



Hierarchical Grouping to Optimize an Objective Function

Joe H. Ward, Jr.

Journal of the American Statistical Association, Vol. 58, No. 301 (Mar., 1963), 236-244.

Stable URL:

<http://links.jstor.org/sici?sici=0162-1459%28196303%2958%3A301%3C236%3AHGT0AO%3E2.0.CO%3B2-9>

Journal of the American Statistical Association is currently published by American Statistical Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/astata.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

HIERARCHICAL GROUPING TO OPTIMIZE AN OBJECTIVE FUNCTION*

JOE H. WARD, JR.

Aerospace Medical Division, Lackland Air Force Base

A procedure for forming hierarchical groups of mutually exclusive subsets, each of which has members that are maximally similar with respect to specified characteristics, is suggested for use in large-scale ($n > 100$) studies when a precise optimal solution for a specified number of groups is not practical. Given n sets, this procedure permits their reduction to $n - 1$ mutually exclusive sets by considering the union of all possible $n(n - 1)/2$ pairs and selecting a union having a maximal value for the functional relation, or objective function, that reflects the criterion chosen by the investigator. By repeating this process until only one group remains, the complete hierarchical structure and a quantitative estimate of the loss associated with each stage in the grouping can be obtained. A general flowchart helpful in computer programming and a numerical example are included.

1. FORMULATION OF THE GROUPING PROBLEM

SITUATIONS often arise in which it is desirable to cluster large numbers of objects, symbols, or persons into smaller numbers of mutually exclusive groups, each having members that are as much alike as possible. Grouping in this manner makes it easier to consider and understand relations in large collections; hence it often increases efficiency of management. Grouping, however, ordinarily results in some loss of information that may be quantified in a "value-reflecting" number.

In earlier formulations of the grouping problem, Cox [2] and Fisher [3] were concerned primarily with grouping based on similarity with respect to a single variable. Cox treated the special case when the variable is normally distributed, and briefly discussed the two-dimensional problem. Fisher described a technique for grouping without regard for the distribution of the variable. His procedure is designed to minimize a weighted sum of squares about group means on one variable. Our formulation of the problem differs from both.

For our purposes, it was necessary to take account of the similarity of group members with respect to many variables. Our desire was to form each possible number of groups, $n, n - 1, \dots, 1$, in a manner that would minimize the loss associated with each grouping, and to quantify that loss in a form that could be readily interpreted. Since it was not feasible to obtain the many optimal solutions that would be required for large-scale studies ($n > 100$), a compromise was necessary. Therefore, we proposed (a) to reduce the number of groups from n to $n - 1$ in a manner that would minimize the loss and then, without modifying the groups formed, to repeat the process until the number of groups was systematically reduced from n to 1, if desired, and (b) to evaluate loss in terms of whatever functional relation best expressed an investigator's criterion for grouping.

* The research reported in this paper was sponsored by 6570th Personnel Research Laboratory, Aerospace Medical Division, under AFSC Project 7734, Task 773403.

Objective Function

Given a set of ratings for 10 individuals, $\{2, 6, 5, 6, 2, 2, 2, 0, 0, 0\}$, a common practice is to use the mean value to represent all the scores rather than to consider individual scores. The "loss" in information that results from treating the 10 scores as one group with a mean of 2.5 can be indicated by a "value-reflecting" number, the error sum of squares (ESS).

The error sum of squares is given by the functional relation,

$$\text{ESS} = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2$$

where x_i is the score of the i th individual. The ESS for the example is

$$\text{ESS}_{(\text{one group})} = \sum_{i=1}^{10} x_i^2 - \frac{1}{10} \left(\sum_{i=1}^{10} x_i \right)^2 = 113 - 62.5 = 50.5.$$

Similarly, if the 10 individuals are classified according to their scores into four sets,

$$\{0, 0, 0\}, \quad \{2, 2, 2, 2\}, \quad \{5\}, \quad \{6, 6\}$$

this grouping can be evaluated as the sum of the four error sums of squares,

$$\text{ESS}_{(\text{four groups})} = \text{ESS}_{(\text{Group 1})} + \text{ESS}_{(\text{Group 2})} + \text{ESS}_{(\text{Group 3})} + \text{ESS}_{(\text{Group 4})}.$$

A functional relation that provides a "value-reflecting" number of this type will be referred to here as an "objective function." In general, an objective function may be any functional relation that an investigator selects to reflect the relative desirability of groupings. In this example, the objective function is "loss of information" as reflected by the error sum of squares. The most desirable level of this objective function is its minimum value, 0.0. The objective-function values that were computed indicate the information lost when the 10 scores are treated as a single set (50.5) and as four sets (0.0). The nature of the problem and the criterion, of course, dictate the choice and interpretation of the objective function.

A number of different objective functions have been developed and put to operational use. These are exemplified in applications of the grouping procedure now programmed by Personnel Research Laboratory to accomplish the clustering of (a) persons to maximize their similarity with respect to measured characteristics, (b) jobs to minimize cross-training time when men are reassigned according to established policies, (c) job descriptions to minimize errors in describing a large number of jobs with a small number of descriptions, and (d) regression equations to minimize the loss of predictive efficiency resulting from reductions in the number of equations. In (a) the objective function is the grand sum of the squared deviations about the means of all measured characteristics. The objective function for (b) is expected cross-training time, assuming random movement among jobs in each cluster; whereas in (c) it is amount of job time incorrectly described and in (d) loss of predictive efficiency. Fortunately, the same general computer program for grouping can be used with many objective functions that appear to be different. In fact, the four

objective functions just described all lead to the use of the same program [1, 7].

In some situations, the desired number of groups can be specified in advance; in others, this is difficult to do.¹ With this grouping procedure, one need not specify the number in advance, for the objective-function value associated with any given number of groups (from n to 1) is available. Indeed, the changes in these values as the number of groups is systematically reduced furnish useful clues to the appropriate number of groups to be used for operational purposes.

Hierarchical Groups

The grouping procedure reported here is based on the premise that the greatest amount of information, as indicated by an objective function, is available when a set of n members is ungrouped. Hence the grouping process starts with these n members, which are termed groups, or subsets, although they contain only one member. The first step in grouping is to select two of these n subsets which, when united, will reduce by one the number of subsets while producing the least impairment of the optimal value of the objective function. The $n-1$ resulting subsets then are examined to determine if a third member should be united with the first pair or another pairing made in order to secure the optimal value of the objective function for $n-2$ groups. This procedure can be continued, if desired, until all n members of the original array are in one group. Since the number of subsets is systematically reduced ($n, n-1, \dots, 1$), the process is termed "hierarchical grouping" and the resulting mutually exclusive groups "hierarchical groups."

Applications of Hierarchical Grouping Procedure

Hierarchical groups, formed in the manner described, are particularly useful for classification purposes [2, 3, 4, 5, 6, 7]. They may be used, for example, to establish taxonomies of plants and animals with respect to genetic background. Such groups also permit organizing and cataloging materials, e.g., library documents, so as to facilitate the storage and retrieval of information. Similarly, hierarchical groupings of jobs may be helpful in identifying job "types" and "subtypes."

As indicated previously, the usefulness of this hierarchical approach is not restricted to classification problems. No serious loss of predictive accuracy resulted, for instance, when a hierarchical grouping of regression equations was used to select a smaller number to replace the 60 equations developed for prediction of success in Air Force technical schools. While a hierarchical arrangement may not always be needed for the "best" partitioning into groups, it is preferred in many practical situations. Even when a nonhierarchical grouping is sought, the approach described here probably will yield a good solution, although it may be one that does not optimize the objective function for the specified number of groups.

A problem often linked with the grouping problem is that of assigning or

¹ When it is desired to assign objects to k , a specified number of groups—without regard for the subdivisions of these groups—the goal is only to identify the k groups that optimize the objective function. If the objective function is a linear form, the problem can be formulated as a linear programming problem. If the objective function is nonlinear in form, the problem can be formulated as a nonlinear programming problem [2]; however certain computational difficulties must be considered.

classifying new objects into accepted groups. This problem, as well as those associated with selecting appropriate objective functions and deciding upon the number of groups to be utilized, will not be discussed here. The purpose of this article is to report a hierarchical grouping procedure believed to be of general interest since it has many applications.

2. HIERARCHICAL GROUPING PROCEDURE

Hierarchical Grouping Cycle

Optimal union of subsets $S(i, n)$. The grouping procedure starts with a universal set (U), $\{e_1, e_2, \dots, e_n\}$, consisting of n one-element subsets, i.e., well-defined objects, symbols, or persons. These are arbitrarily numbered in sequence for convenient identification in computer processing. To reduce the number of subsets to $n-1$, one new subset, which minimizes the change in the objective function's value, is formed by uniting two of the original n subsets, say

$$[S(1, n)] \cup [S(2, n)] = \{e_1, e_2\}.$$

This requires an evaluation of the objective function for each of the $n(n-1)/2$ possible unions of subsets $S(i, n)$, $i=1, 2, \dots, n$, where i refers to the number identifying the set, and the second parameter, n at this stage, refers to the number of sets under consideration. As each union is considered in turn, the value of the corresponding objective function is computed and hypothesized to be "equal to or better than" that of any preceding union. The identity of the "best" union is maintained throughout the sequence of comparisons. This facilitates identification of that union which has an objective-function value "equal to or better than" that of any of the $n(n-1)/2$ possible unions. This union is accepted as an optimal grouping when the number of subsets is reduced from n to $n-1$.

Designation of new subset and its associated objective-function value. For the purpose of printing out a display to show the sequence in which subsets are united in large-scale studies ($n=100$ to 1000), it is convenient in machine processing to give special designations to identify each new subset (resulting from acceptance of a union) and its members. Thus the union resulting in $n-1$ subsets is denoted:

$$S(p_{n-1}, n-1) = [S(p_{n-1}, n)] \cup [S(q_{n-1}, n)]$$

where

p_{n-1} = the smaller of the two numbers used to identify the subset in the n original subsets. This number is used to identify the new subset.

q_{n-1} = the larger of the two numbers used to identify the subset in the n original subsets. This number is "inactive" after it is used at this stage for the printout showing which two subsets have been united.

The objective-function value is denoted in the same manner to identify it with this union:

$$Z[p_{n-1}, q_{n-1}, n-1].$$

The term at the right, $n-1$, shows the number of subsets remaining after the union; the left-hand and center terms are the original identification numbers of the united subsets.

Optimal union of subsets $S(i, n-1), \dots, S(i, n-[n-1])$. With $n-1$ subsets, now, we have to define and consider

$$S(i, n - 1) = S(i, n)$$

when $i=1, 2, \dots, n; i \neq p_{n-1};$ and $i \neq q_{n-1}$

and $S(p_{n-1}, n-1) = [S(p_{n-1}, n)] \cup [S(q_{n-1}, n)]$

when $i = p_{n-1}$.

Selection of an optimal union to reduce the $n-1$ subsets to $n-2$ requires evaluation and comparison of the $(n-1)(n-2)/2$ possible unions in a manner analogous to that used to reduce the n subsets to $n-1$. When this has been done, the accepted union and its associated objective-function value are designated

$$S(p_{n-2}, n - 2) = [S(p_{n-2}, n - 1)] \cup [S(q_{n-2}, n - 1)]$$

and

$$Z[p_{n-2}, q_{n-2}, n - 2] \quad (p_{n-2} < q_{n-2}).$$

The identifications are maintained as before, i.e., p_{n-2} is the identification number with the smaller numerical value and q_{n-2} is that with the larger numerical value.

This grouping cycle can be continued, if desired, until all subsets have been united to form the universal set, U . At any phase in which k mutually exclusive subsets are under consideration, the objective-function value, and the union with which it is associated would be expressed as

$$Z[i, j, k-1] \text{ associated with } [S(i, k)] \cup [S(j, k)]$$

where

$$\begin{aligned} i &= 1, 2, \dots, n - 1 \\ i &\neq q_{n-1}, q_{n-2}, \dots, q_k \\ j &= i + 1, i + 2, \dots, n \\ j &\neq q_{n-1}, q_{n-2}, \dots, q_k. \end{aligned}$$

Following selection of an optimal union, this union and its corresponding objective-function value would be designated

$$S(p_{k-1}, k - 1) = [S(p_{k-1}, k)] \cup [S(q_{k-1}, k)]$$

and

$$Z[p_{k-1}, q_{k-1}, k - 1] \quad (p_{k-1} < q_{k-1}).$$

Furthermore, the elements of any subset, $S(i, k)$, would be designated

$$S(i, k) = \{e_{m_1}, \dots, e_{m_a}, \dots, e_{m_i}\}$$

where

t = number of elements in subset

m_a = identification number of a th element in the subset.

Numerical Example of Hierarchical Grouping Cycle

To illustrate the procedure, let us consider a small problem in which five individuals are to be grouped on the basis of ratings they have given one object (Figure 1). The objective-function value to be minimized is the sum of the squared deviations about the group mean (ESS). While problems like this are readily done by hand and do not require the use of identification numbers and designations, hand-processing of data is seldom economical when $n > 25$. Therefore, we give a general flowchart that may be helpful in computer programming (Figure 2). The pattern for the computational operations for the example is shown in Figure 3.

Numerical Example		Printout Displaying Hierarchical Grouping Results					
Per-son	Object Rating	Person	Number of Groups				
			1	2	3	4	5
1	1	1	1	1	1	1	1
2	7	3	3	3	3	3	3
3	2	2	2	2	2	2	2
4	9	4	4	4	4	4	4
5	12	5	5	5	5	5	5
Objective-Function Value (ESS)			86.80	13.17	2.50	.50	.00
$\Delta 1$			73.63	10.67	2.00	.50	
$\Delta 2$			62.96	8.67	1.50		

FIGURE 1. Summary of results of hierarchical grouping for numerical example.

Figure 1 illustrates the final machine printout summarizing the hierarchical grouping as it is ordinarily provided when an investigator is interested primarily in the last groups formed, although the number of groups can be arranged in the reverse order. (Machine routines to obtain this printout and the values in the $\Delta 1$ and $\Delta 2$ rows are not included in Figures 2 and 3.) This printout gives information on the complete hierarchical structure.

The values in the last three rows of Figure 1 are of particular interest. The ESS row contains the objective-function values which in this case show the error associated with each step in the grouping. Study of the hierarchical structure is facilitated by reference to the $\Delta 1$ entries between columns that

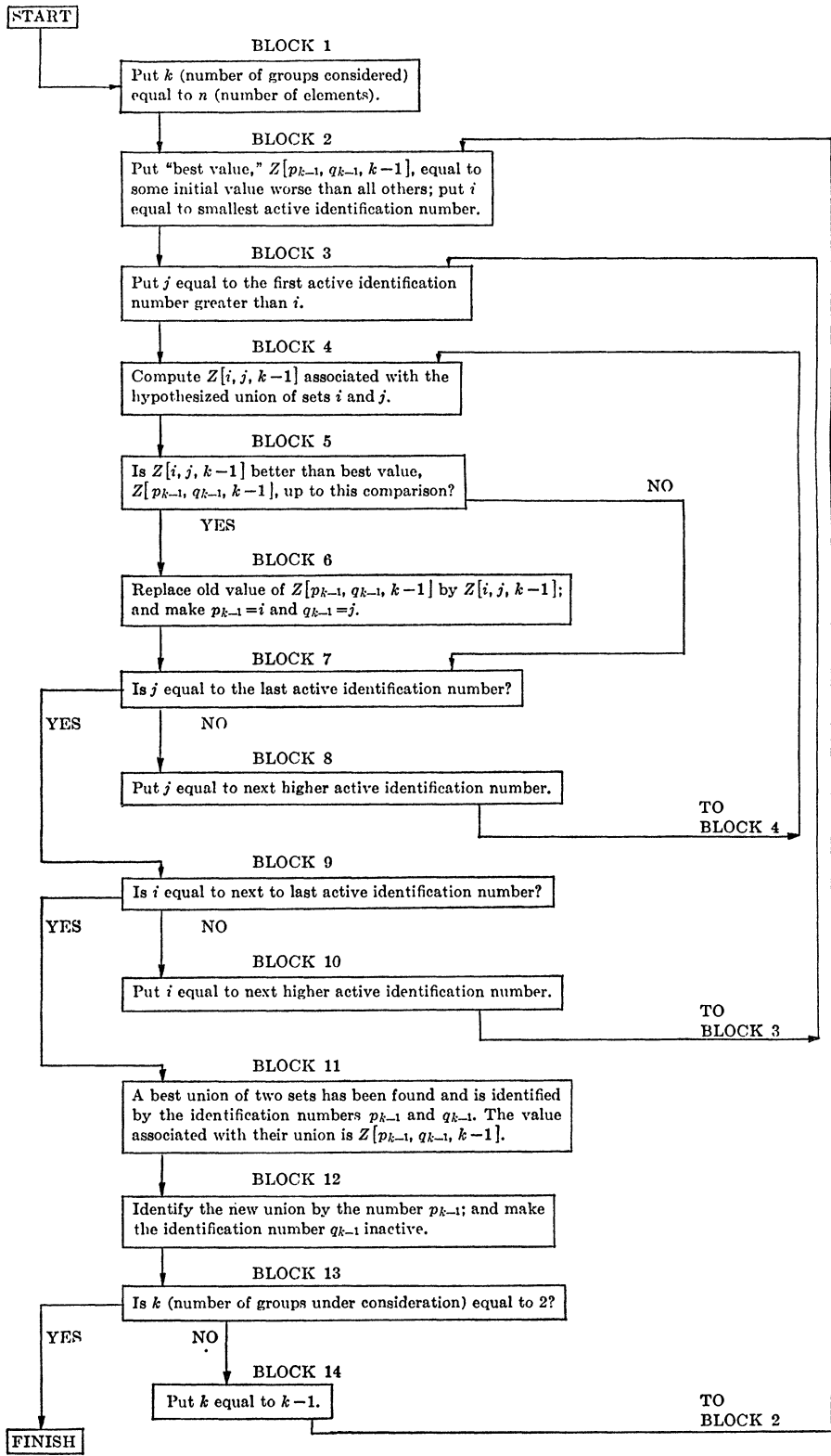


FIGURE 2. Flowchart for hierarchical grouping procedure.

Block Number	Time Through	Operation
1	1	$k = 5$
2	1	$Z[p_4, q_4, 4] = \text{a high value, say, } 100$
2	1	$i = 1$
3	1	$j = 2$
4	1	$Z[1, 2, 4] = [(1)^2 + (7)^2] - .50[(1+7)^2] = 50 - 32 = 18$
5	1	Is $Z[1, 2, 4] < Z[p_4, q_4, 4]$? i.e., Is $18 < 100$? Yes; go to Block 6
6	1	$Z[1, 2, 4]$ replaces $Z[p_4, q_4, 4]$ as temporary best
6	1	18 replaces 100 as temporary best
6	1	($i = 1$) replaces p_4
6	1	($j = 2$) replaces q_4
7	1	Is $j = 5$? i.e., Is $2 = 5$? No; go to Block 8
8	1	Set $j = \text{next higher active number} = 3$. Go to Block 4
4	2	$Z[1, 3, 4] = [(1)^2 + (2)^2] - .50[(1+2)^2] = 5 - 4.5 = .50$
5	2	Is $Z[1, 3, 4] < Z[p_4, q_4, 4]$? i.e., Is $.50 < 18$? Yes; go to Block 6
6	2	$Z[1, 3, 4]$ replaces $Z[p_4, q_4, 4]$ as temporary best
6	2	.50 replaces 18
6	2	($i = 1$) replaces p_4
6	2	($j = 3$) replaces q_4
7	2	Is $j = 5$? i.e., Is $3 = 5$? No; go to Block 8

FIGURE 3. Pattern of computational operations for hierarchical grouping for example.

show the change in the error term at each union and to the Δ_2 entries indicating the acceleration of error. Results of the grouping procedure indicate that Persons 1 and 3 gave similar ratings to the object and that the five individuals form two distinct, relatively homogeneous groups. Detailed information on specific persons and groups is easily extracted.

3. SUMMARY

A procedure has been described for forming hierarchical groups of mutually exclusive subsets on the basis of their similarity with respect to specified characteristics. Given k subsets, this method permits their reduction to $k-1$ mutually exclusive subsets by considering the union of all possible $k(k-1)/2$ pairs that can be formed and accepting the union with which an optimal value of the objective function is associated. The process can be repeated until all subsets are in one group. The computer program described also yields displays of the hierarchical structure showing the sequence in which the subsets have been united. A flowchart and numerical example are provided.

REFERENCES

- [1] Bottenberg, Robert A., and Christal, Raymond E., *An Iterative Technique for Clustering Criteria which Retains Optimum Predictive Efficiency*. Lackland Air Force Base, Texas: Personnel Laboratory, Wright Air Development Division, March 1961. (Technical Note WADD-TN-61-30).
- [2] Cox, D. R., "Note on grouping," *Journal of the American Statistical Association*, 52 (1957), 543-7.
- [3] Fisher, Walter D., "On grouping for maximum homogeneity," *Journal of the American Statistical Association*, 53 (1958), 789-98.

- [4] Tanimoto, T. T., and Loomis, Robert G., *A Taxonomy Program for the IBM 704*. New York: International Business Machines Corporation (Data Systems Division, Mathematics & Applications Department), 1960. (M&A-6, *The IBM Taxonomy Application*)
- [5] Thorndike, Robert L., "Who belongs in the family?," *Psychometrika*, 18 (1953), 267-76.
- [6] Thorndike, Robert L., Hagen, Elizabeth P., Orr, David B., and Rosner, Benjamin, *An Empirical Approach to the Determination of Air Force Job Families*. Lackland Air Force Base, Texas: Air Force Personnel and Training Research Center, August 1957. (*Technical Report AFPTRC-TR-57-5*, ASTIA Document No. 134 239)
- [7] Ward, Joe H., Jr., and Hook, Marion E., *A Hierarchical Grouping Procedure Applied to a Problem of Grouping Profiles*. Lackland Air Force Base, Texas: Personnel Laboratory, Aeronautical Systems Division, October 1961. (*Technical Note ASD-TN-61-55*)