29

Formal methods for the analysis of archaeological data: data analysis vs expert systems

Vanda Vitali

(Royal Ontario Museum, Toronto, Canada, and, C.N.R.S. — U.P.R. 315, Paris, France)

29.1 Introduction

Over the last two decades the use of computer-based methods in archaeology has experienced a considerable increase (Doran & Hodson 1976, Orton 1980, Carr 1985, Ruggles & Rahtz 1988, Rahtz 1988). While at first cautious, the archaeological research community has now accepted computers as one of its essential tools and, as in other fields of scientific endeavor, is exploring the applications of new computer-based methods in research, teaching and communication. This phenomenon has resulted from two mutually interrelated developments.

On one hand, modern archaeology has become much more oriented towards the systematic accumulation of data. Numerical data are now generated as a means of describing various purely archaeological parameters. Also, technical studies of archaeological material, which are being employed more and more in archaeology, are producing numerical information.

On the other hand, the use of computers and computer-based methods has made possible the accumulation, storage and processing of large sets of data. Micro computers are now considered an essential part of any archaeological activity. In addition, developments in statistical software have resulted in techniques for the analysis of large data sets.

Thus, the generation of more and more systematic data in archaeology and the necessity to store and evaluate them have fed on each other and created a need for methods (tools and techniques) for the analysis end evaluation of archaeological data. However, these methods shape both the questions and the design of archaeological research. Therefore, as new approaches become available and accessible, an assessment of their rôle and utility becomes of principal importance. The task that they can perform, their potential and their limitations, all have to be considered.

This paper explores the rôle of two computer-based methods used in archaeological research as formal approaches to the analysis of archaeological data. The two methods which will be under discussion are data analysis and expert systems.

29.2 Definitions

Data analysis can be defined as a set of mathematical/statistical procedures, generally used as computer programs, embracing elementary but particularly multidimensional statistical techniques that require an iterative application in order to statistically process the data and extract information from the data set. This method involves the use of mathematical/statistical rules generally applicable and not subject dependent as procedures for the assessment of data and the acquisition of new information.

An **Expert system** is a problem-solving computer program utilizing a set of prescriptions or 'rules' that provide a model for the reasoning and thus for the solution of a specific type of problem in a particular knowledge domain against which a data base pertaining to the same knowledge domain is evaluated. When dealing with this method it is necessary to distinguish two different aspects of it: the development of an expert system and its use.

1. **The development of an expert system** involves the construction of rules from the already existing knowledge describing the solution of a specific type-problem and the coding of those rules into a computer program.
2. **The use of an expert system** involves the comparison and assessment of information regarding a particular case study against the solution model established by the constructed rules.

Therefore, an expert system method provides a model for reasoning based on existing knowledge in a specific field (i.e. it is subject dependent) for the assessment of data regarding a case study.

29.3 Data analysis method

In general terms, the most important characteristic of the data analysis method is that it represents essentially an exploratory mode of analysis. The examination/exploration is done on the entire data set and an iterative manner of analysis allows for a relatively flexible way of determining the behaviour that best describes the data in question. Disagreements with the postulated general mathematical behaviour are easily visible and thus attention is drawn to 'misfit' or new information. This capacity to highlight new information makes this method one of the most useful tools for analysis in scientific research.

Data analysis procedures utilize numerical data (continuous, categorical or nominal) or data that have been coded into some form of numerical data. Since the principal rôle of the data analysis method is to explore an entire set in order to extract a pattern of behaviour that the data exhibit, the process starts first with the examination of the initial (input) data set. This set is tested for its validity and quality and then it is screened to determine if there is any underlying structure of distribution properties for the data, such as normal, logarithmic, non-parametric, *etc.* On the basis of this information and considering the type of problem that is of archaeological interest, a particular data

analysis technique (protocol) is chosen to yield a general model against which the data is examined.

For instance, when dealing with a data set describing several groups of archaeological ceramics in terms of stylistic and compositional variables, and wanting to establish a way of characterizing those groups and discriminating amongst them, an initial set would first be screened for each variable of each group to establish its range of values and its variance. Correlations between variables would then be examined. Finally, distribution properties of each variable would be determined. On the basis of established characteristics, and in view of the problem, a suitable discriminant analysis procedure would be used to characterize the groups, determine if these groups can be separated from one another, and establish criteria on which the groups might be distinguished.

The exploration and examination of the data is done in an iterative manner that allows for a refinement of the description of the behaviour of the data and the extraction of additional and new information from the data (including a rejection of the proposed behaviour).

For example, for the characterization and discrimination of ceramic groups, various parameters used in characterizing and discriminating between groups would be tested for their validity and efficiency as well as various models for group definitions. In such a process, information about ceramic groups is obtained and refined.

As a method, data analysis has been used in archaeological research for over two decades (Djiandjian forthcoming a). It is criticized principally because it relies on numerical information, because it imposes a relatively rigid, mathematical/statistical structure on the analysis, (Doran 1986) and because it is general (i.e. knowledge free) (Doran 1989). These aspects of the data analysis method are sometimes considered ill-suited to the nature of archaeological investigation. In response to some of these reservations, recent work has shown that as long as non-numerical descriptors can be unambiguously defined, the coding and subsequent use of this type of data do not represent a problem for the application of data analysis techniques (Djiandjian forthcoming b). Nevertheless, the inability of data analysis method to handle certain type of information, such as description of structures, remains. In terms of the considerations of the distribution properties, linking the distribution properties of the input data to the choice of the evaluative technique may have considerable practical merit. If the data follows a certain type of distribution behaviour, then the models built on this type of behaviour are more likely to describe the archaeological reality with greater veracity. Also, today there are more and more data analysis techniques that use non-parametric approaches and require no assumptions about the distribution properties of the data set under investigation. The fact that this method is general rather than domain specific, can equally represent an advantage.

The use of data analysis method, however, requires considerable expertise in mathematics/statistics as well as a knowledge of the archaeological problem.

29.4 Expert system method

29.4.1 Development of an expert system

This phase in the expert system approach involves the structuring of existing knowledge regarding a specific subject (problem) into a series of rule-like statements that provide a model of the task that the system is to perform. According to Reichgelt and van Harmelin (1986), expert systems can perform four single, primitive tasks: classification, monitoring, design, and simulation (or any combination of those tasks), and the formalism employed for expressing the constructed rules is to a certain degree task dependent. Expert systems can be applied to both numerical and non-numerical data as well as to incomplete data, an advantage when dealing with archaeological information.

For example, when considering the problem of interpreting archaeological sites based on site finds (artifacts and ecofacts) and site features (man-made or natural such as walls, ditches, pits, *etc.*), a set of IF. . . THEN. . . statements (rules) is constructed for each type of find and type of feature (and a combination of finds and features) that defines the activity of the site at the time of its occupation (and thus provide a cultural interpretation of the site) in terms of the finds and features considered (Patel & Stutt 1988).

This phase of the expert system approach forces the researcher to fully resolve and expose the logic of the arguments (rules) used. As a consequence, it can be extremely valuable for those who perform it as a means of structuring, standardizing and documenting their reasoning. At the same time, however, it is extremely time consuming.

Construction of an expert system requires an expertise in the field of application, as well as in computer science.

29.4.2 Use of an expert system

This phase of an expert system approach involves the comparison and assessment of a data set pertaining to a specific case study against the reasoning model established by the constructed rules. When using an expert system, an input data set is not generally initially screened for its validity or for its structure in a way comparable to the first step in data analysis but, rather, data are directly tested against the expert system.

For instance, when using an expert system for the interpretation of archaeological sites on the basis of site finds and features, data on site finds and features characteristic of a specific site are compared to the IF. . . THEN. . . rules and conclusions about the activity of the site and its cultural interpretation are reached on the basis of the stated rules.

The principal focus of an expert-system-based analysis is on the agreement of the data with the prescribed model rather than on the exploration of the data themselves.

Mathematical/statistical procedures can be incorporated into an expert system but then they take a form of rule-like statements and the rôle of these procedures ceases to be exploratory as is the case in data analysis.

Finally, the use of an expert system, at least in principle, requires no particular expertise.

Expert systems as tools are relatively recent additions to research methodology. Applications of expert system methodology in archaeology are few in number and the

rôle of the expert system approach in archaeology has been variously assessed (Wilcock 1985, Huggett & Baker 1985, Doran 1986, Doran 1988, Gardin *et al* 1987, Lagrange 1988, Vitali & Lagrange 1988, Vitali 1989). The principal general criticism refer to their inability to handle non-monotonic logic and their way of handling incomplete and uncertain information, all which are of importance to archaeology. Perhaps, with further research and the development of expert systems in general, better understanding and solutions to the issues raised by these criticisms will be found.

29.5 Conclusions

Therefore, it would appear that the two methods being considered here have different rôles in a research process.

Data analysis, as a method, employs generally applicable mathematical/statistical techniques in order to explore a data base under consideration. It is essentially an inductive method in a sense that its principal contribution is in deriving a description of the behaviour of the data and thus a description of the problem under study.

An expert system provides a model for the reasoning and the solution of a specific problem-type against which a data base is evaluated and, thus, its application is deductive in nature. Its principal contributions are first, in structuring the operational knowledge in a specific field and thus imposing the development of a coherent and formalized framework on the field under study, and second and more importantly, in performing routine analysis tasks in a standardized way. However, the quality of these analysis depends on the quality of the expert system and the previously established domain-specific knowledge.

Thus, if a research process is described as a series of steps that begins with a question, continues through the acquisition of materials, and then through the assessment of acquired materials to a final proposition, the data analysis method occupies a place in the assessment of the materials (evidence). The expert system method, on the other hand, takes up the rôle of structuring the final propositions and of performing the routine solution tasks and, thus, does not play a part, strictly speaking, in the research process itself. Both formal methods, as long as they are given their proper rôle, have their place in the archaeological reconstruction of the past.

Acknowledgements

The author is grateful to F. Djindjian, J. Doran, U. M. Franklin, J.-C. Gardin, M.-S Lagrange and H.-P. Francfort for their comments on the earlier version of the manuscript. The funding for the project, of which this article is one result, obtained through CNRS — UPR 315, Paris, France and through L. D. Levine, Royal Ontario Museum, Toronto, Canada, is gratefully acknowledged.

Bibliography

CARR, C., (ed.) 1985. *For Concordance in Archaeological Analysis: Bridging Data Structures, Quantitative Technique, and Theory*. Westport Publishers, Inc., Fayetteville. in cooperation with the Institute for Quantitative Archaeology, University of Arkansas.

DJINDJIAN, F. forthcoming a. "Data Analysis in Archaeology", *Science and Archaeology*. forthcoming .

DJINDJIAN, F. forthcoming b. *Methodes pour l'Archaeologie*. Société Prehistorique Française, Paris. forthcoming .

DORAN, J. 1986. "Formal methods and archaeological theory: a perspective", *World Archaeology*, 18: 21–37.

DORAN, J. 1988. "Expert systems and archaeology: what lies ahead?", *in* Ruggles & Rahtz 1988, pp. 237–241.

DORAN, J. 1989. "Personal communication".

DORAN, J. E. & F. R. HODSON 1976. *Mathematics and Computers in Archaeology*. Harvard University Press, Cambridge.

GARDIN, J. C., O. GUILLAUME, P. Q. HERMAN, A. HESNARD, M. S. LAGRANGE, M. RENAUD, & E. ZADORA-RIO 1987. *Systemes experts et sciences humaines: Le cas de l'archeologie*. Eyrolles, Paris.

HUGGETT, J. & K. BAKER 1985. "The Computerized Archaeologist: The Development of Expert Systems", *Science and Archaeology*, 27: 3–7.

LAGRANGE, M. S. 1988. "Expert Systems in Archaeology and social sciences: A User's View", *Archeo-log*, 3: 7–22.

ORTON, C. 1980. *Mathematics in Archaeology*. Collins, London.

PATEL, J. & A. STUTT 1988. "KIVA: an archaeological interpreter", Technical Report 35, Human Cognition Research Laboratory.

RAHTZ, S. P. Q., (ed.) 1988. *Computer and Quantitative Methods in Archaeology 1988*, International Series 446, Oxford. British Archaeological Reports.

REICHGELT, H. & F. VAN HARMELIN 1986. "Criteria for Choosing Representation Languages and Control Regimes for Expert Systems", *The Knowledge Engineering Review*, 1: 2–17.

RUGGLES, C. L. N. & S. P. Q. RAHTZ, (eds.) 1988. *Computer and Quantitative Methods in Archaeology 1987*, International Series 393, Oxford. British Archaeological Reports.

VITALI, V. 1989. "Expert Systems in Archaeometry", *in* Farquhar, R. M., Hancock, R. G. V., & Pavlish, L. A., (eds.), *Proc. 26th International Archaeometry Symposium, University of Toronto, 16–20th May 1988*.

VITALI, V. & M. S. LAGRANGE 1988. "VANDAL: an Expert System for the Provenance Determination of Archaeological Ceramics based on INAA Data", *in* Rahtz 1988, pp. 369–375.

WILCOCK, J. 1985. "A Review of Expert Systems: Their Shortcomings and Possible Applications in Archaeology", *in Computer Applications in Archaeology 1985*, pp. 139–144, London. Institute of Archaeology, University of London.