

*Sankhyā : The Indian Journal of Statistics*

2006, Volume 68, Part 1, p 174-175

© 2006, Indian Statistical Institute

### **Branch-and-Bound Applications in Combinatorial Data Analysis**

MICHAEL J. BRUSCO AND STEPHANIE STAHL

*Statistics and Computing*

(2005) Springer Verlag, (xii)+221 pp.

Price \$69.95, ISBN 0-387-25037-9

There are many problems in statistics like maximum likelihood estimation, least square methods, gradient-based search methods for multidimensional scaling etc., which require the solution of continuous optimization problems. Solution proceeds via standard calculus-based methods; in some cases closed form solutions are possible whereas in other cases iterative procedures are adopted.

Although somewhat less emphasized in the statistical literature, discrete optimization problems arise in a variety of data analysis scenario, e.g. variable selection for regression analysis, partitioning for cluster analysis etc. For discrete optimization problems, as the functions are not smooth, demonstration of optimality typically requires evaluation, either explicitly or implicitly, of the complete set of feasible solutions of the problem. A systematic method of such exploration and enumeration is very important in this kind of scenario. “Branch and bound” method is an established procedure for handling discrete optimization problems. Operation research experts have been using this method since its proposition in the early sixties for solving popular optimization problems like integer programming, travelling salesman problem, and so on.

This book, to the best of my knowledge, is the first attempt to use this technique, branch and bound, to solve discrete optimization problems that arise in statistical data analysis. The authors are one of the pioneers in this area and authored a series of papers in relevant leading journals. This book is basically a conglomeration of these papers, but written in a more lucid language, and includes a good number of examples, illustrations and case studies.

The book starts with a good description of the branch and bound method and motivates with a nice example of minimum back tracking in a workflow matrix. Thereafter it gets into three kinds of statistical problems: cluster analysis, seriation and variable selection. The relevance of this particular

method to these three problems are explained separately and with clarity.

Chapters 2 to 6 deal with partitioning problem related to clustering. The approaches considered are: minimum-diameter partitioning, minimum within-cluster sums of dissimilarities partitioning, minimum within-cluster sums of squares partitioning and multiobjective partitioning. Chapters 7 to 11 deal with the problem of seriation. (i.e. construction of an order for a collection of objects). Several approaches of seriation: maximization of a dominance index, maximization of gradient indices, unidimensional scaling and multiobjective seriation, have been considered in details. The last three chapters viz 12-14 deal with the problem of Variable Selection. The contexts considered here are cluster analysis and regression analysis.

For each application topic in each area, the book first provides a small numerical example, an algorithm (along with available software), application of the algorithm to larger data sets, its strength and limitations. The algorithms presented are extremely clear and the readers will surely get the insight of it. This book is going to be extremely useful to the researchers or practitioners of combinatorial data analysis. The bibliography presented at the end is a complete one and readers can utilise it for further reading.

BIMAL ROY  
APPLIED STATISTICS UNIT  
INDIAN STATISTICAL INSTITUTE  
203 BARRACKPORE TRUNK ROAD  
KOLKATA 700 108  
E-mail: bimal@isical.ac.in