WHAT COMPUTERS CAN'T DO FOR ARCHAEOLOGISTS.

James Doran.

Computer Centre,
University of Essex,
Colchester, England.

SUMMARY

Over the past decade a variety of techniques of multivariate data analysis have been tested as aids to the interpretation of archaeological data. Prominent have been various forms of cluster analysis, of multidimensional scaling, and of seriation. These techniques have in common that they aim to reveal to the archaeologist simple patterns or "structure" in his data which are likely to be archaeologically significant and which he might not have detected unaided. Such techniques have wide applicability and their value is well established. However there are problems in their use which derive partly from their inherent limitations, and partly from a tendency by users to underestimate these limitations.

Seriation is a case in point. Most seriation procedures require the data first to be prepared as an incidence matrix. In practice this means ignoring most of the available information including, for example, the definitions of the descriptive variables used in the study. This is surely a major limitation. Provided it is borne in mind, and provided the sequence or sequences generated by the seriation procedure are treated by the archaeologist as data patterns to be interpreted as he thinks fit, then no harm will be done. Often, however, the aim is explicitly to recover chronological sequence. It has not always been fully realised that if this is indeed the aim then either the data must be gathered to meet stringent (perhaps impossible) archaeological requirements designed to eliminate non-chronological sources of variation or the seriation procedure must be greatly elaborated to the point at which it can itself detect and eliminate unwanted sources of variation.

The latter possibility is being investigated using a computer program, SOLCEM, written in Algol-60 for the ICL 1906A (Doran and Hodson, 1974). Working from typical cemetery excavation data (grave locations, sex and age of burials, grave inventories, etc.), the program seeks an integrated interpretation of the cemetery involving not only the chronological sequence of the graves, but also their spatial relationships, the relationships between grave contents and grave types and the evolving typology of the artifacts deposited. Essentially the program searches through a very large, implicitly defined, set of possible interpretations of the data, using a variety of data analytic techniques to try to "home-in" on the most plausible complete interpretation.

It must be stressed that SOLCEM is, as it stands, merely a research tool and not of practical use to archaeologists. The range of cemetery interpretations of which it can "conceive" is too limited for its conclusions to be of practical interest. To extend this range other than in a piecemeal and probably largely ineffectual manner requires the solution of a problem of fundamental importance - that of how to provide the program with mechanisms such that it can store and utilize in interpretation the detailed archaeological and general knowledge which is

relevant to its task. The problem is not primarily one of quantity but one of coding - how do we code potentially relevant information about, say, prehistoric metal technology, or house styles or burial practices so that computer programs can do something with them.

In its general form this problem of knowledge representation is currently receiving much attention among computer scientists. Difficult though the problem is, some progress has been made and a number of computer programs have been demonstrated which autonomously utilize specialist knowledge to solve problems in limited but non-trivial contexts. It would be unwise, therefore, to assume that the archaeological version of the problem is inherently impossible.

Nevertheless it seems reasonable to argue that our present inability to program computers to utilize detailed knowledge in the analysis of archaeological data is a general and major limitation on the use of computers as aids to archaeological interpretation. It is apparent, for example, that the techniques of multivariate analysis mentioned at the outset "know" virtually nothing about archaeology. Therefore, they can never be more than very limited aids to the archaeologist. Agreement that knowledge utilization by the computer is fundamental in archaeological data analysis would, for example, prevent vain efforts to find the "best" clustering method, and might go far to reassure the many archaeologists who, with some justification, feel that all the computer can do for them is to simplify their problems to the point of absurdity.

Reference

Doran, J.E.
Hodson, F.R.
1974 Mathematics and Computers in Archaeology. Edinburgh University Press, Section 11.7. (At press).