Hierarchical grouping methods and stopping rules: An evaluation*

R. Mojena

Department of Management Science, University of Rhode Island, Kingston, Rhode Island 02881, USA

Hierarchical clustering procedures have received a great deal of emphasis in recent years, yet research has lagged in their empirical evaluation and in objective means to aid the user in selecting good partitions (rather than good hierarchies). The present paper aims to correct both of these deficiencies, first by empirically testing selected methods which have become popular and, second by proposing and evaluating statistical stopping rules. Results indicate: 1. that methods vary widely in their performance and 2. that the proposed stopping rules can aid the user in selecting partitions.

(Received February 1975)

In recent years cluster analysis in general and hierarchical procedures in particular have enjoyed increasing emphasis in multivariate data analysis. Agglomerative techniques have been especially popularised and extended by Ward (1963), Lance and Williams (1967), Johnson (1967), Wishart (1969), and Hubert (1972, 1973). Surprisingly, however, little effort has been expended in evaluating these methods empirically, the notable exceptions being Cunningham and Ogilvie (1972) and Blashfield (1976).

In general, hierarchical methods apply a *routing* strategy to reproduce *hierarchical* structure in the data, a primary objective in taxonomic studies. In many studies, especially in business and economics, the primary objective is to reproduce *underlying* structure by applying a *clustering* strategy. Although hierarchical methods are specifically designed for the former objective, their use in partitioning is not without precedence, for several reasons: their methodology is conceptually straightforward; computer coded algorithms are readily available; and they purportedly find good, but not necessarily optimal, partitions, e.g. Ward's (1963) method for minimum error sum of squares.

If underlying stucture is defined in terms of identifying sampled items as coming from specific multivariate populations, then the use of a hierarchical method requires a decision regarding the stage (level) which best satisfies or reproduces the underlying structure. In this context, a stopping rule is desired which selects the 'best' number of clusters based on the distribution of a clustering criterion associated with each hierarchical level.

This communication reports on a study which utilised carefully conceived data of known structure to test two statistical stopping rules and the seven hierarchical methods evaluated by Cunningham and Ogilvie (1972).

Hierarchical models

Agglomerative hierarchical clustering models form an initial partition of N clusters (each object is a cluster) and in a stagewise manner proceed to reduce the number of clusters one at a time until all N objects are in one cluster. In the first stage, $N-1$ clusters are formed by enumerating the N taken two at a time possible fusions and selecting the one which optimises the chosen criterion; in the second stage $N-2$ clusters are formed in a similar manner and so on.

All hierarchical models can be characterised by the set of partitions $(P_0, P_1, \ldots, P_{N-1})$ and their corresponding criterion values $\alpha_0, \alpha_1, \ldots, \alpha_{N-1}$, where the stages (subscripts) $0, 1, \ldots, N-1$ correspondingly represent $N, N-1, \ldots, 1$ clusters. For partition P_j the associated configuration of clusters can be represented by $C_1, C_2, \ldots, C_{N-j}$.

Differences among hierarchical methods are due only to the manner in which α is defined. The seven best known hierarchical strategies are based on the generalised transformation model of Lance and Williams (1967) and the extension by Wishart (1969) to include Ward's method:

$$d_{is} = a_p d_{ps} + a_q d_{qs} + b d_{pq} + g|d_{ps} - d_{qs}| \qquad (1)$$

where i represents the index for the new group obtained by fusing groups p and q; d represents the measure of association; s represents a group other than the fused group; and a_p, a_q, b, and g represent parameters whose values depend on the method for defining new associations at each stage (**Table 1**). If a large value for association implies dissimilar groups, then

$$\alpha_j = \min_{i<m} [d_{im}],\ i, m = 1, \ldots, N-j;$$

otherwise,

$$\alpha_j = \max_{i<m} [d_{im}],\ i, m = 1, \ldots, N-j.$$

Stopping rules

Statistical rules for selecting the partition which 'best' approximates the underlying populations can be based on the distribution of the criterion (α) or a suitable transformation of the criterion. If a large association value implies dissimilar entities (case A), the distribution of α is monotonically increasing for Methods 1 to 4 and 7; otherwise (case B), α monotonically decreases for Methods 1 to 4 and 7. It follows that a 'significant' change in α from one stage to the next implies a partition which should not be undertaken. Two stopping rules are proposed which explicitly define what is meant by a 'significant' change in the clustering criterion. In both cases the rules are operational rather than representative of some hypothesised or derived density function.

Rule One

This rule utilises the $N-1$ items in the distribution of α by calculating the mean and standard deviation of the sample and proceeds to define a 'significant' α as one which lies in either the upper tail (case A) or the lower tail (case B) of the distribution.

A precise statement of the rule for case A follows: Select group level corresponding to the first stage $j, j = 1, \ldots, N-2$ satisfying

$$\alpha_{j+1} > \bar{\alpha} + k s_\alpha, \qquad (2)$$

where α_{j+1} represents the value of the criterion in stage $j+1$; k is the standard deviate; $\bar{\alpha}$ and s_α are, respectively, the mean and unbiased standard deviation of the α distribution. For case B, the sense of the inequality is reversed.

If no value for α satisfies the above inequality, then the

*Partially funded by a Grant-in-Aid from the University Research Committee of the University of Rhode Island. The author appreciates the efforts of the referee in providing helpful comments for the revision of the original draft.

Table 1 Grouping methods and parameters associated with equation 1

Method	a_p	a_q	b	g
1. Nearest neighbour (Johnson's min)	$\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$
2. Farthest neighbour (Johnson's max)	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$
3. Simple average	$\frac{1}{2}$	$\frac{1}{2}$	0	0
4. Group (weighted) average	n_p/n_t	n_q/n_t	0	0
5. Median	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{4}$	0
6. Centroid	n_p/n_t	n_q/n_t	$-n_pn_q/n_t^2$	0
7. Ward's error sum of squares	$(n_s + n_p)/(n_s + n_t)$	$(n_s + n_q)/(n_s + n_t)$	$-n_s/(n_s + n_t)$	0

decision maker must choose: (*a*) one cluster; (*b*) the stage j for which the stage $j + 1$ yields the largest standard deviate, i.e. the stage $j + 1$ for which $(\alpha_{j+1} - \bar{\alpha})/\vartheta_\alpha$ is a maximum; or (*c*) some other appropriate heuristic rule, e.g. the first standard deviate in the highest group when the standard deviates have been resolved into three groups of values—low, intermediate, and high. For this rule, choices (*a*) and (*b*) are identical for monotonic α_j, since $\bar{\alpha}$ and ϑ_α are fixed; however, choices (*a*) and (*b*) may differ when $\bar{\alpha}$ and ϑ_α are stochastic, as in the rule below.

Rule Two

The behaviour of α is not unlike that of a time series with trend (except for Methods 5 and 6), which suggests that a more desirable approach is to formulate a one-period forecasting model which defines a 'significant' value of α in stage $j + 1$. The proposed forecasting model is a moving average corrected for linear trend lag. Precisely stated for case A, the partition to be selected is the one corresponding to the *first* stage j, $j = n, n + 1, \ldots, N - 2$ satisfying

$$\alpha_{j+1} > \bar{\alpha}_j + \mathscr{L}_j + b_j + k\vartheta_j \,, \qquad (3)$$

where n is the number of items in the moving average; $\bar{\alpha}_j$ is the moving mean in stage j; L_j is the correction for trend lag in stage j; b_j is the *moving* least squares slope in stage j; k is the standard deviate; and s_j is the moving unbiased estimate of the population standard deviation in stage j. If no value for α satisfies the inequality, the default procedure is as stated in Rule One. As before, the rule for Case B reverses the sense of the inequality.

Note that $\bar{\alpha}_j + \mathscr{L}_j$ represents the expected value of α_j and $\bar{\alpha}_j + \mathscr{L}_j + b_j$ represents the expected value of α_{j+1}. Also ϑ_j is not corrected for lag since it can be proved that, for a ramp, the expected value of ϑ_{j+1} is identical to the expected value of ϑ_j, e.g. see Mojena (1971).

Statistically, this rule is superior to the preceding rule, as candidates for 'significance' do not bias the sample statistics. The following formulas for the moving least squares slope and the correction for trend lag can be derived easily, e.g. see Mojena (1971):

$$b_j = \frac{6\left[\sum_{i=j-n+1}^{j} w_i\alpha_i - (n+1)\sum_{i=j-n+1}^{j}\alpha_i\right]}{n(n^2-1)} \qquad (4)$$

$$\mathscr{L}_j = (n-1)\,b_j/2 \,, \qquad (5)$$

where n and α are as before and $w_i = w_{i-1} + 1$, $i = j - n + 2, \ldots, j$, given that $w_{j-n+1} = 1$.

Data matrices

Based on the conceptualisation of clusters as compact swarms in Euclidian space, twelve data matrices of N entities, five attributes, and K populations were generated randomly according to an experimental design which controls two key parameters: the number of underlying populations and the degree of statistical 'overlap' among the populations. The latter variable is of special importance because the degree of homogeneity among the populations directly affects the ability of a clustering procedure to discriminate samples (clusters) from the populations.

Gamma process generators were used to simulate 30 items per cluster based on five orthogonal attributes. Gamma *pdf's* were chosen because of their shape flexibility. The size for each cluster represents a compromise between computer costs (time and storage) and sampling error. Samples of size 30 afford a maximum error of 17% at 90% confidence in estimating the means of populations.

A lower bound for the probability that an item from population i will *not* be classified in population j is given by

$$L_{ij} = \prod_{m=1}^{5} G_{pim}(X^*_{jim};\, \mu_{im},\, \sigma_{im}) \qquad (6)$$

where $G_{1im}(X_{im};\, \mu_{im},\, \sigma_{im})$ is the gamma distribution function for the m^{th} attribute (X) in population i with mean μ_{im} and standard deviation σ_{im}; G_{2im} is the inverse distribution function $(1 - G_{1im})$; $p = 1$ if $\mu_{im} < \mu_{jm}$; $p = 2$ if $\mu_{im} > \mu_{jm}$; $G_{pim} = 1$ if $\mu_{im} = \mu_{jm}$; and X^*_{jim} represents the midpoint between the centroids of populations i and j as projected on the $_m{}^{\text{th}}$ coordinate axis.

In effect, low values for L_{ij} imply that a high proportion of items which are sampled from population i will lie on the far side of a bisecting hyperplane between the centroids of populations i and j. For K populations, a measure which expresses the average potential for misclassifying items is given by

$$M = \frac{\sum_{i,j}(1 - L_{ij})}{K(K-1)} \,. \qquad (7)$$

High values for M imply high overlap among populations (i.e. a high probability that items will be misclassified by a clustering procedure), and vice versa.

Means and standard deviations for each attribute in each population (μ_{im} and σ_{im}, $i = 1, \ldots, K$, $m = 1, \ldots, 5$) were randomly generated and specific combinations yielding low to high values of M were selected. Characteristics of the twelve

Table 2 Description of data sets

Number	*Populations* (K)	*Items* (N)	*Overlap**
1	2	60	1
2	2	60	2
3	2	60	3
4	2	60	4
5	3	90	1
6	3	90	2
7	3	90	3
8	3	90	4
9	4	120	1
10	4	120	2
11	4	120	3
12	4	120	4

*The numbers (1, 2, 3, 4) denote increasing order of average overlap among items in different populations according to Equation (7).

Table 3 Conditional matching coefficients

Data Set	Method 1	2	3	4	5	6	7
1	0·492	1·000	1·000	1·000	1·000	0·492	1·000
2	0·492	0·559	1·000	1·000	0·497	0·967	0·967
3	0·492	0·492	0·492	0·492	0·492	0·492	0·935
4	0·492	0·492	0·492	0·492	0·492	0·492	0·790
5	0·342	0·739	0·768	0·768	0·768	0·768	0·957
6	0·342	0·683	0·652	0·584	0·389	0·342	0·791
7	0·342	0·543	0·464	0·342	0·359	0·357	0·710
8	0·342	0·603	0·342	0·408	0·581	0·350	0·719
9	0·270	0·555	0·579	0·591	0·666	0·595	0·849
10	0·270	0·636	0·382	0·580	0·714	0·580	0·833
11	0·269	0·688	0·646	0·612	0·319	0·344	0·809
12	0·278	0·652	0·327	0·278	0·269	0·269	0·716

sets of data are described in **Table 2.**

The described procedure for generating artificial sets of data, which are metric, represents an approach which is less *ad hoc* than that of Cunningham and Ogilvie (1972) and more rigorous with respect to the control of an experimental variable for overlap. This conceptualisation, however, requires the specification of the Euclidian distance measure for describing associations between any two clusters.

Evaluation of clustering results

If the underlying structure (i.e. the population identity of items) of a particular set of data is known, a clustering result based on that set of data can be evaluated directly. The evaluative measure used here, although derived differently, is identical to the one proposed by Rand (1971).

The *incidence matrix* for a given set of data is the $N \times N$ matrix of 0 or 1 cell entries: an entry of zero in cell (i, j) signifies that entities i and j are *not* in the same population (or cluster); otherwise, the cell entry is one. The simple *matching coefficient* can be used as a measure which reflects the disparity between the *known* incidence matrix for an artificial set of data and the incidence matrix given by the clustering result. Strictly working with the lower or upper triangle of the incidence matrices (main diagonals excluded), the criterion for evaluating a clustering result is given by

$$C = \frac{\sum_{i,j} c_{ij}}{N\ (N-1)/2} \qquad (8)$$

where $c_{ij} = 1$ if the corresponding cells of the incidence matrices are equal; otherwise, $c_{ij} = 0$. It follows that high values for the matching coefficient imply good clustering results, e.g. a value of unity translates as a perfect partitioning of the data set. This measure, unlike the goodness-of-fit measures used by Cunningham and Ogilvie (1972), can be used for nonhierarchical clustering models.

Results

An association vector based on standardised *Euclidian distance* was generated for each of the twelve data matrices as input to the clustering routines. To preserve the geometrical interpretations of intercluster associations, squared Euclidian distances were used for Methods 5 to 7.

Comparison of methods

Table 3 describes the performance (based on C) of each method on each data set *given* (conditional on) the underlying number of populations, e.g. the two clusters of Method 7 for the first data set exactly reconstructed the samples from the two underlying populations and the three clusters for the fifth data set matched 95·7% of the binary comparisons. **Table 4** summarises these results by method, degree of overlap, and number of populations. Several results are worth noting.

1. Mean performance tends to decrease as either the degree of overlap or the number of underlying populations increases. A three-factor (overlap, populations, method) ANOVA verified all factors as highly significant ($<0{\cdot}001$).
2. Method 7 (Ward's error sum of squares) gave a superior performance across all data sets. Matched t-tests showed very significant differences ($<0{\cdot}001$) between Method 7 and all other methods.
3. Method 1 (nearest neighbour) was significantly inferior to all other methods. The performance of this method verifies its chaining tendencies and subsequent heterogeneity in the size of clusters as discussed by Lance and Williams (1967), Hubert (1973), and Blashfield (1976), e.g. for the first four data sets the two clusters were split 59 and 1.
4. Methods 2 to 6 were not significantly different in their performances. Method 2 (farthest neighbour) was the best of this group, especially with regard to stability over a varying number of population overlap configurations. Methods 5 (Median) and 6 (Centroid) were especially volatile in the latter stages, e.g. for the first data set, Method 6 gave an excellent three-cluster configuration ($C = 0{\cdot}984$) which was not preserved at the two-cluster level ($C = 0{\cdot}492$).

Evaluation of stopping rules

Next, the stopping rules were evaluated for Ward's method (Method 7), with α defined in terms of standardised Euclidian distance. Rule One was tested using values of 2·00, 2·25, . . ., and 4·0 for k. Rule Two was evaluated for the same values of k in combination with a range of 0·1 to 0·9 in increments of 0·1 for $\mathscr{P}$, the number of items in the moving average as a proportion of the total number of objects, i.e., $\mathscr{P} = n/N$. (Card decks of the complete FORTRAN program—clustering methods and stopping rules—are available on request.) **Table 5** provides performance data for Rule One. The mean absolute deviations (MAD's) are based on the predicted number of clusters versus the known underlying number of populations.

5. In terms of the *predicted number of clusters*, values of k in the range of 2·75 to 3·50 gave the best overall results. Values for k in the range 2·00 to 2·75 resulted in predictions which were too high.
6. More importantly, with respect to *matching coefficients*, values of k in the range 2·25 to 3·50 essentially gave the same average results. In most cases, performance degrades only slightly when predictions are off (especially on the high side), *as long as the predicted number of clusters is in the same*

Table 4 Means of conditional matching coefficients

Overlap / Populations	Method 1	2	3	4	5	6	7
1	0·368	0·765	0·782	0·768	0·812	0·618	0·935
2	0·368	0·626	0·678	0·721	0·533	0·630	0·864
3	0·368	0·574	0·534	0·482	0·390	0·398	0·818
4	0·371	0·582	0·387	0·393	0·447	0·370	0·742
Populations							
2	0·492	0·636	0·746	0·746	0·620	0·611	0·923
3	0·342	0·642	0·556	0·526	0·524	0·454	0·794
4	0·272	0·633	0·484	0·515	0·492	0·447	0·802
Grand	0·369	0·637	0·595	0·596	0·546	0·504	0·840

Table 5 Performance of rule one—method 7*

Data Set	k 2·0	2·25	2·50	2·75	3·00	3·25	3·50	3·75	4·00	Number of Clusters 2	3	4	5	6	7
1	3	2	2	2	2	2	2	2	2	1·000	0·875	—	—	—	—
2	3	2	2	2	2	2	2	2	2	0·967	0·855	—	—	—	—
3	4	4	2	2	2	2	2	2	2	0·934	0·824	0·809	—	—	—
4	4	4	4	3	2	2	2	2	2	0·790	0·717	0·704	—	—	—
5	4	4	3	3	3	2	2	2	2	0·775	0·957	0·913	—	—	—
6	4	4	4	3	3	3	3	2	2	0·725	0·791	0·761	—	—	—
7	4	4	3	3	3	3	3	2	2	0·699	0·710	0·690	—	—	—
8	5	5	4	3	3	3	3	2	2	0·644	0·719	0·707	0·688	—	—
9	5	5	3	3	3	3	3	3	3	0·482	0·766	0·849	0·834	—	—
10	5	4	4	4	3	3	3	3	3	0·557	0·822	0·833	0·815	—	—
11	7	6	5	4	4	4	3	3	3	0·686	0·780	0·808	0·815	0·790	0·774
12	7	6	5	5	5	5	4	3	3	0·512	0·636	0·716	0·720	0·734	0·738
MAD	19/12	14/12	7/12	3/12	3/12	4/12	5/12	8/12	8/12						
Mean C	0·788	0·810	0·823	0·827	0·832	0·817	0·814	0·795	0·795						

*Entries in the left-hand side of the table represent the number of clusters predicted by the rule; entries in the right-hand side represent values of C for the corresponding number of clusters as given by Equation (8).

neighbourhood as the underlying number of populations, e.g. Ward's Method on data set 11 yielded matching coefficients of 0·790 and 0·808 for six and four clusters, respectively. In fact, sometimes the matching coefficient is higher when the predicted number of clusters deviates from the 'ideal' number of clusters, e.g. Model 7 on data set 12 gave 0·716 for four clusters and 0·734 for six clusters.

Table 6 provides the predicted number of clusters given by Rule Two for selected values of k and $\mathscr{P}$.

7. A value of 3·50 for k and a range of 0·70 to 0·90 for $\mathscr{P}$ yield the best overall results for this rule. As expected, the predicted number of clusters decreases as overlap among the populations increases. Furthermore, the rule was unstable (very sensitive) at low values of $\mathscr{P}$, a result which is not surprising given that the rule is based on a moving average (or equivalently an exponential smoothing) forecasting model. On the average, Rule One outperformed Rule Two.

Discussion

Based on metric data, Cunningham and Ogilvie (1972) found little difference in the performances of Methods 2 to 7. Blashfield (1976), however, concluded that Method 7 outperforms Methods 1, 2, and 4. In the present study, based on more demanding metric data sets and a different measure of goodness-of-fit, Ward's Method clearly stands out as the best. In part, however, its performance *vis-a-vis* the other methods is a self-fulfilling prophecy, for the data sets were constructed in a manner which is consistent with the conceptualisation of clusters as compact swarms in Euclidian space. Still, the poor performances of Methods 2, 5, and 6 are surprising. Farthest neighbour or Johnson max is a space dilating method which is not inconsistent with the variance definition used to define clusters in this study. For this reason, of course, it was the best of Methods 1 to 4. Its performance relative to Ward's Method can be excused on the basis that it is designed for ultrametric input data. The median and centroid methods, however, cannot be dismissed so easily, as they treat clusters in a manner which is consistent with the geometrical interpretations of this study.

The results clearly show the importance of consistency among the conceptualisation of clusters, the measure of association, the type of input data and the clustering method. Given the above conceptualisation of clusters, the availability of metric data and the condition that solutions need not be monotone invariant, Ward's Method appears to be an excellent choice.

The stopping rules appear worthy of further consideration as a pragmatic means of objectively assessing the selection of a clustering partition. Rule One, in particular, provides stable results. Since the rules require the specification of values for k and $\mathscr{P}$, it is wise to consider reasonable ranges for these parameters. For the most part the selection of an appropriate partition is aided greatly by an examination for central tendency in the predicted number of clusters. It is worth emphasising that the predicted number of clusters can vary slightly from the true number (especially as overestimations) without appreciably altering the relationship among objects, as given by the matching coefficient based on incidence matrices.

Table 6 Performance (number of clusters) of rule two—method 7

Data Set	k / $\mathscr{P}$	2·50 0·70	2·50 0·80	2·50 0·90	3·00 0·70	3·00 0·80	3·00 0·90	3·50 0·70	3·50 0·80	3·50 0·90
1		3	3	3	2	2	3	2	2	2
2		5	5	5	2	2	2	2	2	2
3		4	4	4	2	4	4	2	2	2
4		2	2	2	2	2	2	1	1	2
5		5	5	6	4	4	5	3	4	4
6		4	4	4	4	4	4	4	4	4
7		3	4	4	3	3	3	3	3	3
8		5	5	5	2	3	3	2	2	2
9		9	9	9	5	5	9	5	5	5
10		8	8	9	4	4	4	4	4	4
11		7	7	7	7	7	7	4	4	4
12		6	6	6	5	5	5	2	5	5
MAD		25/12	26/12	28/12	8/12	9/12	15/12	6/12	6/12	5/12

References

BLASHFIELD, R. K. (1976). Mixture Model Tests of Cluster Analysis: Accuracy of Four Agglomerative Hierarchical Methods, *Psychological Bulletin*, Vol. 83, pp. 377-388.

CUNNINGHAM, K. M. and OGILVIE, J. C. (1972). Evaluation of Hierarchical Grouping Techniques: A Preliminary Study, *The Computer Journal*, Vol. 15, pp. 209-213.
HUBERT, L. (1972). Some Extensions of Johnson's Hierarchical Clustering Algorithms, *Psychometrika*, Vol. 37, pp. 261-274.
HUBERT, L. (1973). Monotone Invariant Clustering Procedures, *Psychometrika*, Vol. 38, pp. 47-62.
JOHNSON, S. C. (1967). Hierarchical Clustering Schemes, *Psychometrika*, Vol. 32, pp. 241-254.
LANCE, G. N. and WILLIAMS, W. T. (1967). A General Theory of Classificatory Sorting Strategies, I. Hierarchical Systems, *The Computer Journal*, Vol. 9, pp. 373-380.
MOJENA, R. (1971). An Approach to the Problem of Evaluation in Multivariate Classification, unpublished Ph.D. dissertation, University of Cincinnati.
RAND, W. M. (1971). Objective Criteria for Evaluating Clustering Methods, *J. Amer. Stat. Assoc.*, Vol. 66, pp. 846-850.
WARD, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function, *J. Amer. Stat. Assoc.*, Vol. 58, pp. 236-244.
WISHART, D. (1969). An Algorithm for Hierarchical Classifications, *Biometrics*, Vol. 25, pp. 165-170.

Book reviews

Digital Computer Fundamentals, 4th edition, by T. C. Bartee, 1977; 563 pages. (*McGraw-Hill*, £11·20)

This is a revised edition of a now well-established text which appeared first in 1960. This edition covers recent developments such as microprocessors, floppy disc memories, large scale integrated circuits and microprogramming.

After an opening chapter on the operation of the computer, which orientates the reader, there are chapters on number systems and circuit logic. The elaboration of these principles in the design of integrated circuits is explained before the underlying electronic principles. Four chapters then deal with the next level of organisation: the arithmetic, memory, input/output and the control units. The final chapter deals with the architecture of a machine.

The student for the BCS Part I examination will find the material on computer hardware more than adequately covered in this book. The way it is presented is somewhat dry but it is helpfully organised so that the book may be read at several glances, taking in more detailed and difficult material each time. The student will find each chapter is supported by an ample number of questions for a selection of which the answers are supplied. Index, bibliography and an abundance of clear diagrams and pictures will help to commend the book to students.

R. K. STAMPER (London)

Encyclopedia of Computer Science and Technology, Volume 4, edited by J. Belzer, A. G. Holzman and A. Kent, 1976; 512 pages. (*Marcel Dekker*, SFr 285.00)

The chapters in Volume 4 deal with the following topics: brain models, branch and bound techniques, budgeting for electronic data processing, the Burroughs Corporation, California Computer Products Inc. (Calcomp), computers in Canada, cathode ray devices, Cauchy methods, character recognition systems, Chebyshev methods, chemical education, and the computer handling of chemical structures. Useful information on suitable mathematical techniques will be found in the articles on 'branch and bound techniques' (for solving optimisation problems, such as integer programming, the travelling salesman, scheduling problems, etc.) 'Cauchy methods' (for partial differential equations with given initial values, e.g. the Laplace equation, the heat equation, wave equation, etc.) and 'Chebyshev methods' (here considered mainly as a technique for approximating to a given real function by polynomials, but one which is also valuable in numerical integration and the solution of ordinary differential equations.) The budgeting chapter will help in selecting the most suitable data processing equipment and departmental structure. Other articles are mainly descriptive, and are written in a clear informative style.

CEDRIC A. B. SMITH (London)