BOOK REVIEWS

SPÄTH, Helmut: **Cluster-Analyse-Algorithmen zur Objektklassifizierung und Datenreduktion** (in German). München–Wien: R. Oldenbourg Verlag [1]1975. [2]1977. 217 p., ISBN 3-486-39912-8.

"Cluster-Analysis" (synon.: automatic classification, numerical taxonomy etc.) is a collective name designating numerous mathematical and statistical procedures for grouping a great set of objects described by a series of (numerical, ordinal, nominal or binary) variables into a small number of different classes (groups, clusters, types) such that each class is as homogeneous as possible and comprises only "similar" objects whilst on the other side, different classes should be well separated and "dissimilar" objects should be attributed to different classes. Such a classification (together with a corresponding interpretation of classes) gives more insight into the underlying structure of the set of objects; other motivations are: data reduction, segmentation for organisational purposes, formation of hypotheses etc.–

The book of H. *Späth* is an introduction in cluster analysis from the programmer's and user's viewpoint. It describes the most essential methods in a formally exact and well readable way, gives the corresponding FORTRAN computer programs (in a modular form) and illustrates the use of each program by one or more real data examples.–

Chap. 1: Introduction. Chap. 2 introduces various measures for the similarity and dissimilarity (distance) between objects for several types of variables. Chap. 3 presents exact enumerative and heuristic exchange procedures for minimizing a grouping criterion (e.g. trace criterion, determinantal criterion; dynamic programming approach). A method for solving the location-allocation problem is given. Chap. 4 comprises 9 hierarchical (divisive and agglomerative) procedures (no program for a dendrogram plot). Chap. 5 presents a method for optimizing the representation of profiles (e.g. cluster centers in multidimensional space) and a computational version of the "shaded diagram" method.–

On the basis of the given examples the author gives recommendations regarding the procedure to use. He argues for the procedure with the best interpretable result and evidently prefers the criterion minimizing procedures of chap. 3. The reviewer misses some method for reducing chaining effects (e.g. of *Wishart, Ling*) and for the simultaneous clustering of objects and variables. A table of contents is lacking. The examples are mostly two-dimensional whilst the very profit of cluster analysis lies in multidimensional applications. Some errors must be corrected (e.g.: (2.1.36); (2.2.6a) is the coefficient of *Jaccard* (comp. p. 25, line 6); (2.219); the criterion (3.3.4) is a constant lm; the splitting 3/9 is lacking in B 29).–

The book is a good guideline for people (able to read mathematical formulae) desiring an overview on cluster analysis or needing a corresponding short computer program.

H.-H. Bock

SPÄTH, Helmuth (Ed.): **Fallstudien Cluster-Analyse**. München–Wien: R. Oldenbourg Verlag 1977. 189 p., ISBN 3-486-20771-7.

Cluster analysis is a collective name designating numerous mathematical and statistical methods for dissecting a great set of objects into several small homogeneous and well separated classes (groups, clusters) and thereby using exclusively some numerically given information about properties, similarities or dissimilarities between these objects. The book "Fallstudien Cluster-Analyse" contains 11 papers of different authors which apply these methods to the solution of a practical problem (or subproblem) taken from the economical, sociological, administrational or psychological field. Generally the papers follow all the same scheme (presentation of the problem/description of data/motivation of the procedure/resulting classification/interpretation and valuation), they are easily readable and written from a non-mathematical view point. They differ considerably in their precision and their level of argumentation. I miss a real comparison between several clustering procedures whose selection is often motivated more by the available software than by substantial reasoning. The book illustrates the use of clustering methods and gives an idea of the kind of problems which can be tackled by these methods (ignoring fields like medicine, biometry, pattern recognition, data files etc.).

Contents: *D. Steinhausen/J. Steinhausen: Cluster-Analyse als Instrument der Zielgruppendefinition in der Marktforschung.* (Life-style data of 4000 persons are analysed by principal component analysis and minimal distance procedures; result: 15 groups of persons which differ by their consumer behaviour.) –

G. Blaschke/G. Liesegang: Die Klassifizierung von Nachfragekurven zur Verbesserung der kurzfristigen Absatzprognose in einem Betrieb mit modeabhängigem Produktionsprogramm. (Using a weighted euclidean distance between demand curves of textiles the hierarchical method of Ward reveals the existence of several groups of textiles; this permits the forecasting of a future demand and an improved ordering strategy based on preliminary data.) –

F. Bingemer/H.-A. Tauschwitz: Ein Modell zur Optimierung der Struktur von Absatzstellen. –

H. Späth: Partitionierende Cluster-Analyse für große Objektemengen mit binären Merkmalen am Beispiel von Firmen und deren Berufsgruppenbedarf. (An iterative exchange procedure (FORTRAN) is given for optimizing a clustering criterion with binary data; it is applied to the grouping of enterprises according to the professions in their staff.) –

W. Schläger: Die Klassifikation von Verweildauerhäufigkeiten stationärer Patienten der Medizinischen Klinik der Universität Erlangen–Nürnberg mittels Cluster-Analyse. –

W. Klösgen: Einsatz von Gruppierungsverfahren für Organisationsuntersuchungen. Cluster-Analyse und alternative Verfahren. (The sections of a ministry are grouped on the basis of their activities and their interactions; factor analysis, complete linkage method and maximal cliques are used.) –

H. Fakiner/E. Krieger/H. Rohmeier: Regional differenzierte Analyse und Prognose des Wasserbedarfs der

https://doi.org/10.5771/0943-7444-1978-2-112-1 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

privaten Haushalte in der Bundesrepublik Deutschland. (541 departments are classified into 10–15 clusters according to 31 demographic or sociological attributes (variance criterion, exchange algorithm); by cluster-specific regression methods one obtains a prognosis for the demand for water in 2000.) –

W. D. Rase/E.-M. Paech: Klassifizierung der Kreise der Bundesrepublik Deutschland nach ihrer Versorgung mit Basis-Freizeiteinrichtungen. (Departments are classified on the basis of their equipment in sportsgrounds.) –

B. Hamacher/K. Preiser: Eine Infrastrukturtypologie am Beispiel des Landes Bremen. (By using the procedures of Ward and Wishart the 78 sections of a town are grouped into five easily interpretable classes according to sociological indicators.) –

H. T. Forst: Anwendung der Cluster-Analyse zur Typisierung des Freizeitverhaltens von Jugendlichen. (On the basis of 21 qualitative leisure time attributes 154 juveniles are classified by a hierarchical algorithm.) –

W. Schneider: Taxonomie der Gedächtnisleistungen schwacher und normaler Rechtschreiber. (The hypothesis of distinguishability between legasthenic and normal pupils by their memory abilities is examined by constructing and interpreting 6 groups of pupils.) –

H.-H. Bock

SCHEELE, Martin: **Ordnung und Wortschatz des Wissens.** Entwurf zu einem Überblick über das menschliche Wissen auf der Grundlage der Wörter. 1. Bd.: Das Ordnungssystem. Universelle Facetten-Classifikation (UFC). (The organization and vocabulary of knowledge. A blueprint for an overview of human knowledge on the basis of the vocabulary of language. Vol.1: The ordering system. Universal Faceted Classification UFC). Schlitz/Hessen: Verlag H. Guntrum II, 1977. 208 p., ISBN 3-921739-01-2.

According to the author this is a universal faceted classification for two purposes, namely,
(1) to serve as an orientation or guide into the entire domain of human knowledge, a classified guide into alphabetically arranged encyclopedias, and
(2) to serve as an index language for bibliographic files, especially for personal bibliographic files.
The entire work is projected to consist of three parts, namely,

Part A – what this reviewer calls a "core classification".

Part B – a thesaurus of words of the German language expressed as combinations of semantic factors taken from the core classification and arranged by these factors, resulting in a classified sequence of the words.

Part C – an alphabetical index to part B.

The book under review contains the introduction to the entire work, the core classification (Part A), an alphabetical index of all terms used in the core classification, and some samples of entries of Part C. Thus, the subject of this review is a critical analysis of the core classification.

One might perhaps admire the courage of an author who single-handedly attempts to create a new universal classification. However, the attempt failed. Scheele produced an addition to the long list of classifications that can serve as examples of how *not* to develop a classification.

As will be demonstrated in the following analysis the main reasons for this failure are three:
(1) The author lacks insight into the principles of conceptual organization. This manifests itself in two ways:
  (1a) The scheme adheres to a rigidly monohierarchical structure without any cross references. This is all the worse in view of the fact that many concepts are not elemental but compound and thus are bound to have more than one broader concept.
  (1b) There are quite a few instances where compound concepts are included in the scheme but one or more of their more general elemental components are missing.
(2) The conceptual organization is strictly subordinate to a rigid four-digit decimal notation.
(3) In many areas the author simply does not have the expertise to produce a meaningful structure.

One must expect from a classification that it provides adequate coverage of the concepts needed for the purpose at hand, that it contains all useful hierarchical and associative relationships, and that it displays these relationships in a useful way. We shall take these points up in reverse order.

First we shall discuss the linear arrangement and relationships displayed by it.

The classification uses ten main classes:
0 General concepts
1 Matter
2 Live (living organisms)
3 Man
4 Society
5 Technology
6 Fine arts
7 The earth
8 The universe
9 Metaphysics (primarily religion)

(Like others before him, Scheele draws the absurd conclusion that since we are using the decimal system of numbers, therefore, the optimal division of human knowledge is into 10 main classes.) Scheele stresses that this subdivision by phenomena rather than by traditional scientific and scholarly disciplines is more in keeping with modern developments. However, the choice of the primary characteristic of subdivision must be predicated upon the intended use of a classification, and should perhaps be left to the user. Therefore, we shall not argue this matter further. However, we shall examine the helpfulness of the arrangement with respect to the few information science concepts included in the scheme by simply listing them with a little of their context:

019 Other general concepts
  0190 Information
  0191 Sensitivity
  . . . .
  0197 Tendency
  0198 Inventory
339 Other service professions
  3395 Librarian
  3396 Archivist
  3397 Documentalist (Dokumentar)

https://doi.org/10.5771/0943-7444-1978-2-112-1 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY