

Monotonicity of an asymmetric agglomerative hierarchical clustering algorithm

Akinobu Takeuchi	Hiroshi Yadohisa	Koichi Inada
College of Social Relations	Dept. of Math. & Comp. Sci.	Dept. of Math. & Comp. Sci.
Rikkyo University	Kagoshima University	Kagoshima University
akitake@rikkyo.ac.jp	yado@sci.kagoshima-u.ac.jp	inada@sci.kagoshima-u.ac.jp

Keywords: Asymmetric (dis)similarity data, Clustering algorithm, Monotonicity.

Abstract

Similarity data for cluster analysis is usually assumed to be symmetric, that is, a similarity (or dissimilarity) from object A to object B is considered to be the same as that from object B to object A. However, the data used in actual analysis is intrinsically asymmetric, requiring the data to be symmetrized before use in a usual symmetric clustering algorithm (Hubert, 1973). An interpretation of asymmetry as having elemental meaning has also been proposed, in which the researcher analyzes given data using an algorithm that considers the asymmetry. According to this idea, agglomerative hierarchical clustering algorithms for asymmetric similarity data have been proposed (Fujiwara, 1980; Okada and Iwashita, 1996). Yadohisa (2002) suggested an extended updating formula, paralleling the symmetric case proposed by Lance and Williams (1967), in order to handle a wide range of asymmetric agglomerative hierarchical clustering algorithms. However, the detailed character of the proposed algorithms for asymmetric data was not discussed in their paper.

In the present, the monotonicity of the updating formula is defined for an asymmetric clustering algorithm, and the relationship between the monotonicity and the parameters of the extended updating formula proposed by Yadohisa (2002) is discussed. When two clusters are merged and the similarities between the combined cluster and another clusters are updated, the updated similarity may be greater (or less) than the original similarity. The monotonicity states that the updated similarity monotonically decreases, which allows the clustering algorithm to avoid reversal of the updated distance.

References

- Fujiwara, H. (1980). Methods for cluster analysis using asymmetric measures and homogeneity coefficient, *The Japanese Journal of Behaviormetrics*, **7**, 12–21 (in Japanese).
- Hubert, L. (1973). Min and max hierarchical clustering using asymmetric similarity measures, *Psychometrika*, **38**, 63–72.
- Lance, G. N. & Williams, W. T. (1967). A general theory of classificatory sorting strategies: 1. hierarchical systems, *The Computer Journal*, **9**, 373–380.
- Okada, A & Iwamoto, T. (1996). University enrollment flow among the Japanese prefectures: A comparison before and after the Joint First Stage Achievement Test by asymmetric cluster analysis, *Behaviormetrika*, **23**, 169–185.
- Yadohisa, H. (2002). Asymmetric agglomerative hierarchical clustering and graphical representation of its result, To appear in *Bulletin of the Computational Statistics of Japan* (in Japanese).