PSYCHOMETRIKA—VOL. 32, NO. 3
SEPTEMBER, 1967

HIERARCHICAL CLUSTERING SCHEMES*

STEPHEN C. JOHNSON

BELL TELEPHONE LABORATORIES,
MURRAY HILL, NEW JERSEY

Techniques for partitioning objects into optimally homogeneous groups on the basis of empirical measures of similarity among those objects have received increasing attention in several different fields. This paper develops a useful correspondence between any hierarchical system of such clusters, and a particular type of distance measure. The correspondence gives rise to two methods of clustering that are computationally rapid and invariant under monotonic transformations of the data. In an explicitly defined sense, one method forms clusters that are optimally "connected," while the other forms clusters that are optimally "compact."

Introduction

In many empirical fields there is an increasing interest in identifying those groupings or clusterings of the "objects" under study that best represent certain empirically measured relations of similarity. For example, often large arrays of data are collected, but strong theoretical structures (which might otherwise guide the analysis) are lacking; the problem is then one of discovering whether there is any structure (i.e., natural arrangement of the objects into homogeneous groups) inherent in the data themselves. Recent work along these lines in the biological sciences has gone under the name "numerical taxonomy" [Sokal, 1963].

Although the techniques to be described here may find useful application in biology, medicine and other fields as well, we shall use psychology as an illustrative field of application. In that field, the "objects" under study might, for example, be individual human or animal subjects, or various visual or acoustic stimuli presented to such subjects. We might want to use measures that we have obtained on the similarities (or psychological "proximities") among the "objects" to classify the objects into optimally homogeneous groups; that is, similar objects are assigned to different groups.

Suitable data on the similarities among the objects (from which such a natural grouping might be derived) may be obtained directly or indirectly. For example, sometimes one obtains for every pair of objects a subjective

*I am indebted to R. N. Shepard and J. D. Carroll for many stimulating discussions about this work, and for aid in preparing this paper.

rating of similarity, or, (what is often very closely related) a measure of the confusion or "interchangeability" of the objects. Less directly, we may measure a number of attributes of the objects (often termed a profile of measures) and combine them to form a single measure of similarity. Various kinds of measures of profile similarity can be used for this purpose (e.g., product-moment-correlation, covariance, or the sum of squared or absolute differences between corresponding components of the profiles).

The problem of course, is that if the number of objects is large, the resulting array of similarity measures (containing, as it does, one value for each *pair* of objects) can be so enormous that the underlying pattern or structure is not evident from inspection alone. This paper discusses procedures which, when applied to such an array of similarity measures, constructs a hierarchical system of clustering representations, ranging from one in which each of the n objects is represented as a separate cluster to one in which all n objects are grouped together as a single cluster.

An algorithm for finding such a clustering representation was sought that would have the following features:

1. The input should consist solely of the $n(n-1)/2$ similarity measures among the n objects under study. This is in contrast to some previous methods which additionally require that each object be initially represented as a point in Euclidean space. (In many applications the restriction to a representation of the grouping in the concrete, spatial sense of an Euclidean metric seems unnecessarily and undesirably severe).
2. There should be a clear, explicit, and intuitive description of the clustering; i.e., the clusters should mean something. Some of the published clustering methods have nice algorithms, but when they have been carried out it is difficult to see exactly what problem has been solved.
3. The clustering procedure should be essentially invariant under monotone transformations of the similarity data. Often in psychology we have confidence in our data only up to rank-order; the absolute numbers obtained from the experiments may lie along virtually any scale. The method of Ward [1963], which inspired much of this current study, is indeed so general as to permit monotone invariant methods, but they are not explicitly treated.

The notion of a hierarchical clustering scheme, the central idea of this paper, was abstracted from examples given by Ward [1963]. We first consider such schemes, and develop a correspondence between hierarchical clustering schemes and a certain type of metric. Two recursive methods are then given for obtaining hierarchical clustering schemes from a given similarity matrix, and finally the significance of these two methods is discussed and illustrated by application to real data.