ANALYZING INDICATOR MATRICES USING DISTANCE ASSOCIATION METHODS

JAN DE LEEUW

So far, there are separate programs for what we call "logistic unfolding" and what we call fitting "distance association models", although it is clear there is a close relationship between the two. In this note we make the relationship explicit and indicate that "logistic unfolding" can be done by slightly modifying the "distance association" techniques.

The key insight is that if we have a partitioning $G = (G_1 \mid \cdots \mid G_m)$ of the data (frequency matrix), then we can fit a Poisson model of the form

$$\mathbf{E}(g_{i\ell}^j) = \alpha_i^j \beta_\ell^j \exp\{-\phi(x_i, y_\ell^j)\},$$

where ϕ is Euclidean distance or Euclidean squared distance. Superscript j is used to index the submatrices of the data matrix. The submatrices can be (binary) indicator matrices, but they can also be subtables of a general frequency matrix.

This is the same model we fit using distance association technique, with the minor difference that each subset of columns has its own row scores α^j. Since row scores (along with column scores) are fitted separately by iterative proportional fitting steps, this generalization does not really complicate the algorithm (except for some book keeping of index vectors).

Department of Statistics, University of California, Los Angeles, CA 90095-1554

E-mail address, Jan de Leeuw: deleeuw@stat.ucla.edu

URL, Jan de Leeuw: http://gifi.stat.ucla.edu

Date: February 18, 2006.

2000 *Mathematics Subject Classification.* 62H25.

Key words and phrases. Multivariate Analysis, Correspondence Analysis.