Preface of the volume

This volume contains a selection of papers presented at the workshop on Symbolic Data Analysis of the PKDD’2000 conference which was held in Lyon the 12 September 2000. This book contains original research contributions, innovative applications and overview papers in various aspects of Symbolic Data Analysis.

When observations in large data sets are aggregated into smaller more manageable data sizes, the resulting descriptions of the new units invariably involve "symbolic data". By symbolic data, we mean that rather than a specific categorical or numerical value, an observed value can be a set of categories or numbers, an interval or a probability distribution or any kind or more complex information than the usual one. In addition there may be rules or taxonomies. Hence, Symbolic Data Analysis generalises classical methods of exploratory, statistical and graphical data analysis to more complex data issued from huge Relational Data Bases. Now the domain is enhanced by two prototype software called SODAS and ISO3'D issued from two European projects involving 21 European industrial or academic teams from ten countries. Several papers of this book illustrate these software.

The volume presents first an overview and the state of the art in Symbolic Data Analysis. Areas which received more attention in this book are applications to financial data and long time series, generalising standard methods of non supervising Data Analysis with the example of Principal Component Analysis. An example of supervised method for extracting symbolic descriptions from categorical data is also given. Finally, the last papers are devoted to the important question of visualising symbolic descriptions of concepts obtained from queries to huge Relational Data Bases or from clustering large data sets by Hierarchies or special overlapping clusters called "Pyramids".

Acknowledgement

First of all we wish to express our gratitude towards the authors of the papers in the present volume, not only for their contribution but also for their diligence and their timely production of their papers.

We also thank the organizer of PKDD’2000 for their invitation to organise this Workshop and specially D. Zighed for his help in this publication.
Paris, August 2000

Edwin Diday, Oldemar Rodríguez

PKDD WORKSHOP ON SYMBOLIC DATA ANALYSIS SCHEDULE

9h 30 INTRODUCTION TO SYMBOLIC DATA ANALYSIS E. Diday (Dauphine University-INRIA)
10h SYMBOLIC ANALYSIS OF FINANCIAL DATA M. Touati (Dauphine University)
11h SYMBOLIC REPRESENTATION OF LONG TIME SERIES. G. Hbrail, B.Hugueney (EDF, France)

10h 30 BREAK

11h GENERALIZATION PRINCIPAL COMPONENT ANALYSIS FOR HISTOGRAM DATA O. Rodriguez, E. Diday, S. Winsberg
11H 45 EXTRACTING SYMBOLIC DESCRIPTION FROM SUPERVISED CATEGORICAL DATA M. Gettler-Summa (Dauphine University, SOFRES)

"TUTORIAL" PART OF THE WORKSHOP:
14H THE SODAS AND ISO3’D SOFTWARE: General presentation E. Diday, C. Mazel
14H30 FROM DATA BASE TO SYMBOLIC OBJECTS IN SODAS F. Vautrain (Dauphine University)
15H SYMBOLIC DATA ANALYSIS METHODS IN SODAS BY EXAMPLES M. Touati, M. Summa, F. Vautrain
16H TEMPORAL SYMBOLIC DESCRIPTIONS GRAPHICS IN ISO3’D M. Noirhomme (Namur FUNDP University, C. Mazel (TGS, Bordeaux)
16H 30 CLUSTERING LARGE DATA SETS AND VISUALISATION OF SYMBOLIC DATA ANALYSIS METHODS IN ISO3’D C. Mazel, M. Noirhomme, V.Batagelj, 17h 15 Practice on examples