APCLUST: A versatile Apple II program to perform agglomerative hierarchical clustering analysis

ANDREW R. GILPIN
University of Northern Iowa, Cedar Falls, Iowa

A variety of algorithms for performing agglomerative hierarchical clustering analysis have been embodied in programs on mainframe computers (Blashfield, Aldenderfer, & Morey, 1982), but few clustering programs are available as yet for microcomputers (Whaley, 1982; recently Gilpin, 1985, described a program which performs hierarchical clustering analysis using a divisive procedure). Program APCLUST performs hierarchical clustering analyses on the Apple II family of computers.

Requirements. The program is written in APPLESOFT BASIC, a dialect of Microsoft BASIC, and should run without changes on other members of the Apple II family (II, IIe, and IIc), provided at least 48K bytes of RAM are available (the program requires 28K bytes of free memory). It is written to direct output to a printer if one is available, but may be used on systems lacking a printer as well. Conversion to other systems running Microsoft BASICs should be straightforward; a section of the APPLESOFT BASIC manual is devoted to relevant peculiarities of the Apple II version. A disk drive running DOS 3.3 is necessary if the file input option, discussed below, is to be used.

Input/Output Options. The program APCLUST requires as input the upper half (including diagonal) of a triangular matrix of values representing dissimilarities between pairs of up to 60 objects (variables). Data may be entered in three ways: (1) via DATA statements added to the program; (2) via entry from the keyboard during program execution; and (3) via a previously created data file on diskette (a utility program, APCLUST-FILEBUILDER, is provided to create the file). Output may be directed to the screen or to a printer in any slot (port) number. If desired, the program will pause periodically during output, awaiting a response from the user, to prevent results from scrolling off the screen. APCLUST assumes that the data matrix represents a single two/way

one-mode proximities matrix; an additional option will transform similarity data by subtraction from the largest value in the matrix. A data editing routine is incorporated into APCLUST.

Methodology. APCLUST is an interactive implementation of Späth's (1980) algorithm, written in FORTRAN for a batch environment. Any of seven methods of analysis may be selected, including Johnson's (1967) "Min" (single-link) and "Max" (complete-link) methods; Sokal & Sneath's (1963, p. 182) average method; three methods described by Gower (1967), using various forms of weighted averages; and Ward's (1963) procedure. Output includes a copy of the raw data matrix (and the transformed version if that option was selected), a table summarizing the clusterings and distances at which they are formed, and a dendrogram representing the same information in a tree form. Runs of APCLUST on data sets reported by Späth (1980) indicate close agreement between his program and APCLUST.

Availability. A user's manual, including program listings for APCLUST and APCLUSTFILEBUILDER, and a sample run of the former program, is available without charge by writing to the author. Copies of the program on diskette (DOS 3.3) may be obtained by sending a blank diskette plus $2 to cover postage and handling costs (make checks payable to the University of Northern Iowa).

REFERENCES

BLASHFIELD, R. K., ALDENDERFER, M. S., & MOREY, L. C. (1982). Cluster analysis software. In P. R. Krishnaish & L. N. Kanal (Eds.), *Handbook of statistics, Vol. 2: Classification, pattern recognition, and reduction of dimensionality.* Amsterdam: North-Holland.

GILPIN, A. R. (1985). APPLECOR: An Apple II implementation of hierarchical clustering analysis using the CONCOR algorithm. *Behavior Research Methods, Instruments, & Computers, 17,* 140.

GOWER, J. C. (1967). A comparison of some methods of cluster analysis. *Biometrics, 23,* 623-637.

JOHNSON, S. C. (1967). Hierarchical clustering schemes. *Psychometrika, 32,* 241-254.

SOKAL, R. R., & SNEATH, P. H. A. (1963). *Principles of numerical taxonomy.* New York: Freeman.

SPÄTH, H. (1980). *Cluster analysis algorithms for data reduction and classification of objects.* Chichester, West Sussex, England: Ellis Horwood.

WARD, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58,* 236-244.

WHALEY, C. P. (1982). Interactive clustering. *Behavior Research Methods & Instrumentation, 14,* 170-175.

The author's mailing address is: Department of Psychology, University of Northern Iowa, Cedar Falls, IA 50614.