

Preface

New Frontiers in Mining Complex Patterns (NFMCP 2013)

Modern automatic systems are able to collect huge volumes of data, often with a complex structure (e.g. multi-table data, XML data, web data, time series and sequences, graphs and trees). This fact poses new challenges for current information systems with respect to storing, managing, and mining these sets of complex data. The Second International Workshop on New Frontiers in Mining Complex Patterns (NFMCP 2013) was held in Prague in conjunction with the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD 2013) on September 27, 2013. It was aimed to bring together researchers and practitioners of data mining who are interested in the advances and latest developments in the area of extracting patterns from complex data sources like blogs, event or log data, medical data, spatio-temporal data, social networks, mobility data, sensor data and streams, and so on.

This book features a collection of revised and significantly extended versions of papers accepted for presentation at the workshop. These papers went through a rigorous review process to ensure their high quality and to have them comply with Springer-Verlag publication standards. The individual contributions of this book illustrate advanced data mining techniques which preserve the informative richness of complex data and allow efficient and effective identification of complex information units present in such data.

The book is composed of four parts and a total of sixteen chapters.

Part I gives a view of **Data Streams and Time Series Analysis** by illustrating some complex situations involving temporal and spatio-temporal data. It consists of five chapters. Chapter 1 describes data driven parameter estimation measures for mining flock patterns with a validation procedure to measure the quality of these extracted patterns. Chapter 2 presents a simple yet effective and parameter-free feature construction process for time series classification. Chapter 3 proposes a learning algorithm, combining conditional log-likelihood with Bayesian parameter estimation, designed for analyzing multivariate streaming data. Chapter 4 investigates the problem of mining frequent trajectories by resorting to frequent sequential pattern mining. Chapter 5 details a process mining approach that uses predictive clustering to equip an execution scenario with a prediction model.

Part II analyses issues posed by **Classification, Clustering, and Pattern Discovery** in presence of complex data. It consists of six chapters. Chapter 6 describes a novel methodology that combines principal component analysis and support vector machines in order to detect emotions from speech signals. Chapter 7 illustrates a sequential pattern mining algorithm that allows us to discover lengthy noise-tolerant sequential patterns over item-indexable databases. Chapter 8 studies the problem of efficiently mining frequent partite episodes that satisfy partwise constraints from an

input event sequence. Chapter 9 proposes an estimation algorithm for the copula of a continuous multivariate distribution. Chapter 10 focuses on extending the ReliefF algorithm for regression, in order to address the task of hierarchical multi-label classification (HMC). Chapter 11 addresses the task of learning both global and local models for predicting structured outputs.

Part III presents technologies and applications where complex patterns are discovered from **Graphs, Networks, and Relational Data**. It contains three chapters. Chapter 12 describes a generative model for random graphs with discrete labels and weighted edges. Chapter 13 proposes a pre-processing strategy to simplify Semantic Similarity Networks based on a hybrid global-local thresholding approach based on spectral graph theory. Chapter 14 investigates the translation of the complex data represented by natural language text to complex (relational) patterns that represent the writing style of an author.

Finally, Part IV gives a general overview of **Machine Learning and Music data**. It contains two chapters. Chapter 15 explores a set of personalized emotion classifiers that are learned using feature data extracted from audio and tagged with a set of emotions by volunteers. Chapter 16 addresses the problem of using random forests, in order to identify multiple musical instruments in polyphonic recordings of classical music.

We would like to thank all the authors who submitted papers for publication in this book and all the workshop participants and speakers. We are also grateful to the members of the Program Committee and external referees for their excellent work in reviewing submitted and revised contributions with expertise and patience. We would like to thank João Gama for his invited talk on “Evolving Data, Evolving Models”. Special thanks is due to both the ECML PKDD workshop chairs and to the members of ECML PKDD organizers who made the event possible. We would like to acknowledge the support of the European Commission through the project MAESTRA - Learning from Massive, Incompletely annotated, and Structured Data (Grant number ICT-2013-612944). Last but not the least, we thank Alfred Hofmann of Springer for his continuous support.

January 2014

Annalisa Appice
 Michelangelo Ceci
 Corrado Loglisci
 Giuseppe Manco
 Elio Masciari
 Zbigniew W. Ras

New Frontiers in Mining Complex Patterns
Second International Workshop, NFMCP 2013, Held in
Conjunction with ECML-PKDD 2013, Prague, Czech
Republic, September 27, 2013, Revised Selected
Papers

Appice, A.; Ceci, M.; Loglisci, C.; Manco, G.; Masciari, E.;
Ras, Z. (Eds.)

2014, XII, 261 p. 80 illus., Softcover

ISBN: 978-3-319-08406-0