), et la méthode AVM (agrégation par la Vraisemblance du Lien moyen entre deux classes) .

On présente i
basée sur des données

Cette classif
chaîne) et optimale. E
des graphes : la liais
symétrique donnée par

Dans le cas ou
"tout ou rien", les cla
hiérarchique correspon
graphe des similarités.

VALEURS PROPRES ET VECTEURS PROPRES

EN ANALYSE DES PREFERENCES

Auteur : M. GONDRAN M.
EDF, Direction des Etudes et Recherches
1, Avenue du Général de Gaulle
92141 CLAMART CEDEX

Trois méthodes d'analyse des préférences sont présentées. Elles consistent à déterminer les valeurs propres et les vecteurs propres de la matrice des préférences par couple dans une structure algébrique bien choisie : les dioïdes. Si le dioïde est $(\mathbb{R}^+, +, \times)$ on trouve un ordre moyen ; si le dioïde est $(\mathbb{R}^+, \max, \times)$ on trouve la classe des éléments les plus liés entre eux ; si le dioïde est (\mathbb{R}, \max, \min) on trouve une classification hiérarchique de préordres.

CROISEMENT DE C

LERMAN I.C.
Laboratoire de
I.R.I.S.A. - Un
35042 - RENNES

On montre que la donnée d'un ensemble d'attributs descriptifs permet, en fait, de définir une relation d'ordre relative d'un individu donné à une matrice de préférences et la mise en évidence d'une classification des individus décrits. On développe dans ce papier la notion de deux classifications floues par l'analyse de la donnée aux cases du croisement conformément à la théorie de la généralisation adéquate de la théorie des ordres. Ce qui permet de situer, l'une par rapport à l'autre, le même ensemble S d'individus, respectivement une classification monothétique et une classification polythétique à partir de données descriptifs. Une telle approche est particulièrement utile pour l'analyse des questionnaires pouvant être effectuée en première comprenant les attributs de la première et la seconde, les attributs d'attributs.

UN PROBLEME DE LA CLASSIFICATION AUTOMATIQUE POSE PAR
L'INTERROGATION DES RASSEMBLEMENTS DOCUMENTAIRES

Auteur: F. MURTAGH,
Laboratoire de Statistiques Mathématiques,
Université de Paris VI.

Nous proposons un traitement rigoureux de la classification automatique appliquée à l'interrogation des rassemblements documentaires. En particulier on expose une nouvelle approche à l'analyse probabiliste des résultats d'une classification ascendante hiérarchique. Quelques résultats nécessaires à la mise au point de cette théorie sont fournis. Bien que cette étude s'applique à - et s'explique par - l'exemple de l'interrogation des systèmes documentaires, on en suggère la généralisation aux autres applications de la classification automatique.

APPLICATION DES METHODES D'ANALYSE
A L'ETUDE DE LA CHARGE DE TRAVAIL

Auteur : M. RALAMBONDRAIN H.,
LABORIA - BP 105 -781

RESUME : Alors qu'on envisage une nouvelle configuration, il est nécessaire de faire la charge de l'IRIS 80 de cette étude.

LES DONNEES : Le fichier de consommations machines donner un ordre de grandeur. On signale qu'il a été fait 13000 travaux batch. Les mois : janvier à octobre. Ont été extraits divers fichiers utilisateurs, fichiers de partage, etc,...

LES ANALYSES EFFECTUEES :

- . Etude des consommations machines par types.
- . Typologie des utilisateurs.
- . Suivi des utilisateurs.
- . Etude de la charge de travail.
- . Typologie des travaux.
- . Typologie des sessions.

METHODES MISES EN OEUVRE :

- . Analyse factorielle.
- . Analyse en composantes principales.
- . Classification hiérarchique.
- . Nuées dynamiques.