Note on "A Comparison of Four Clustering Methods"

Maurice Lorr
Catholic University

Blashfield and Morey (1980) reported on a comparison of four clustering methods using monte carlo procedures to generate data sets resembling three MMPI score profile patterns. The four methods compared were (1) average linkage (Sokal & Michner, 1958); (2) Ward’s (1963) hierarchical procedure; (3) inverse factor analysis; and (4) a so-called clumping procedure called Buildup (Lorr & Radhakrishnan, 1967; Suziedelis, Lorr, & Tonesk, 1976).

The purpose of this note is to alert readers that Blashfield and Morey misinterpreted the criteria of inclusion and exclusion required in the Buildup procedure. They applied the criteria in an order that was reverse to the order specified. The clustering procedure calls for a high inclusion cutoff to determine membership in a cluster, followed by a lower exclusion cutoff to remove outliers. Blashfield and Morey, however, began with an inclusion cutoff of $r = .60$ and followed it with an exclusion cutoff of $r = .74$. The first cutoff is thus lower than the second.

The Buildup procedure is nonhierarchical, uses correlations, and represents a form of average linkage analysis (Sneath & Sokal, 1973, p. 182). Membership in a cluster is determined by two criteria—inclusion and exclusion cutoffs—that are specified in advance. Beginning with a nucleus of three profiles, the matrix is searched for the profile that, on the average, correlates highest with the nucleus and satisfies the inclusion cutoff $C_{in}$ (e.g., $r = .60$). The process continues until no more profiles can be added to the cluster. Next an exclusion $C_{ex}$ is set, which defines dissimilarity, keeps outliers out, and prevents cluster overlap. Any profile in the residual matrix is deleted if it correlates, on the average, $C_{ex}$ or higher (e.g., $r = .50$) with the first cluster. Profiles excluded are not considered again for inclusion in subsequent cycles. The second cluster is generated from the reduced matrix in the same manner. Profiles are added to a nucleus if they satisfy the inclusion limit. Again, the exclusion limit $C_{ex}$ is applied to delete profiles that are too close to the second cluster. The process is continued until no cluster with at least four members can be formed.

The inclusion limit suggested is the value at which a coefficient of correlation based on $K$ independent variates is significant at $p = .025$. The exclusion limit is set at $p = .05$. Experimental trials with simulated data have shown these values to be optimal (Suziedelis et al., 1976). Deletion of profiles in the matrix that are too close to a later emerging cluster does not exclude profiles included in preceding clusters. Accordingly, if a cluster is not the first, then all profiles in the cluster formed in a preceding cycle that
correlate at or above $Cex$ with the cluster in question are also deleted.

Thus, two errors were involved in Blashfield and Morey's application of Buildup. First, their inclusion cutoff was lower than their exclusion cutoff. Second, the exclusion cutoff should have been applied to the residual matrix instead of the complete matrix. The effect of this application of cutoff can be surmised by examining the findings. Blashfield and Morey used two criteria—misclassifications and coverage—to evaluate how well a clustering procedure solved the MMPI mixture of three basic categories. A misclassification is defined as a profile assigned to a cluster whose dominant membership is composed of profiles from a different parent sample. Coverage is the number of profiles in a data set that are not assigned to any cluster. In the first mixture of three, six clusters were found; and on the replication data set seven were found. It is here that the effects of the misapplication of cutoffs probably occurred. The exclusion limit was designed to prevent emergence of unwarranted clusters. It seems likely that only three clusters would be found with the proper application of the two criteria.

References


Author's Address

Send requests for reprints or further information to Maurice Lorr, Boys Town Center, Catholic University, Washington DC 20064.