Interpreting graphical representation of asymmetric data

Takayuki Saito
Graduate School of Decision Science and Technology
Tokyo Institute of Technology
gsaito@valdes.titech.ac.jp

Keywords: Escoufier & Grorud procedure, Euclidean distance, Gower's procedure, Hermitian matrix, proximity, singular value decomposition, skew-symmetry.

Abstract

Given an asymmetric matrix ($n \times n$) of observations $O = (o_{jk})$, construct a symmetric matrix $S = (O + O')/2$ and a skew-symmetric matrix $A = (O - O')/2$. Applying Gower's procedure or Escoufier & Grorud procedure to those data matrices, we obtain a graphical representation of objects in 2 dimensions. For interpretation of the configuration, this paper suggests that we can examine Euclidean distance which corresponds to some quantity in terms of data.

In Gower's approach, starting with O whose o_{jj} may or may not be defined, we perform the singular value decomposition of A. Assume that $rank\, A = 2r$. Let $\boldsymbol{u}_t = (u_{jt})$ and $\boldsymbol{v}_t = (v_{jt})$ be the pair of singular vectors associated with singular value μ_t. Then we have a plot of n points in terms of (u_{jt}, v_{jt}) on the u-v plane $(t = 1, 2, \ldots, r)$. For the t-th plane, denote the Euclidean distance between points j and k by $d_{jk}(t)$. For data transformed by (1), we obtain relation (2).

$$c_{jk} = \frac{1}{4}\sum_{i=1}^{n}(o_{ij} - o_{ik} + o_{ki} - o_{ji})^2 \quad (1)$$

$$c_{jk} = \sum_{t=1}^{r}\mu_t^2\, d_{jk}^2(t) \quad (2)$$

In Escoufier & Grorud' approach, starting with O whose o_{jj} should be defined, we deal with an Hermitian matrix $H = S + i\,A\,(i^2 = -1)$. Write the eigenvalue problem as $H\boldsymbol{z}_t = \lambda_t \boldsymbol{z}_t$ where $\boldsymbol{z}_t = \boldsymbol{x}_t + i\,\boldsymbol{y}_t$, $\boldsymbol{x}_t = (x_{jt})$ and $\boldsymbol{y}_t = (y_{jt})$. Assume that $rank\, H = r$. In the plot of n objects in terms of (x_{jt}, y_{jt}) for each $t (= 1, 2, \ldots, r)$, denote the Euclidean distance between points j and k by $d_{jk}(t)$ on the x-y plane. For data transformed by (3), we obtain relation (4).

$$p_{jk} = o_{jj} + o_{kk} - o_{jk} - o_{kj} \quad (3)$$

$$p_{jk} = \sum_{t=1}^{r}\lambda_t\, d_{jk}^2(t) \quad (4)$$

References

Gower, J. C. (1977) The analysis of asymmetry and orthogonality. In: Barra, J. R. et al. (eds.), *Recent Developments in Statistics*, 109-123, North-Holland.
Escoufier,Y. & Grorud,A.(1980) Analyse factorielle des matrices carrees non-symétriques. In Diday, E. (Eds.) *Data Analysis and Informatics*, 263-276, North Holland.