

Preface

Following a biannual tradition of organizing joint meetings between classification societies, the Classification and Data Analysis Group of the Italian Statistical Society, CLADAG, has organized its international meeting together with the German Classification Society, GfKl, at Firenze, Italy, September 8–10, 2010. The Conference was originally conceived as a German-Italian event, but it counted the participation of researchers from several nations and especially from Austria, France, Germany, Great Britain, Italy, Korea, the Netherlands, Portugal, Slovenia, and Spain. The meeting has shown once more the vitality of data analysis and classification and served as a forum for presentation, discussion, and exchange of ideas between the most active scientists in the field. It has also shown the strong bonds between the two classification societies and has greatly helped to deepen relationships.

The conference program included 4 Plenary, 12 Invited, and 31 Contributed Sessions. This book contains selected and peer-reviewed papers presented at the meeting in the area of “Classification and Data Mining.” Browsing through the volume, the reader will see both methodological articles showing new original methods and articles on applications illustrating how new domain-specific knowledge can be made available from data by clever use of data analysis methods. According to the title, the book is divided into three parts:

1. Classification and Data Analysis
2. Data Mining
3. Applications

The methodologically oriented papers on classification and data analysis deal, among other things, with robustness, analysis of spatial data, and application of Monte Carlo Markov Chain methods. Variable selection and clustering of variables play an increasing role in applications where there are substantially more variables than observations. Support vector machines offer models and methods for the analysis of complex data structures that go beyond classical ones. Special discussed topics are association patterns and correspondence analysis.

Automated methods in data mining, producing knowledge discovery in huge data structures such as those associated with new media (e.g., Internet), digital images,

or genomes in Genetics, continue to represent, in the near future, a big challenge for data analysis. Information is readily retrieved in these fields; however, interpreting it and identifying relevant results is not a straightforward task at all. Especially data produced by the Internet, genetics studies on genomes, and proteomes have a particular appeal as objects of analysis and are studied in this book. Furthermore, there are applications of the Markov chains model, to a new brand of problems such as the knowledge discovery in the Internet, the analysis of large biomedical data sets, and in more general sensor data. Moreover, the automatic online processing of data streams is becoming increasingly important. In sociology and market research, opinion mining on a large number of expressed preferences plays an important role. All these data typologies require algorithmic methods in the interface between statistics and computer science. Other contributions in the book focus on the application of the singular value decomposition to structural learning in Bayesian networks and on molecular simulation for drug design.

The last part of the book contains interesting applications to various fields of research such as sociology, market research, environment, geography, and music: estimation in demographic data, description of professional profiles, metropolitan studies such as income in municipalities, labor market research, environmental energy consumption, geographical data such as seismic time series, auditory models in speech and music, application of mixture models to multi-state data, and visualization techniques.

We hope that this short description stimulates the reader to take a closer look at some of the articles. Our thanks go to Andrea Giommi and his local organizing team who have done a great job (Bruno Bertaccini, Matilde Bini, Anna Gottard, Leonardo Grilli, Alessandra Mattei, Alessandra Petrucci, Carla Rampichini, Emilia Rocco). We gratefully acknowledge the Faculty of Economics and the “Ente Cassa di Risparmio di Firenze” for financial support, and desire to express our special thanks to Chiara Bocci for her valuable contribution to the organization of the meeting and for her assistance in producing this book. Also on behalf of our colleagues we may say that we have very much enjoyed having been their guests in Firenze. The dinner with a view to the Dome was excellent and we appreciate it very much.

We wish to express our gratitude to the other members of the Scientific Programme Committee: Daniel Baier, Reinhold Decker, Filippo Domma, Luigi Fabbris, Christian Hennig, Carlo Lauro, Berthold Lausen, Hermann Locarek-Junge, Isabella Morlini, Lars Schmidt-Thieme, Gabriele Soffritti, Alfred Ultsch, Rosanna Verde, Donatella Vicari, and Claus Weihs.

We also thank the section organizers for having put together such strong sections. The Italian tradition of discussants and rejoinders has been a new experience for GfKI. Thanks go to the referees for their important job. Last but not least, we thank all speakers and all who came to listen and to discuss with them.

Florence, Italy
Passau, Germany
Rome, Italy

Antonio Giusti
Gunter Ritter
Maurizio Vichi

Classification and Data Mining

Giusti, A.; Ritter, G.; Vichi, M. (Eds.)

2013, XIV, 286 p. 85 illus., 49 illus. in color., Softcover

ISBN: 978-3-642-28893-7