Studies in Computational Intelligence 471

Fabrice Guillet
Bruno Pinaud
Gilles Venturini
Djamel Abdelkader Zighed (Eds.)

Advances in Knowledge Discovery and Management

Volume 3

Springer

Studies in Computational Intelligence 471

Editor-in-Chief

Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

For further volumes:
http://www.springer.com/series/7092

Fabrice Guillet, Bruno Pinaud, Gilles Venturini,
and Djamel Abdelkader Zighed (Eds.)

Advances in Knowledge Discovery and Management

Volume 3

Springer

Editors
Fabrice Guillet
LINA (CNRS UMR 6241)
Polytechnic School of Nantes University
Nantes Cedex 3
France

Gilles Venturini
Université François-Rabelais de Tours
Polytech'Tours, Dpt Informatique
Tours
France

Bruno Pinaud
Univ. Bordeaux 1, LaBRI
Talence Cedex
France

Djamel Abdelkader Zighed
Laboratoire ERIC
Université Lumiére Lyon 2
Bron
France

ISSN 1860-949X
e-ISSN 1860-9503
ISBN 978-3-642-35854-8
e-ISBN 978-3-642-35855-5
DOI 10.1007/978-3-642-35855-5
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012955287

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'2011 Conference held in Brest, France, on January 2011. These 10 best papers have been selected from the 34 papers accepted in long format at the conference. These 34 long papers were themselves the result of a peer and blind review process among the 131 papers initially submitted to the conference in 2011 (acceptance rate of 26% for long papers). This conference was the 11^{th} 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) 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, Business Intelligence, Knowledge Engineering and Semantic Web. It is intended to be read by all researchers interested in these fields, including PhD or MSc students, and researchers from public or private laboratories. It concerns both theoretical and practical aspects of KDM.

This book has been structured in two parts. The first part, entitled "Data Mining, classification and queries", deals with rule and pattern mining, with topological approaches and with OLAP. Three chapters study rule and pattern mining and concern binary data sets, sequences, and association rules. Chapters related to topological approaches study different distance measures and a new method that learns a

hierarchical topological map. Finally, one chapter deals with OLAP and studies the mining of queries logs.

The second part of the book, entitled “Ontology and Semantic”, is more related to knowledge-based and user-centered approaches in KDM. One chapter deals with the enrichment of folksonomies and the three other chapters deal with ontologies.

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 slected papers.

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

Nantes, Bordeaux, Tours, Lyon — Fabrice Guillet, Bruno Pinaud
October 2012 — Gilles Venturini, Djamel Abdelkader Zighed

Review Committee

All published chapters have been reviewed by 2 or 3 referees and at least one non-french speaking referee (2 for most papers).

- Tomas Aluja (UPC, Spain)
- Nadir Belkhiter (Univ. of Laval, Canada)
- Sadok Ben Yahia (Univ. of Tunis, Tunisia)
- Omar Boussaid (Univ. of Lyon 2, France)
- Paula Brito (Univ. of Porto, Portugal)
- Francisco de A. T. De Carvalho (Univ. Federal de Pernambuco, Brazil)
- Gilles Falquet (Univ. of Geneva, Switzerland)
- Carlos Ferreira (LIAAD INESC Porto LA, Portugal)
- Jean-Gabriel Ganascia (Univ. of Paris 6, France)
- Joao Gama (Univ. of Porto, Portugal)
- Fabien Gandon (INRIA, France)
- Robert Hilderman (Univ. of Regina, Canada)
- Philippe Lenca (Telecom Bretagne, France)
- Henri Nicolas (Univ. of Bordeaux, France)
- Monique Noirhomme (FUNDP, Belgium)
- Jian Pei (Simon Fraser Univ., Canada)
- Pascal Poncelet (Univ. of Montpellier, France)
- Zbigniew Ras (Univ. of North Carolina)
- Jan Rauch (Univ. of Prague, Czech Republic)
- Chiara Renso (KDDLAB — ISTI CNR, Italy)
- Lorenza Saitta (Univ. of Torino, Italy)
- Florence Sédes (Univ. of Toulouse 3, France)
- Dan Simovici (Univ. of Massachusetts Boston, USA)
- Ansaf Salleb-Aouissi (Columbia Univ., USA)
- Yannick Toussaint (Univ. of Nancy, France)
- Stefan Trausan-Matu (Univ. of Bucharest, Romania)
- Rosanna Verde (Univ. of Naples 2, Italy)
- Christel Vrain (Univ. of Orléans, France)
- Jef Wijsen (Univ. of Mons-Hainaut, Belgium)
- Michaël Mcguffin (Ecole de Technologie Supèrieure, Canada)

Associated Reviewers

Hanane Azzag, Nahla Benamor, Julien Blanchard, Marc Boullé, Sylvie Guillaume, Pascale Kuntz, Patrick Marcel, Mathieu Roche, Julien Velcin, Nicolas Voisine.

Contents

List of Contributors

Nathalie Abadie is a geographical and cartographical state works engineer, and a PhD candidate at IGN. She is working on specifications formalisations of geographic databases for their integration. She also works on the implementation and evaluation of the proposed models on the IGN databases.

Rafik Abdesselam is Professor of statistics and data analysis at the University of Lyon. His research and teaching interests include supervised classification, topological learning and methods for data mining. He is also member of various national program committees.

Julien Aligon is PhD student at the Computer Science Laboratory of Université François-Rabelais Tours, France. His research interests include On-Line Analytical Processing, data-warehouses, query personalization and recommendation in databases.

Hanane Azzag is currently associate professor at the University of Paris 13 (France) and a member of machine learning team A3 in LIPN Laboratory. Her main research is in biomimetic algorithms, machine learning and visual data mining. Graduated from USTHB University where she received his engineer diploma in 2001. Thereafter, in 2002 she gained an MSC (DEA) in Artificial Intelligence from Tours University. In 2005, after three years in Tours, she received her PhD degree in Computer Science from the University of Tours.

Giuseppe Berio is professor of computer science at the University of South Brittany, France, and affiliated to the Lab-STICC laboratory. Previously, he was senior researcher at the University of Turin, Italy, working in the Department of Computer Science. Prof. Berio holds the "Laurea" degree (1990) in Computer Science cum Laude from University of Turin, Master (1991) and PhD (1995) degrees both in Information Systems from Polytechnic of Turin. In 1997, he was granted the "Marie Curie Fellowship" from European Communities for a two year assignment to Laboratory for Industrial Engineering and Mechanical Production (LGIPM) of University of Metz, France. His main research work is in information

systems and in interoperability of enterprise software applications. He was in the core members of the UEML Thematic Network and INTEROP Network of Excellence (`http://www.inteop-noe.org`), both projects funded by the European Commission. Prof. Berio is currently member of IFIP Working Group 8.1 "Design and Evaluation of Information Systems" and IFAC Technical Committee 5.3 "Enterprise Integration and Networking".

Marc Boullé graduated from Ecole Polytechnique (France) in 1987 and Sup Telecom Paris in 1989. He is currently a senior researcher in the data mining research group of Orange Labs. His main research interests include statistical data analysis, data mining, especially data preparation and modelling for large databases. He developed regularized methods for feature preprocessing, feature selection and construction, model averaging of selective naive Bayes classifiers and regressors.

Ahmed Bounekkar is Associate Professor at the University Claude Bernard Lyon 1. He works in the field of data analysis, especially in spatial data analysis and diffusion models of pandemic influenza. He also works on multi-objective problems of optimization.

Sandra Bringay received her Ph.D. in 2006 at the University of Picardie Jules Verne in Medical Informatics. She was then a temporary lecturer (ATER) for the CERIM (Center for Studies and Research in Medical Informatics) at the University of Lille 2. Since 2007, she is a lecturer at the University of Montpellier III and she integrated the LIRMM laboratory in the data mining group (TATOO). She works specifically on data mining techniques dedicated to health data. She takes part of knowledge extraction projects dedicated to health data as well as collaborations with private partners.

Michel Buffa teaches Computer Engineering at the University of Nice Sophia-Antipolis. He conducts his research work on Socio-Semantic Web and more specifically on collaborative Knowledge Platforms. Previously He used to work on Underwater Virtual Reality with Professor Peter Sander. In 1994, He was a visiting scientist at the Robotic Institute of the Carnegie Mellon University in Pittsburgh.

Lionel Chauvin is a postdoctoral research assistant in Computer Science at the University of Nantes (France). He received his Ph.D in Computer Science from the University of Angers in 2010. His Ph.D thesis is about cognitive maps, ontologies and conceptual graphs. His current research are about semantic similarities and ontologies.

Fabien Gandon is a enior Researcher at Inria, Leader of the Wimmics research team in the Inria Research Center of Sophia-Antipolis (France). Fabien has a Ph.D. and HDR in Informatics and Computer Science and is a Graduated Engineer in Applied Mathematics from INSA Rouen. His professional interests include: Semantic Web, Ontologies, Knowledge Engineering and Modelling, Mobility, Privacy, Context-Awareness, Web Services and Multi-Agents Systems. His main domain of

application is organizational memories (companies, communities, etc.) and knowledge management in general. His personal objectives are to research and teach in the field of applied computer science and informatics, in an international environment. He also participates to W3C working groups.

Dominique Gay received a PhD degree in Computer Science in 2009 from Université de la Nouvelle-Calédonie (Nouméa, New-Caledonia) and Institut National des Sciences Appliquées (Lyon, France). He is currently a researcher in the data mining research group of Orange Labs. His main research interests are about local pattern mining and its use for classification purpose.

David Genest received his Ph.D in Computer Science from the University of Montpellier in 2000. He joined the LERIA at the University of Angers in 2001 as an assistant professor. His research interests are in the area of knowledge representation and especially graphical models with specific interests in conceptual graphs and cognitive maps.

Toader Gherasim is a PhD student at the University of Nantes and affiliated to the LINA laboratory (Computer Science Laboratory of Nantes Atlantique). Toader's research interests include knowledge engineering and ontology learning and evolution. Toader graduated from the Faculty of Automatic Control and Computers of the Politehnica University of Bucharest in 2007 and received a Master in Data Mining from The Polytechnic School of University of Nantes.

Sylvie Guillaume is an Assistant Professor at the University of Auvergne, France and researcher at the Laboratory of Computer Modeling and Optimization Systems (LIMOS). She hold a Ph.D. in Computer Science in 2000 from the University of Nantes, France. Her research domain is positive and negative association rules and quality measures in data mining. Since 2012, she is also working in data mining applied to specific biological problems and particularly in selection of plant phenolic compounds for the formulation of additives in ruminant feed.

Mounira Harzallah graduated from the Ecole Nationale d'Ingénieurs de Tunis (Tunisia). She received a Ph.D. degree in industrial engineering (2000) from the University of Metz (France). She is currently Assistant Professor at the University of Nantes and affiliated to the LINA laboratory (Computer Science Laboratory of Nantes Atlantique). Her research is in the field of enterprise modeling and knowledge engineering, especially focusing on human competence modeling, enterprise interoperability and similarity measures for ontology building and validation. Dr. Harzallah has been involved in various national and international research projects, among them the INTEROP Network of Excellence (`http://www.inteop-noe.org`) funded by European Commission.

Pascale Kuntz received the M.S. degree in Applied Mathematics from Paris-Dauphine University and the Ph.D. degree in Applied Mathematics from the Ecole des Hautes Etudes en Sciences Sociales, Paris in 1992. From 1992 to 1998 she was assistant professor in the Artificial Intelligence and Cognitive Science Department

at the Ecole Nationale Superieure des Telecommunications de Bretagne. In 1998, she joined the Polytechnic School of Nantes University (France), where she is currently professor of Computer Science in the LINA laboratory (UMR 6241). She was the head of the team "KOD - KnOwledge and Decision" for eight years (2003–2011). She is member of the board of the French Speaking Classification Society. Her research interests include classification, graph mining and graph visualization, and post-mining.

Mustapha Lebbah is currently Associate Professor at the University of Paris 13 and a member of Machine learning Team A3, LIPN. His main researches are centred on machine learning (Self-organizing map, Probabilistic and Statistic, unsupervised learning, cluster analysis. Graduated from USTO University where he received his engineer diploma in 1998. Thereafter, he gained an MSC (DEA) in Artificial Intelligence from the Paris 13 University in 1999. In 2003, after three year in RENAULT R&D, he received his PhD degree in Computer Science from the University of Versaille. He is also member of the IEEE, INNS, SFDS, EGC and AML group.

Aymeric Le Dorze is a Ph.D student of the LERIA at the University of Angers since 2010. His master thesis is about expressing Semantic Web ontologies with Answer Set Programming. His Ph.D thesis is about the use of preferences in order to merge cognitive maps

Israël-César Lerman is an Emeritus Professor from the Rennes 1 University, France and a researcher at the IRISA institute of Rennes in the Data and Knowledge Management Department. His research domain is Data Classification and Data Mining. His most important contribution adresses the problem of probabilistic comparison between complex structures in data analysis and in data mining. He recieved the diploma of "Docteur ès Sciences Mathématiques" in 1971 at the Paris 6 University. He wrote two books in 1970 and 1981 and more than one hundred papers, mostly in French. His second book "Classification et Analyse Ordinale des Données" (Dunod, 1981) is on the site `http://thames.cs.rhul.ac.uk/~bcs/books`.

Freddy Limpens is a doctor in Informatics from Nice — Sophia Antipolis University. He conducted his research at Inria research center on possible ways to bridge Social Web and Semantic Web. He is interested in hacker philosopy and DIY, alternative uses of technology, Art/Science/Philosophy connections, social tagging, and collaborative organisation of shared knowledge.

Stéphane Loiseau is a professor in computer science. After a Ph.d at the university of Paris 11-Orsay, he received his Accreditation to Supervise Research (HDR) in January 1998. He is full professor at the Angers university. His major interests are Knowlegde base validation, visual knowledge model (conceptual graphs, semantic map, ...) and human-machine interaction.

Patrick Marcel is assistant professor at the Computer Science Department and the Computer Science Laboratory of Université François-Rabelais Tours, France, since

1999. He received is PhD from INSA Lyon in 1998. He has more than 30 publications in referred conferences and journals. He has been reviewer for several conferences. His research interests include database query languages, On-Line Analytical Processing, Knowledge Discovery in Databases, query personalization and recommendation in databases.

Ammar Mechouche obtained an engineering degree in computer science from the Science and Technology University in Algiers, in 2004. In 2005, he obtained a master degree in Artificial Intelligence from the Pierre et Marie Curie University in Paris. After that, he obtained his PhD from the University of Rennes 1 in 2009. During his thesis, he worked on semanic web, ontologies and the semantic annotation of brain MRI images. He then look up a postdoc position at the French Mapping Agency (IGN), where he worked at the Cogit laboratory on geodata discovering and integration using ontologies. He also worked on methodologies for comparing heterogenuous ontologies. In late 2010, he joined the Aix-Marseille University and the LSIS laboratory in Marseille as a research assistant. Since 2011, he works at Thales group as a research engineer.

Sébastien Mustière is a geographical and cartographical state works engineer, and the leader of the Cogit laboratory at IGN. He obtained his PhD in computer science from the Pierre et Marie Curie University in 2001. He carries out his research activities on generalisation at IGN. He also carried out a postdoc position in data processing and GIS at Laval University in Canada.

Elsa Negre received her Ph.D. in CS in 2009 from Université François-Rabelais Tours, France. She is currently an Assistant Professor at Université Paris-Dauphine, France. Her research interests include query recommendation and personalization, data warehousing and social network analysis.

Pascal Poncelet is a Professor and the head of the data mining research group (TATOO) in the LIRMM laboratory. Professor Poncelet has previously worked as lecturer (1993–1994), as associate professor respectively in the Mediterannée University (1994–1999) and Montpellier University (1999–2001), as Professor at the Ecole des Mines d'Alès in France where he was also head of the KDD (Knowledge Discovery for Decision Making) team and co-head of the Computer Science Department (2001–2008). His research interest can be summarized as advanced data analysis techniques for emerging applications. He is currently interested in various techniques of data mining with application in Web Mining and Text Mining. He has published a large number of research papers in refereed journals, conference, and workshops, and been reviewer for some leading academic journals. He was also co-head of the French CNRS Group I3 on Data Mining.

Emeric Prouteau obtained his Master degree from the University of Le Mirail in 2010. He was internship at the Cogit laboratory from april 2010 to september 2010.

Julien Rabatel received his Ph.D. degree in Computer Science from the University of Montpellier II, France, in 2011. He was then a member of the data mining group (TATOO) of the LIRMM Laboratory. He is currently a post-doctoral researcher in the Katholieke Universiteit Leuven, Belgium. His research activities mainly concern pattern mining in structured data such as sequential or graph data, as well as its applications in various health (drug design, chemoinformatics) and industrial (sensor data analysis) contexts.

Djamel Abdelkader Zighed is Professor in computer science at the Lyon 2 University. He is the head of the Human Sciences Institute and he was Director of the ERIC Laboratory (University of Lyon). He is also the coordinator of the Erasmus Mundus Master Program on Data Mining and Knowledge Management (DMKM). He is also member of various international and national program committees.

About the Editors

Fabrice Guillet is a CS professor at Polytech'Nantes, the graduate engineering school of University of Nantes, and a member of the "KnOwledge and Decision" team (COD) of the LINA laboratory. He received a PhD degree in CS in 1995 from the "École Nationale Supérieure des Télécommunications de Bretagne", and his Habilitation (HdR) in 2006 from Nantes university. He is a co-founder of the International French-speaking "Extraction et Gestion des Connaissances (EGC)" society. His research interests include knowledge quality and knowledge visualization in the frameworks of Data Mining and Knowledge Management. He has recently co-edited two refereed books of chapter entitled "Quality Measures in Data Mining" and "Statistical Implicative Analysis — Theory and Applications" published by Springer in 2007 and 2008.

Bruno Pinaud received the PhD degree in Computer Science in 2006 from the University of Nantes. He is currently assistant professor at the University of Bordeaux I in the Computer Science Department since September 2008. His current research interests are visual data mining, graph rewriting systems, graph visualization and experimental evaluation in HCI (Human Computer Interaction). He successfully organized the 2012 edition of the EGC Conference.

Gilles Venturini is a CS Professor at François Rabelais University of Tours (France). His main researches interests concern visual data mining, virtual reality, 3D acquisition, biomimetic algorithms (genetic algorithms, artificial ants). He is at the head of the Fovea research group of the CS Laboratory of the University of Tours. He is co-editor in chief of the French New IT Journal (Revue des Nouvelles Technologies de l'Information) and was recently elected as President of the EGC society.

Djamel Abdelkader Zighed is a CS Professor at the Lyon 2 University. He is the head of the Human Sciences Institute and he was Director of the ERIC Laboratory (University of Lyon). He is also the coordinator of the Erasmus Mundus Master Program on Data Mining and Knowledge Management (DMKM). He is also member of various international and national program committees.