

Table of Contents of the Wokshop proceedings on SDA

Knowledge discovery from symbolic data and the SODAS software

E. Diday

Symbolic Analysis of Financial Data

F. Goupil, M. Touati, E. Diday and H. Van Der Veen

Generalization of the Principal Components Analysis to Histogram Data

O. Rodriguez, E. Diday and S. Winsberg

Symbolic Representation of Long Time-Series

G. Hebrail and B. Hugueney

Pyramidal Clustering Algorithms in ISO-3D Project

O. Rodriguez and E. Diday

Clustering Large Datasets and Visualizations of Large Hierarchies and Pyramids:
Symbolic Data Analysis Approach

V. Batagelj, E. Pavletič, M. Zaveršnik and S. Korenjak-Cerne

Temporal Symbolic Descriptions Graphics in ISO-3D

M. Noirhomme, A. Nahimana and C. Mazel

Marking and Generalization by Symbolic Descriptions in the Symbolic Official Data
Analysis Software

M. Gettler-Summa

Preface

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".

Acknowledgements

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 Rodriguez

Program 12 September

- 9h30 Knowledge discovery from symbolic data and the SODAS software
E. Diday
- 10h Symbolic Analysis of Financial Data
F. Goupil, M. Touati, E. Diday and H. Van Der Veen
- 10h30 Coffee break
- 11h00 Symbolic Representation of Long Time-Series
G. Hebrail and B. Hugueney
- 11h30 Generalization of the Principal Components Analysis to Histogram Data
O. Rodriguez, E. Diday and S. Winsberg
- 12h00 Marking and Generalization by Symbolic Descriptions in the Symbolic Official Data Analysis Software
M. Gettler-Summa
- 12h 30 LUNCH
- 14h 30 The SODAS and ISO3'D Software: General presentation
E. Diday, C. Mazel
- 15h From Data Base To Symbolic Objects in SODAS
F. Vautrain (Dauphine University)
- 15h 30 Symbolic Data Analysis Methods in SODAS by Examples
M. Touati, M. Summa, F. Vautrain
- 16h Coffe Break
- 16h30 Pyramidal Clustering Algorithms in ISO-3D Project
O. Rodriguez and E. Diday
- 16h50 Clustering Large Datas sets and Visualizations of Large Hierarchies and Pyramids: Symbolic Data Analysis Approach
V. Batagelj, E. Pavletić, M. Zaveršnik and S. Korenjak-Cerne
- 17h30 Temporal Symbolic Descriptions Graphics in ISO-3D
M. Noirhomme, A. Nahimana and C. Mazel
- 17h30 General discussion