

Preface

The recent and novel research contributions collected in this book are extended and reworked versions of a selection of the best papers that were originally presented in French at the EGC'2014 and EGC'2015 conferences respectively held in Rennes (France) in January 2014 and Luxembourg in January 2015. The papers have been selected among the papers accepted in long format at the conferences. For the conferences, the long papers are themselves the result of a double-blind peer-review process among the 106 papers initially submitted to the conference in 2013 and 83 papers in 2015 (conference acceptance rate for long papers of 26 % in 2014 and 27 % for 2015). These conferences were the 14th and 15th edition of this event, which takes place each year and which is now successful and well-known in the French-speaking community. This community was structured in 2003 by the Foundation of the International French-speaking EGC society (EGC in French stands for "Extraction et Gestion des Connaissances" and means "Knowledge Discovery and Management", or KDM). This society organizes every year its main conference (about 200 attendees) also workshops and other events with the aim of promoting exchanges between researchers and companies concerned with KDM and its applications in business, administration, industry or public organizations. For more details about the EGC society, please consult <http://www.egc.asso.fr>.

Structure of the Book

This book is a collection of representative and novel works done in Data Mining, Knowledge Discovery, Clustering and Classification. It is intended to be read by all researchers interested in these fields, including Ph.D. or M.Sc. students, and researchers from public or private laboratories. It concerns both theoretical and practical aspects of KDM.

This book has been structured into three parts. The first four chapters are related to optimization consideration while mining data. The second part presents four

chapters dealing with specific quality measures, dissimilarities and ultrametrics. The five remaining chapters focus on semantics, ontologies and social networks.

Mining Data with Optimization

Chapter 1, *Online Learning of a Weighted Selective Naive Bayes Classifier with Non-convex Optimization*, is concerned with improving supervised classification for data streams with a high number of input variables. It focuses on direct estimation of weighted naïve Bayes classifiers using a sparse regularization of the model log-likelihood which takes into account knowledge relative to each input variable.

Chapter 2, *On Making Skyline Queries Resistant to Outliers*, aims to reduce the impact of exceptional points when computing skyline queries, so that outliers do not “hide” more interesting answers. The approach relies on the notion of fuzzy typicality and makes it possible to compute a graded skyline answers. A GPU-based parallel implementation is also described.

Chapter 3, *Adaptive Down-Sampling and Dimension Reduction in Time Elastic Kernel Machines for Efficient Recognition of Isolated Gestures*, addresses both the dimensionality reduction of the feature vector describing multidimensional motion time series and the dimensionality reduction along the time axis by the means of adaptive down-sampling used in conjunction with time Elastic Kernel Machines.

Chapter 4, *Exact and Approximate Minimal Pattern Mining*, presents a generic framework for exact and approximate minimal patterns mining by introducing the concept of minimizable set system, and it also demonstrates that minimal patterns mining is polynomial-delay and polynomial-space.

Quality Measures, Dissimilarities and Ultrametrics

Chapter 5, *Comparison of Proximity Measures for a Topological Discrimination*, proposes a methodology to make a clustering of proximity measures in the context of discrimination using a topological structure, and to choose the best discriminant measure for considered data.

Chapter 6, *Comparison of Linear Modularization Criteria Using the Relational Formalism, an Approach to Easily Identify Resolution Limit*, deals with the comparison of linear modularization criteria by using the Mathematical Relational analysis (MRA). MRA allows to compare numerous criteria on the same type of formal representation in order to facilitate their understanding and their usefulness in practical contexts.

Chapter 7, *A Novel Approach to Feature Selection Based on Quality Estimation Metrics*, proposes an adaptation of the Feature maximization (F-max) criterium in order to perform more efficient feature selection and feature contrasting within the framework of supervised classification. The comparison with other feature selection

techniques shows a significant improvement of the performances, notably in the case of unbalanced, highly multidimensional and noisy textual data.

Chapter 8, *Ultrametricity of Dissimilarity Spaces and Its Significance for Data Mining*, evaluates the extent to which a dissimilarity is close to an ultrametric by introducing the notion of ultrametricity of a dissimilarity, and examines their influence on the accuracy of a classification or the quality of a clustering.

Semantics, Ontologies and Social Networks

Chapter 9, *SMERA: Semantic Mixed Approach for Web Query Expansion and Reformulation*, uses implicate and explicate concepts to automatically improve web queries. This approach handles several challenges related to query expansion, such as selective choice of expansion terms, named entities treatment, and concept-based query representation.

Chapter 10, *Multi-layer Ontologies for Integrated 3D Shape Segmentation and Annotation*, introduces an original framework where annotation and segmentation of 3D meshes are performed conjunctly. An expert's knowledge of the context is used while minimizing the use of geometric analysis, and a multi-layer ontology is designed to conceptualize 3D object features from the point of view of their geometry, topology, and possible attributes.

Chapter 11, *Ontology Alignment Using Web Linked Ontologies as Background Knowledge*, proposes an ontology matching method for aligning a source ontology with target ontologies already published and linked on the Linked Open Data (LOD) cloud. The evaluation was achieved on two well-known ontologies in the field of life sciences and environment: AgroVoc and Nalt.

Chapter 12, *LIAISON: reconciliAtion of Individuals Profiles Across SOcial Networks*, describes an algorithm that uses the social network topology and the publicly available personal information to iteratively determine the profiles that belong to the same individuals across several social networks.

Chapter 13, *Clustering of Links and Clustering of Nodes: Fusion of Knowledge in Social Networks*, compares two network clustering approaches: the search for communities and the extraction of frequent conceptual links, in order to understand both the intersections that can exist between them and the knowledge that emerges from their fusion.

Acknowledgments

The editors would like to thank the chapter authors for their insights and contributions to this book.

The editors would also like to acknowledge the members of the review committee and the associated referees for their involvement in the review process of the

book. Their in-depth reviewing, criticisms and constructive remarks have significantly contributed to the high quality of the selected papers.

Finally, we thank Springer and the publishing team, and especially T. Ditzinger and J. Kacprzyk, for their confidence in our project.

Nantes Cedex 3, France
Talence Cedex, France
Tours, France
April 2016

Fabrice Guillet
Bruno Pinaud
Gilles Venturini

Advances in Knowledge Discovery and Management

Volume 6

Guillet, F.; Pinaud, B.; Venturini, G. (Eds.)

2017, XXI, 278 p. 81 illus., 61 illus. in color., Hardcover

ISBN: 978-3-319-45762-8