Ability of correspondence analysis to find latent structures: Comparison with association model

Tatsuo Otsu[1]
Department of Behavioral Science
Hokkaido University
t_otsu@pop02.odn.ne.jp

Keywords: dimension reduction, horse-shoe problem, association model

Abstract

Correspondence analysis (CA) is widely used for finding latent structure of two-way table. Guttman(1950) pointed out an interesting property of the method. If the table has latent one-dimensional structure, the obtained scores are similar to orthogonal polynomials of the index. The scores for the first non-trivial dimension are linear for the index. The second scores are quadratic, and the third scores are cubic. Assume the upper triangular elements of an $N \times N$ table are 1's, and lower off-diagonal elements are zeros. In this typical case, the singular-values are $r_k = 1/(k+1),\ (k = 1, ..., N-1)$.

Guttman called these solutions "intensity" and "closure". Although he suggested using them for inspecting structures of the data, this property may cause serious problems in practical data analysis. If the data have a prominent latent component, the "resonance" sequence of the largest singular-values disturbs detection of less prominent components.

If the latent structure of the two-way data is multivariate normal, Goodman's association model fits well to the data. This is brought by quadratic nature of the log-likelihood function of the distribution. With this property, association models are able to avoid "resonance" singular-values. But we confront computational difficulties caused by local maximums in large size tables.

I will show analytical properties of the problem and examples for comparing the two methods.

References

Guttman,L. (1950). The principal components of scale analysis. in Stouffer,S.A., Guttman,L., Suchman,E.A., Lazarsfeld,P.F., Star,S.A. & Clausen,J.A. *Measurement and Prediction.* Princeton University Press.

Goodman,L. (1991). Measures, models, and graphical displays in the analysis of cross-classified data (with discussions). *Journal of the American Statistical Association*, 86, 1085–1138.

Otsu,T. and Matsuo,H. (2002). MTV and MGV: Two criteria for nonlinear PCA. in Baba,Y., Hayter,A.J., Kanefuji,K. and Kuriki,S. eds. *Recent Advances in Statistical Research and Data Analysis.* Tokyo: Springer-Verlag.

[1]The author's affiliation after April,2003: Research Devision, National Center for University Entrance Examinations, 2-19-23, Komaba, Meguro-ku, Tokyo, 153-8501, Japan.