

Preface

The recent and novel research contributions collected in this book are extended and reworked versions of a selection of the five best papers that were originally presented in French at the EGC'2013 Conference held in Toulouse, France, on January 2013 and one paper from the EGC'2014 Conference held in Rennes, France, on January 2014. The five papers from the 2013 edition of the conference have been selected from the 26 papers accepted in long format at the conference. These 26 long papers were themselves the result of a peer and blind review process among the 123 papers initially submitted to the conference in 2013 (acceptance rate of 26 % for long papers). This conference was the 13th 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 not only its main conference (about 200 attendees) but 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 two parts. The first three chapters are related to novel applications on real datasets of various origins. The second part of this book presents three methodological chapters on the foundations of knowledge extraction and management.

Chapter “[A Study of the Spatio-Temporal Correlations in Mobile Calls Networks](#)” proposes an analysis of phone-call detailed records collected during five months in France. MODL, a nonparametric method, is applied to solve two different problems: first, partitioning antennas leading to territory segmentation; and second, discretizing time aiming at determining changes in users’ behavior. A set of visualizations, emphasizing the most interesting patterns, eases the analysis and the interpretation of the results. Chapter “[Co-Clustering Network-Constrained Trajectory Data](#)” study the problem of clustering moving object trajectories in a road network environment. A bipartite graph representation is used to model the relationships between trajectories and road segments visited. The authors propose three approaches to clustering the vertices of such a graph. Using synthetic data, they demonstrate how the data can be used to gain insight about mobility in road networks such as detecting frequent routes, characterizing road segment roles, etc. The work by Grabar and colleagues presented in Chap. “[Medical Discourse and Subjectivity](#)” proposes a contrastive study of corpora from the medical field. The corpora contain documents that are differentiated by their specialization level: documents written by medical experts and by patients. The differentiation features are related to medical notions, uncertainty, emotions, and negation. These features appear to be relevant for the distinction between the types of documents aimed. The authors then discuss the roles played by uncertainty, emotions, and negation in these documents.

Chapter “[Relational Concept Analysis for Relational Data Exploration](#)” deals with Relational Concept Analysis (RCA) which is an unsupervised classification method producing a set of connected concept lattices by considering relations between objects from different contexts. While designed to be intuitive to extract knowledge from relational data, dealing with many relations with RCA implies scalability problems. This article presents an adaptation of RCA, tested on environmental data, to explore relations in a guided way in order to increase the performance and the pertinence of the results. In Chap. “[Dynamic Recommender System: Using Cluster-Based Biases to Improve the Accuracy of the Predictions](#)”, the authors propose a methodology for recommender systems based on Matrix Factorization (MF) that reduces the loss of quality of the recommendations over time. MF is very popular because it gives good scalability at the time of recommending while allowing remarkable prediction accuracy. However, one drawback of MF is that once its model has been generated, it delivers recommendations based on a snapshot of the incoming ratings frozen at the beginning of its generation. To take into account the new ratings, the model has to be computed periodically. The proposed solution to this problem improves the scalability of MF by reducing the frequency of model recomputations. Chapter “[Mining \(Soft-\) Skypatterns Using Constraint Programming](#)” introduces a softness in the skypattern mining problem. Skypatterns enable to express a user-preference point of view w.r.t. a dominance

relation. First, the authors show how softness can provide valuable patterns that would be missed otherwise. Then, thanks to CP, they propose a generic and efficient method to mine (soft-)skypatterns. Finally, the relevance and the effectiveness of the proposed approach through an experimental study is shown.

Nantes

Bordeaux

Tours

Lyon

June 2015

Fabrice Guillet

Bruno Pinaud

Gilles Venturini

Djamel Abdelkader Zighed

Advances in Knowledge Discovery and Management
Volume 5

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

2016, XVIII, 137 p., Hardcover

ISBN: 978-3-319-23750-3