

Commentaires sur l'article de G. Celeux et C. Robert

Pierre Cazes

Université Paris 9-Dauphine

1. Le premier but de l'ACP et de l'ACM n'est pas de fournir une classification ou une partition. Si l'on veut donc comparer réellement les 5 codages utilisés (les deux codages continus et les trois codages en 0-1) du point de vue de la recherche d'une partition, il vaut mieux, me semble-t-il, utiliser une technique de classification (la distance entre 2 individus étant calculée avec la métrique déduite de l'analyse factorielle associée ACP ou ACM, qui se réduit (à une constante près en ACM) à la métrique usuelle, sauf pour le codage additif).
2. Il semble normal que le dernier codage (variables considérées comme qualitatives) n'ait rien donné d'intéressant compte tenu du grand nombre de modalités (11 ou 12) de chaque variable par rapport au nombre d'individus. Nous n'en parlerons plus par la suite, bien que pour être logique, les auteurs auraient dû essayer avec ce découpage trivial, le codage additif, qui leur aurait sans doute permis de retrouver la partition désirée.
3. Aussi bien en ACP, qu'en ACM, les auteurs n'utilisent que les 2 ou 3 premiers axes factoriels, sans donner les pourcentages d'inertie (qui, il est vrai, n'ont pas grand sens en ACM). Or, il y a au plus 4 (3 pour la 2ème ACP) facteurs non triviaux en ACP et 8 (ou 11 dans le cas du codage additif) en ACM. Peut-être, pourrait-on dans l'espace des 4 ou 5 premiers axes factoriels issus de l'ACM (sur codage disjonctif complet) déceler la partition que l'on recherche ; mais cela ne peut plus se faire visuellement, mais à l'aide d'un algorithme de classification (on rejoint ainsi la remarque du 1)).
4. De toute façon, compte tenu du nombre relativement faible d'individus, si l'on veut avoir des résultats stables, il ne faut pas dans l'ACM utiliser le codage disjonctif complet, mais un codage flou, par exemple un codage linéaire par morceaux (cf. Cazes, 1990). Avec ce dernier codage, on retrouverait probablement la partition recherchée dans l'espace des 2 ou 3 premiers axes factoriels issus de l'ACM.
5. Je suis entièrement d'accord avec les auteurs, quand ils disent dans le dernier paragraphe de l'article que lorsqu'on discrétise des variables continues, il est bon de guider le choix du codage par des informations a priori, ou des informations issues d'analyses préliminaires. Cela me semble fondamental.

Bibliographie

- [1] Cazes P. (1990) *Codage d'une variable continue en vue de l'analyse des correspondances*. Revue de Statistique Appliqué **38**, 35-51.