Y. LE FOLL

Sur le choix des aînés et benjamins des nœuds en vue du tracé optimum d'un arbre

Les cahiers de l'analyse des données, tome 8, n° 2 (1983), p. 237-238

<http://www.numdam.org/item?id=CAD_1983__8_2_237_0>

Les Cahiers de l'Analyse des Données
Vol. VIII 1983 n° 2 p. 237 238

SUR LE CHOIX DES AÎNÉS ET BENJAMINS DES NŒUDS EN VUE DU TRACÉ OPTIMUM D'UN ARBRE

[A.B.N. ARBRE]

par Y. Le Foll (1)

Un chapitre du traité de l'Analyse des Données (TI C n° 13 [*PEURS*]) commence par cette mise en garde.

"En commentant les résultats issus d'un algorithme de classification il faut prendre garde que d'une part tant l'amplitude des intervalles séparant les classes que la cohésion même de celles-ci ne peuvent être éprouvées, et que d'autre part l'ensemble des classes suspendues aux ramifications de l'arbre est semblable à ces essaims d'objets mobiles que l'art expose au souffle de l'air pour nous en proposer la changeante figure ; en ce sens que, quant à l'ordre global des classes, une multitude de combinaisons hypothétiques est possible ."

Un article paru dans les *Cahiers* (cf. C.A.D. Vol I n° 4 pp 441 sqq.) suggère de préciser pour chaque paire (a(n) b(n)) les points entre lesquels se réalise la distance minima ce qui crée comme un système de liens entravant les branches et permet de dessiner une sorte d'arbre de longueur minima associée à la hiérarchie. Nous proposons ici une autre voie : choisir pour chaque noeud n celui de ses deux descendants qu'on appelle aîné a(n) et celui qu'on appelle benjamin b(n) de telle sorte que dans l'impression usuelle de l'arbre les contiguïtés créées entre les classes respectent le mieux possible la disposition spatiale.

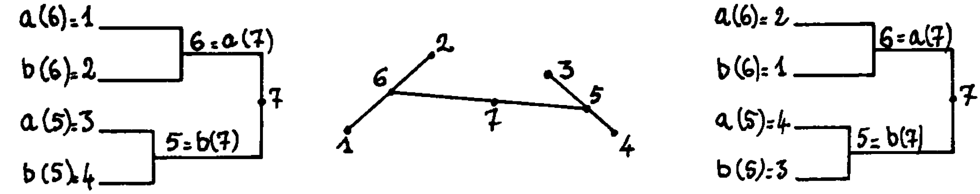
Partons d'un exemple très simple qui suffira à introduire le problème et en suggérer la solution. Nous supposerons pour fixer les idées que l'arbre s'imprime en plaçant toujours l'aîné a(n) d'un noeud au-dessus de son benjamin b(n). Soit les points 1, 2, 3,4 ; qui s'organisent suivant la hiérarchie suivante :

$$7 = \{6,5\} \ ; \ 6 = \{1,2\} \ ; \ 5 = \{3,4\}$$

Pour chacun des 3 noeuds, 7, 6, 5, le choix de l'aîné et du benjamin peut se faire de deux façons possibles soit au total $2^3 = 8$ possibilités (en général, pour une hiérarchie binaire sur n éléments, il y a n-1 noeuds donc 2^{n-1} possibilités). Voici la figure supposée plane des 4 points, avec deux écritures possibles de l'arbre.

---

(1) *Docteur 3° cycle. Ingénieur à l'E.N.S.T. - Paris.*



Il est manifeste que le tracé proposé à gauche est le plus satisfaisant en ce qu'il donne la succession (1, 2, 3, 4) qui est justement celle de la figure plane. D'ailleurs le choix de a(7) et b(7) (étiquetage des successeurs immédiats du sommet) est indifférent car il suffit de changer simultanément tous les arbres pour passer d'une figure à sa symétrique. Reste donc à baser sur un indice formel le choix de a(n) et b(n) pour tout n autre que le sommet.

Pour cela nous procédons par voie descendante : supposant choisi pour un noeud n un aîné an et un benjamin bn, nous nous proposons de choisir parmi les deux successeurs san et tan de an (ou parmi les deux successeurs de bn) celui qui sera aîné et celui qui sera benjamin : le critère sera simplement le suivant : bien ordonner le dipole (aan, ban) relativement au dipole (an, bn). On demandera simplement que :

$$\text{signe } \langle (aan\ , ban),\ (an\ , bn)\rangle = +$$

où $\langle , \rangle$ désigne le produit scalaire des vecteurs.

Il est facile de vérifier sur la figure le bien fondé de cette règle d'ailleurs très facile à appliquer dans un programme : le produit scalaire $\langle \overrightarrow{12}, \overrightarrow{65}\rangle$ est positif ; ainsi que $\langle \overrightarrow{34}, \overrightarrow{65}\rangle$ d'où les choix a6 = 1, b6 = 2 ; a5 = 3, b5 = 4 ; étant entendu que a7 = 6 et b7 = 5.

Quel que soit l'intérêt de cette suggestion pour le dessin des arbres, il reste que rien ne dispense le lecteur d'un arbre de recourir à l'analyse factorielle pour se figurer le plus précisément possible la disposition des classes.

Note 1 : Le calcul des produits scalaires est particulièrement simple si on possède les coordonnées des noeuds sur un système d'axe orthonormé, par exemple sur les axes factoriels.

On a alors :

$$\langle (aan,ban),(an,bn)\rangle = \Sigma_\alpha (F_\alpha(aab) - F_\alpha(ban))(F_\alpha(an) - F_\alpha(bn)) .$$

Le calcul est un peu plus complexe si on effectue la classification d'après un tableau de contingence et utilise la métrique du χ^2. On retrouve en tout cas des calculs requis dans l'interprétation par FACOR ou VACOR.

Note 2 : Outre l'article des C.A.D. déjà cité, on pourra consulter :

G. Brossier : Représentation ordonnée des classifications hiérachiques (in *Stat. et A. des D.* 1980-2 pp 31-44). En bref l'auteur s'intéresse au cas particulier où l'arbre de ℓ minima associée à une C.A.H. (selon l'article des C.A.D.) est une suite ordonnée de points : dans ce cas la représentation construite ici est particulièrement adéquate.