

Clustering in Contemporary Mixed-valued Data

L. Billard

University of Georgia, Athens, Georgia USA lynne@stat.uga.edu

Jaejik Kim

Georgia Health Sciences University, Augusta, Georgia USA jaejik@gmail.com

Contemporary databases can be too large to be analysed by traditional methods. One approach is to aggregate the data according to whatever scientific question(s) are driving the analysis. The resulting data are perforce intervals, lists, histograms, and the like, which formats are but examples of symbolic data (Diday, 1987). This work will look at clustering for data sets in R^p of mixed types whereby some random variables take interval realizations, some histogram realizations, and/or some are modal or non-modal multi-valued realizations.

Key Words: agglomerative, interval, histogram, list