



A flexible clustering method based on UPGMA and ISS

1988

M. J. Dallwitz

Abstract. A formula is given for a flexible combinatorial clustering strategy, having UPGMA and ISS as special cases.

When numerical clustering methods are used to investigate the structure of taxonomic groups, it is common practice to try several methods, having different clustering intensities. For example, one might use single linkage (weak), UPGMA (medium), and complete linkage (strong) (Sneath and Sokal 1973). Lance and Williams (1967) proposed a 'flexible' clustering strategy in which the intensity of clustering can be varied continuously by means of a parameter, β . This strategy has the WPGMA method as a special case ($\beta=0$). This could be considered a disadvantage, as the UPGMA method is generally preferred over WPGMA (Sneath and Sokal 1973). Furthermore, experiments with data on grass genera (Watson, Clifford and Dallwitz 1985) have tended to show that the Lance and Williams flexible method, with parameter values giving intense clustering, is inferior (Watson and Dallwitz, unpublished) to other intensely clustering methods such as ISS (Burr 1970) and information analysis (Williams, Lambert and Lance 1966). Abel and Williams (1985) also argue that UPGMA and ISS are preferable to the Lance and Williams flexible method.

The flexible clustering strategy suggested here has UPGMA and ISS as special cases. In the notation of Sneath and Sokal (1973), the formula is

$$U_{(J,K),L} = [(t_J + \sigma t_L)U_{J,L} + (t_K + \sigma t_L)U_{K,L} - \sigma t_L U_{J,K}] / (t_J + t_K + t_L)$$

where σ is the intensity parameter. $\sigma=0$ gives UPGMA and $\sigma=1$ gives ISS.

The formula is easy to incorporate into clustering programs which use combinatorial fusion strategies.

References

- Abel, D.J. and Williams, W.T. 1985. A re-examination of four classification fusion strategies. *Comput. J.* 28, 439–443.
- Burr, E.J. 1970. Cluster sorting with mixed character types. II. Fusion strategies. *Aust. Comput. J.* 2, 98–103.
- Lance, G.N. and Williams, W.T. 1967. A general theory of classificatory sorting strategies. I. Hierarchical systems. *Comput. J.* 9, 60–64.
- Sneath, P.H.A. and Sokal, R.R. 1973. Numerical taxonomy — the principles and practice of numerical classification. (W. H. Freeman: San Francisco.)
- Watson, L., Clifford, H.T. and Dallwitz, M.J. 1985. The classification of Poaceae: subfamilies and supertribes. *Aust. J. Bot.* 33, 433–484.
- Williams, W.T., Lambert, J.M. and Lance, G.N. 1966. Multivariate methods in plant ecology. V. Similarity analyses and information-analysis. *J. Ecol.* 54, 427–445.