A Comparison of Monothetic Divisive and Polythetic Agglomerative Classifications of Archaeological Data

J.D.Wilcock

Research Centre for Computer Archaeology
North Staffordshire Polytechnic.

Abstract

The paper explores the essential differences between the monothetic divisive and polythetic agglomerative approaches to automatic classification. It is deduced that the monothetic method is capable of producing an infinite number of "classifications"; but that by careful choice of weighting for presence as opposed to absence and of weights for individual attributes the monothetic method may be made to produce similarly acceptable results to the polythetic method, with the added advantage that the specific attributes responsible for the discrimination between groups are identified at all stages of the process.

Introduction

The polythetic agglomerative approach to automatic classification has long been accepted in archaeological studies. The method is "polythetic" because all attributes are used simultaneously in the calculation of similarity measures between pairs of archaeological entities (artefacts, complete assemblages, cultures, etc.), and "agglomerative" because all entities start off distinct and similar items are gradually formed into groups until all the entities finally merge into a single large group. Various methods of forming the links between similar items have been employed, such as single link, double link, multiple link and average link. It is not the purpose of this paper to discuss the advantages and disadvantages of these linkage methods; suffice it to say that the average link method gives results which are generally the most acceptable from an archaeological point of view, and this method has been used for the current study.

The monothetic divisive approach, on the other hand, starts with all entities in a single large group, dividing and subdividing at each stage of the analysis according to the values of a single selected presence/absence (binary) attribute, until all items are distinct. The attribute chosen at each stage must be one that has not previously been used to define the subgroup under consideration, and some criterion is used to select the attribute which partitions the subgroup most efficiently to give the "best" archaeological results, from among those attributes remaining at each stage. This method has not been used a great deal in archaeology, for the results, while logical, have not in general provided the bases for practical classification systems. This paper shows that the monothetic divisive method may be modified to generate a potentially infinite set of "classifications", some of which produce similarly acceptable results to the polythetic agglomerative analysis of the same data, with the added advantage that the specific attributes responsible for the discrimination between subgroups are identified at all stages of the process.

Criteria for attribute selection

The analysis may be applied to any set of archaeological data where entities are described solely in terms of the presence (binary 1) or absence (binary 0) of a number of defined attributes. Thus Iron Age pits may be

described in terms of the presence/absence of animal bones, human bones, metallic objects, quern stones, sling stones, pottery, etc., and divided into types and subtypes. Assemblages from Palaeolithic to Industrial Archaeology may be so described; the specific example is taken from a collection of electrical insulators comprising 58 distinct types of insulator developed between 1840 and 1877, chosen because of its large number of attributes (55 features were identified as being significant or potentially significant in the design of the insulators, and the analysis was made easier because the actual working drawings and patent records were available). Features were divided into those describing the shape of the insulators, the design of the *sheds* which provided the chief insulation property, other functional features, the attachment of the insulator to its support, the attachment of the electrical wire, and the materials employed. For simplicity the subset describing the shed design is given in Table 1 and illustrated further in Figure 1.

The polythetic agglomerative analysis of these data using the average-link weighted pair group method is shown by the skyline plot of Figure 2. Note that the chief discriminating attributes appear to be numbers 1-3, which are concerned with the number of sheds. Next most important are the shed hem attributes (numbers 9 and 10), while the shed relative length attributes (numbers 6 and 7) occur to a lesser extent. Attributes 4,5 and 8 do not appear to be useful in the polythetic classification, while 11 is important only for group 31. It is therefore apparent that some attributes will be far more important than others, and a monothetic classification which treats all attributes as being of equal weights cannot be expected to give comparable results to a polythetic classification.

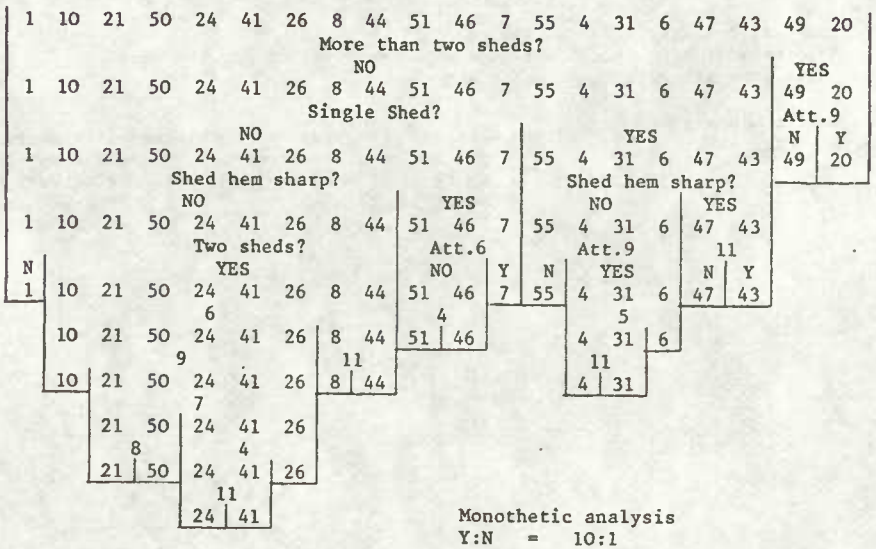
The next thing to realise is that archaeologists subconsciously attach far more importance to the *presence* of an attribute than to its *absence*. However, the conventional monothetic analysis logically treats absence as of equal significance to presence for all attributes. It is not surprising therefore that the monothetic method does not produce results which are "useful" in an archaeological sense, for artefacts are just as likely to be classified according to the absence of attributes as to their presence (in an extreme case an artefact can be classified such that all the selected attributes are *absent*), and this does not strike the archaeologist as being useful. For example, a monothetic analysis of the data of Table 1 with equal attribute weights and with absence treated as of equivalent weight to presence results in artefact 49, a highly distinctive 3-shedded insulator, being discriminated as "*not* 2-shedded", "*not* shed hem not sharp", "*not* single-shedded" and only at the lowest level of discrimination as "more than two sheds". The attributes treated as absent in such analyses could well be irrelevant to the artefact under consideration. Thus another useful modification of the monothetic method would be to weight the *presence* of an attribute more highly than its *absence*.

By different choices for the weights of individual attributes relative to one another, and different choices for the weight of presence of any attribute relative to absence (of the same attribute) an infinite number of monothetic "classifications" may be derived. The question "which weights are the correct ones?" may be answered "those which give a result which corresponds with the polythetic analysis". For the data of Table 1, a weight ratio of 10:1 for presence relative to absence, and weights 10 for attributes 1-3 (number of sheds), 5 for 9 and 10 (shed hem attributes), 2 for 6 and 7 (shed relative lengths) and 1 for the remainder gave a monothetic analysis identical with the polythetic analysis (Table 2). The criticism that these weights have been chosen subjectively may be answered by pointing out that the weights could be allocated automatically according to the lowest phenon level at which they are used in the polythetic analysis. From Figure 2, attribute 2 is last used at

TABLE 1

Artefact or Group number	Number in group	Attribute 12	345	678	9 10 11
1	9	00	000	000	000
49	1	00	100	000	010
20	2	00	100	000	100
10	1	01	000	000	000
21	1	01	000	000	100
50	1	01	000	001	100
51	1	01	000	010	010
24	6	01	000	010	100
41	1	01	000	010	101
7	3	01	000	100	010
8	1	01	000	100	100
44	1	01	000	100	101
46	1	01	010	010	010
26	1	01	010	010	100
55	1	10	000	000	000
47	1	10	000	000	010
43	1	10	000	000	011
4	18	10	000	000	100
31	6	10	000	000	101
6	1	10	001	000	100

TABLE 2



Monothetic analysis
Y:N = 10:1
Attribute weights:

1 - 3	10
9, 10	5
6, 7	2
others	1

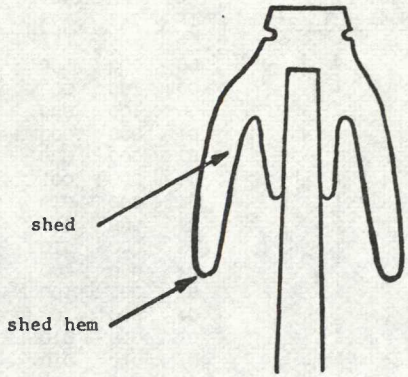


Figure 1

Attribute List

1. Single shed)
2. Two sheds) Number of sheds
3. More than two sheds)

4. Separable sheds)
5. More than two parallel-sided segments) Minority features

6. Inner shed longer than outer shed)
7. Outer shed longer than inner shed) Shed lengths for two sheds
8. Inner shed same length as outer shed)

9. Shed hem not sharp)
10. Shed hem sharp) Shed hem (although complements, both these attributes are included, absence of both indicating that there is no shed hem, and presence of both being impossible)

11. Inverted shed

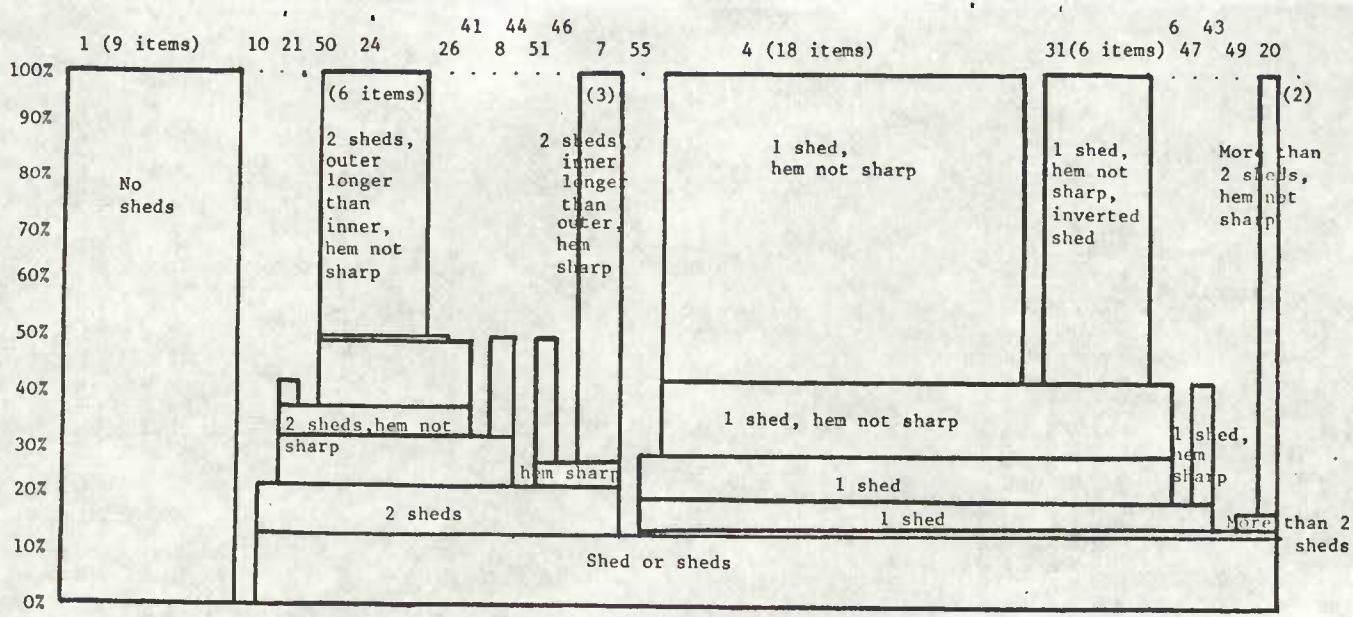


Figure 2. Skyline plot for polythetic agglomerative grouping of the data, employing average link, with common attributes as deduced from the corresponding monothetic divisive analysis.

the 13% phenon, 1 and 3 at 14%, 9 at 18%, 10 at 20%, 6 and 7 at 27%, and the remainder at much higher levels. Thus the lower the final phenon level for an attribute, the higher is its weight. A presence/absence ratio of about 10:1 seems to give good results, and suitable discrimination formulae are given in Appendix 1.

Conclusion

It has been shown for one set of archaeological data that the monothetic divisive method may be made to produce an identical classification to that produced by the polythetic agglomerative method which hitherto has been more acceptable archaeologically. With further development in the objective allocation of attribute weights there seems to be no reason why the monothetic method may not be applied to any archaeological presence/absence data, carrying with it the advantage that the specific attributes responsible for discrimination are identified at all stages.

Research Centre for Computer Archaeology
North Staffordshire Polytechnic
Blackheath Lane
STAFFORD
ST18 OAD
England

APPENDIX 1

Discrimination formulae

Let A = number of absences for the current attribute

P = number of presences for the current attribute

W = weight of presence with respect to absence

C = weight of current attribute

N = number of artefacts in current set to be partitioned

If $A \leq WN/(W+1)$ then the usefulness of the current attribute in partitioning the set of N artefacts is given by

$$D = CA\ ((W-1)N/(W+1)+P)$$

and if $A > WN/(W+1)$ by

$$D = CW^2(A-(W-1)N/(W+1))P$$

The result of these formulae is to bias the selection of an attribute so that the "presence" subset is smaller than the "absence" subset for $W > 1$, thus attaching greater importance to presence. Attributes with large C also receive preferential selection. The attribute with the largest value for the discrimination function D is selected at each stage of the analysis. The discrimination function has a maximum at

$$A = WN/(W+1)$$