Book Review

B. Mirkin, *Mathematical Classification and Clustering*, Kluwer Academic Press: Boston-Dordrecht, 1996, 448 p.

The book presents a comprehensive description of current mathematical theories in combinatorial clustering supplemented with a general discussion of goals and scope of classification and clustering and a review of clustering algorithms, as well. The author consistently pursues the idea that clustering is a (set of) techniques purported for transformation of data into clustering structures.

Traditionally, clustering is considered as an activity , for revealing "coherent" groups (clusters) of points in a geometrical space. The author develops another approach which suggests that clustering is an approximation of a data set by a cluster structure. This approach appears much more productive for creating a sound mathematical theory unifying many aspects of the traditional approaches and addressing some problems which couldn't be solved or even formalized before. It should be noted that the author adds to these developments (mostly due to himself and his collaborators) many results found recently by the other authors (especially in the aspects concerning hierarchies and their extensions) and which seem not exactly belonging in the principal line of the approximation approach. This may make it difficult for a potential reader to go through all details of the material presented. On the other hand, the fact that the author somehow managed to put in many of the most interesting and recent developments in mathematical clustering makes the book an invaluable reference text.

The book consists of 7 chapters: 1. Classes and clusters. 2. Geometry of data. 3. Clustering algorithms: a review. 4. Single cluster clustering. 5. Partition: square data table. 6. Partition: rectangular data table. 7. Hierarchy as a clustering structure.

In the first chapter, the author reviews classification goals and forms to define clustering as a data-driven phase of classification. The goals are, basically, those related to forming concepts (classes) and describing interrelations between them. The author mentions this as shaping and keeping knowledge, structurizing phenomenae, and relating different aspects to each other. The latter goal usually is not considered a part of clustering; it is covered by different techniques from regression analysis to pattern recognition to machnine learning. The author suggests some material about this (as in Sections 3.3 and 6.3); in particular, he explains the specifics of the clustering approach, but this topic seems more a subject of future research rather than a more or less complete treatment.

Journal of Global Optimization **12:** 105–108, 1998.

A dozen well selected small examples of clustering problems (ranging from Fisher's Iris data set to a 8·5 Russian masterpieces mixed data table) along with a review of data types completes the chapter.

The author systematically considers three data types: column-conditional (entity-to-variable), comparable (similarity/dissimilarity) and aggregable (mostly contingency) data (the latter kind is an author's innovation) which are considered as mathematical objects on their own (almost with no statistical thinking about them). Geometrical representation and singular-value-based decompositions of these three data types are the contents of chapter 2. Depending on what is considered as the space elements - the rows, the columns, or the data matrix itself - different clustering approaches are outlined as those related to traditional "geometric" clustering, conceptual clustering or approximation clustering (which is the principal concern of the author), respectively.

The third chapter can be considered a book on its own. The review systematizes numerous ideas, methods and algorithms in a most natural way: with regard to input data types and output cluster structures. The list of the cluster structures includes not only the conventional hierarchical vs. nonhierarchical alternatives but also single clusters and many other structures such as neural networks or mixtures of distributions.

The other four chapters are mostly devoted to the description of the approximation clustering approach. Three types of clusterings are considered: single clusters, partitions and hierarchies. The basic idea is this: for a given data type, a cluster structure is represented by a matrix of the same size; that is, as a point of the space whose elements are data sets. Then, the clustering problem is to find a closest cluster structure to a given data "point". The between-"point" distances considered are Euclidean (least-squares method) and, sometimes, city-block (least-moduli method) ones. The optimization problems involving both discrete clustering structures and "smoothifying" quantitative parameters is a major mathematical framework. There is yet another mathematical aspect in the problems concerning relations of the solutions found to intuitive meaning of clusters as coherent and cohesive groups, i.e., "real" clusters of the entities.

Mostly locally-optimal methods are employed through all the material: alternating optimization, local search and greedy algorithms. However, the author proves that the resulting clusters are "real" which may be considered a substantiation of use of local optimization techniques, in the present context. Another substantiation relies upon the fact that some of the most used heuristical techniques such as K-Means and agglomerative Ward clustering appear to be those local algorithms. This latter fact allows for naturally extending the heuristical algorithms to getting "deeper" minima.

In particular, Chapter 4 is devoted to single cluster clustering, a term probably coined by P. Hansen in the early eighties. It was the first time, that the problem of separating a subset of "bright coloured" entities from the entire set received a treatment in a clustering text. In the framework of approximation approach, the author presents models and methods for revealing five single cluster structures: principal clusters, ideal fuzzy types, additive clusters, star clusters, and box clusters. He suggests also extending these methods to finding overlapping multi-cluster solutions by sequentially extracting the clusters found from the data. In spite of their "greediness" the methods seem quite effective as illustrated with a set of examples treated also by other authors who used different methods. Two heuristical single cluster clustering techniques - seriation and moving center separation - appear equivalent to local search approximation, which allows to set some parameters of the algorithms (initial setting, radia and stopping rule) and to prove their convergence. Amazingly, in this context, some of the best known combinatorial problems (such as min cut or maximum density subgraph) emerge as those of approximation clustering. A new class of greedily-wise globally optimized criteria, the so-called quasi-convex set functions, is described.

Partitioning problems for square similarity or mobility data matrices are extensively treated in Chapter 5. A set of relevant results about properties of Lance-Williams agglomerative partitioning techniques are described. The author's contributions go in three lines: (1) uniform partitioning which appears a theoretic development substantiated by earlier Monte-Carlo experiments by G. Milligan (1981), (2) block-model structuring which is finding a "small" graph approximating a "large" similarity matrix, and (3) aggregation of mobility or other flow data. Each of these present a careful study of the criteria emerged along with examples of application of the locally optimizing techniques.

In chapter 6, approximation partitioning problems for rectangular entity-to-variable and contingency data are considered. An attempt is made to extend "geometric" clustering to the case of tables comprising both quantitative variables and qualitative categories. The basis is the preliminary standardization of such data based on two postulates about the data scatter. The postulates seem controversial; however, the author derives a nice interrelation from them of his developments with contingency data analysis. In particular, some well known contingency coefficients such as Pearson chi-squared appear to be contributions of cluster partitions to the data scatter. An extensive, though also controversial, description of using of the author's outfinds in choosing parameters of K-Means clustering strategy is presented.

Chapter 7 (also a book on its own) presents an update account of the "state of the art" in presenting data with hierarchical structures. The subjects covered are: representing and comparing hierarchies, monotone-invariant clustering methods, ultrametrics, tree metrics, Robinson matrices and pyramids, weak hierarchies and split decompositions. The last section is devoted to description of a quite original approach to approximating hierarchical clustering based on linear embedding of discrete binary hierarchies.

The structure of the book as presented in the Preface (p. xiv) is a hierarchy itself, which makes it possible for a potential user to read the book by parts. Also, concluding discussions to the sections and opening features to the chapters make it easier a task of going through all the enormous material included in the book. All the algorithms are box-framed and examples are carefully separated which also helps in acquiring and coding them.

However, there are many misprints which sometimes make getting through difficult (as, for instance, confusing denotations p_{i+} and p_i for the same marginal frequencies on p. 85 and the description of the five-cluster solution in the caption to Figure 5.4 (p. 281) contradicting to that in the tree presented). A potential reader may find it curious to try to develop a method for separation of the single cluster of the misprints in the book.

Overall, the author's claim that the book can be considered in a threefold way (as a reference and textbook and presentation of his own approach) seems correct. The monograph is a significant step in transforming clustering from a set of ad hoc techniques into a sound mathematically supported theory which is related to discrete mathematics, multivariate statistics, machnine learning, artificial intelligence, non-convex and combinatorial optimization. The material presented seems a set of benchmarks for further developments. The book should be recommended as an inspiring reading to the students and specialists in the fields listed above. The numerous algorithms suggested can be exploited by data analysis practitioners in various application areas.

M. LEVIN

C.A. Floudas, *Nonlinear and Mixed-Integer Optimization. Fundamentals and Applications*, Oxford University Press, 1995, 462 pp, 75 USD, ISBN 0-19-510056-5 (Topics in Chemical Engineering)

During the last two decades, the field of mixed-integer nonlinear optimization has been witnessing a dramatic growth, with substantial theoretical advances and algorithmic developments, and continuously increasing popularity, with a wide spectrum of applications in many areas of engineering, applied mathematics, applied science and operations research, such as process synthesis, scheduling, planning and design of processes/facilities, molecular design, process and control design and topology of transportation networks to name a few. With a large number of publications continuously appearing in major scientific journals, the apparent lack of a book documenting the fundamentals in mixed-integer nonlinear optimization was clearly felt. The book of C.A. Floudas, directed to researchers and users of mixed-integer nonlinear optimization, clearly bridges this gap - coming from a leading authority in the field, such an edition represents a documental milestone: it contains all major original results and mathematical developments in mixed-integer nonlinear optimization theory along with a number of important application areas in chemical engineering.

The book is effectively divided in three major parts - Part 1, covering fundamentals in nonlinear optimization, provides the necessary background to Part 2,

Journal of Global Optimization **12:** 108–110, 1996.