Complex Data Analytics with Formal Concept Analysis

Rokia Missaoui • Léonard Kwuida
Talel Abdessalem
Editors

Complex Data Analytics with Formal Concept Analysis

Springer

Editors
Rokia Missaoui
University of Quebec in Outaouais
Gatineau, QC, Canada

Talel Abdessalem
Place Marguerite Perey
Télécom-Paris, Institut Polytechnique
Palaiseau, France

Léonard Kwuida
Business School
Bern University of Applied Sciences
Bern, Switzerland

ISBN 978-3-030-93277-0 ISBN 978-3-030-93278-7 (eBook)
https://doi.org/10.1007/978-3-030-93278-7

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

In memory of Vincent Duquenne
27.06.1950–26.02.2020.

Foreword

It was in 1970s that Galois connections and respective lattices of closed sets (studied before in Mathematics by Garrett Birkhoff and Oystein Ore) were found useful for modeling information structures and processes [3,26]. After several decades of research in formal concept analysis (FCA) no one can say now that FCA proposes hardly scalable techniques for the analysis of binary data. Highly efficient FCA algorithms with various options for approximation strategies are now widely used for the analysis of complex voluminous heterogeneous data.

In spite of intrinsic complexity of computational problems related to unrestricted generation of both formal concepts and implication bases [15,16], several efficient FCA algorithms were found already around year 2000 [10,19] and new efficient implementations show excellent scalability [1]. Numerous approaches to partial generation of concepts and implications were proposed, based on interestingness constraints [17] and probabilistic considerations [2,5].

Models for treating complex data with FCA-based approaches are manifold. Several approaches were proposed and developed for relational data, first considered through the prism of conceptual scaling, which reduces complex data to binary (or unary, in terms of Rudolf Wille). Most popular approaches to treating complex data in FCA "directly", i.e., without binarizing (scaling, in FCA terms) them, are logical concept analysis [7], pattern structures [9], fuzzy concept analysis [4], relational concept analysis [11], triadic concept analysis [13,20,22,23], polyadic CA [25], and probabilistic FCA [14]. Recent interest in natural language processing, knowledge graphs, and social network analysis inspired development of new FCA-based approaches [6,8,12,21].

The recent wave of interest in deep neural networks is tempered by the problems of explainability and robustness of proposed solutions. For some applied domains, like medicine, law, and finance, these issues are crucial: experts would not accept efficient accurate solutions that do not provide acceptable explanations. FCA can propose a broad scope of tools for finding interpretable solutions, since explainability is in the core of FCA. Several attempts were made already to combine neural network efficiency with explainability provided by FCA-based approaches [18,24].

This volume presents an important step in all the above-mentioned directions: meeting the challenge of big and complex data, combining FCA-based approaches with methods based on neural networks to guarantee explainability of results.

Moscow, Russia
August 2020

Sergei O. Kuznetsov

References

1. Simon Andrews. A 'best-of-breed' approach for designing a fast algorithm for computing fixpoints of galois connections. *Information Sciences*, 295:633–649, 2015.
2. Albert Atserias, José L. Balcázar, and Marie Ely Piceno. Relative entailment among probabilistic implications. *Log. Methods Comput. Sci.*, 15(1), 2019.
3. Marc Barbut and Bernard Monjardet. *Ordre et classification: algèbre et combinatoire*. Hachette, Paris, 1970.
4. Radim Belohlávek. Lattices of fixed points of fuzzy galois connections. *Math. Log. Q.*, 47(1):111–116, 2001.
5. Daniel Borchmann, Tom Hanika, and Sergei Obiedkov. Probably approximately correct learning of horn envelopes from queries. *Discrete Applied Mathematics*, 273:30–42, 2020.
6. Sébastien Ferré and Peggy Cellier. Graph-fca: An extension of formal concept analysis to knowledge graphs. *Discret. Appl. Math.*, 273:81–102, 2020.
7. Sébastien Ferré and Olivier Ridoux. A logical generalization of formal concept analysis. In Bernhard Ganter and Guy W. Mineau, editors, *Conceptual Structures: Logical, Linguistic, and Computational Issues, 8th International Conference on Conceptual Structures, ICCS 2000, Darmstadt, Germany, August 14–18, 2000, Proceedings*, volume 1867 of *Lecture Notes in Computer Science*, pages 371–384. Springer, 2000.
8. Boris A. Galitsky, Dmitry I. Ilvovsky, and Sergey O. Kuznetsov. Detecting logical argumentation in text via communicative discourse tree. *J. Exp. Theor. Artif. Intell.*, 30(5):637–663, 2018.
9. Bernhard Ganter and Sergei O. Kuznetsov. Pattern structures and their projections. In Harry S. Delugach and Gerd Stumme, editors, *Conceptual Structures: Broadening the Base, 9th International Conference on Conceptual Structures, ICCS 2001, Stanford, CA, USA, July 30-August 3, 2001, Proceedings*, volume 2120 of *Lecture Notes in Computer Science*, pages 129–142. Springer, 2001.
10. Robert Godin, Rokia Missaoui, and Hassan Alaoui. Incremental concept formation algorithms based on galois (concept) lattices. *Computational Intelligence*, 11:246–267, 1995.
11. Mohamed Rouane Hacene, Marianne Huchard, Amedeo Napoli, and Petko Valtchev. Relational concept analysis: mining concept lattices from multi-relational data. *Ann. Math. Artif. Intell.*, 67(1):81–108, 2013.

12. Mohamed Hamza Ibrahim, Rokia Missaoui, and Abir Messaoudi. Detecting communities in social networks using concept interestingness. In Iosif-Viorel Onut, Andrew Jaramillo, Guy-Vincent Jourdan, Dorina C. Petriu, and Wang Chen, editors, *Proceedings of the 28th Annual International Conference on Computer Science and Software Engineering, CASCON 2018, Markham, Ontario, Canada, October 29–31, 2018*, pages 81–90. ACM, 2018.
13. Dmitry I. Ignatov, Dmitry V. Gnatyshak, Sergei O. Kuznetsov, and Boris G. Mirkin. Triadic formal concept analysis and triclustering: searching for optimal patterns. *Mach. Learn.*, 101(1–3):271–302, 2015.
14. Francesco Kriegel. Implications over probabilistic attributes. In Karell Bertet, Daniel Borchmann, Peggy Cellier, and Sébastien Ferré, editors, *Formal Concept Analysis - 14th International Conference, ICFCA 2017, Rennes, France, June 13–16, 2017, Proceedings*, volume 10308 of *Lecture Notes in Computer Science*, pages 168–183. Springer, 2017.
15. Sergei Kuznetsov. On the intractability of computing the duquenne-guigues base. *Journal of Universal Computer Science*, 10:927–933, 01 2004.
16. Sergei O. Kuznetsov. On computing the size of a lattice and related decision problems. *Order*, 18(4):313–321, 2001.
17. Sergei O. Kuznetsov and Tatiana P. Makhalova. On interestingness measures of formal concepts. *Information Sciences*, 442–443:202–219, 2018.
18. Sergei O. Kuznetsov, Nurtas Makhazhanov, and Maxim Ushakov. On neural network architecture based on concept lattices. In Marzena Kryszkiewicz, Annalisa Appice, Dominik Slezak, Henryk Rybinski, Andrzej Skowron, and Zbigniew W. Ras, editors, *Foundations of Intelligent Systems - 23rd International Symposium, ISMIS 2017, Warsaw, Poland, June 26–29, 2017, Proceedings*, volume 10352 of *Lecture Notes in Computer Science*, pages 653–663. Springer, 2017.
19. Sergei O. Kuznetsov and Sergei A. Obiedkov. Comparing performance of algorithms for generating concept lattices. *J. Exp. Theor. Artif. Intell.*, 14(2–3):189–216, 2002.
20. Fritz Lehmann and Rudolf Wille. A triadic approach to formal concept analysis. In *ICCS*, pages 32–43, 1995.
21. Rokia Missaoui, Sergei O. Kuznetsov, and Sergei A. Obiedkov, editors. *Formal Concept Analysis of Social Networks*. Lecture Notes in Social Networks. Springer, 2017.
22. Rokia Missaoui and Léonard Kwuida. Mining Triadic Association Rules from Ternary Relations. In *LNAI*, volume 6628, pages 204–218. 2011.
23. Rokia Missaoui, Pedro H. B. Ruas, Léonard Kwuida, and Mark A. J. Song. Pattern discovery in triadic contexts. In Mehwish Alam, Tanya Braun, and Bruno Yun, editors, *Ontologies and Concepts in Mind and Machine - 25th International Conference on Conceptual Structures, ICCS 2020, Bolzano, Italy, September 18–20, 2020, Proceedings*, volume 12277 of *Lecture Notes in Computer Science*, pages 117–131. Springer, 2020.
24. Sebastian Rudolph. Using FCA for encoding closure operators into neural networks. In Uta Priss, Simon Polovina, and Richard Hill, editors, *Conceptual Structures: Knowledge Architectures for Smart Applications, 15th International*

Conference on Conceptual Structures, ICCS 2007, Sheffield, UK, July 22–27, 2007, Proceedings, volume 4604 of *Lecture Notes in Computer Science*, pages 321–332. Springer, 2007.
25. George Voutsadakis. Polyadic concept analysis. *Order*, 19(3):295–304, 2002.
26. Rudolf Wille. Restructuring lattice theory: An approach based on hierarchies of concepts. In Ivan Rival, editor, *Ordered Sets*, volume 83 of *NATO Advanced Study Institutes Series*, pages 445–470. Springer Netherlands, 1982.

Preface

With the advent of complex and big data and the increasing number of studies towards their management and analysis, it becomes important to get a better insight into existing studies, trends, and challenges and rely on promising theories such as formal concept analysis (FCA) together with recently developed technologies to design new, accurate, and scalable solutions for big data analytics facilities.

FCA is an important formalism that is associated with a variety of research areas such as lattice theory, knowledge representation, data mining, machine learning, and semantic Web, to name a few. It is successfully exploited in an increasing number of application domains such as software engineering, information retrieval, social network analysis, and bioinformatics. The mathematical power of FCA comes from its concept lattice formalization in which each element captures a formal concept while the whole structure represents a hierarchy that offers conceptual clustering, browsing, and association rule mining.

Although there are significant theoretical and practical contributions within the FCA community, including the design and implementation of efficient algorithms and tools for concept lattice computation and exploitation, this book examines a set of important and relevant research directions in complex data management and updates the contribution of the FCA community for analyzing complex and large data. For example, formal concept analysis and some of its extensions are exploited, revisited, and coupled with recent processing paradigms to maximize the benefits in analyzing large data. This book is a follow-up project of the workshop BigFCA'2019—Formal Concept Analysis in the Big Data Era—which was jointly organized with the ICFCA'2019 Conference in Frankfurt (see https://icfca2019.frankfurt-university.de/bigfca.html).

This volume of eleven chapters is meant to cover the state of the art of the research on the intersection of FCA and complex data analysis in a more systematic and detailed manner than it was done in the workshop proceedings mentioned above.

Gatineau, QC, Canada — Rokia Missaoui
Bern, Switzerland — Léonard Kwuida
Palaiseau, France — Talel Abdessalem
September 2020

Acknowledgments

The production of this volume would not have been possible without the valuable involvement and efforts of the contributing authors and the following reviewers: Simon Andrews, Jaume Baixeries, Karell Bertet, Agnès Braud, Víctor Codocedo, Pablo Cordero, Miguel Couceiro, Christophe Demko, Manuel Enciso, Sébastien Ferré, Tom Hanika, Mohamed Ibrahim, Dmitry Ignatov, Mehdi Kaytoue, Léonard Kwuida, Florence Le Ber, Pierre Martin, Ángel Mora Bonilla, Emilio Muñoz-Velasco, Amedeo Napoli, Manuel Ojeda-Aciego, Domingo López-Rodríguez, and Uta Priss. We highly appreciate the efforts and commitment of all the authors and reviewers. Our warm thanks go to Sergei O. Kuznetsov who wrote the foreword. We also would like to express our gratitude to the Editorial Director Ronan Nugent and the Senior Editor Paul Drougas from Springer for their help in the preparation of this volume. Finally, the production of this volume was greatly facilitated thanks to two excellent and well-known platforms, namely *EasyChair* for organizing paper submission and review, and *Overleaf* as an efficient and user-friendly collaborative LaTeX editor.

Contents

List of Contributors

Jaume Baixeries
Computer Science Department, Universitat Politècnica de Catalunya, Barcelona, Spain

Meryem Bendella
Aix-Marseille Université, Marseille, France

Karell Bertet
Laboratory L3i, La Rochelle University, La Rochelle, France

Salah Boukhetta
Laboratory L3i, La Rochelle University, La Rochelle, France

Agnès Braud
ICube UMR 7537, Universite de Strasbourg, CNRS, ENGEES, Strasbourg, France

Michael Broughton
Defence Science and Technology Group, Adelaide, SA, Australia

Víctor Codocedo
Departamento de Informática, Universidad Técnica Federico Santa María, Campus San Joaquin, Santiago de Chile, Chile

Pablo Cordero
Universidad de Málaga, Málaga, Spain

Miguel Couceiro
Université de Lorraine, CNRS, Inria, LORIA, Nancy, France

Christophe Demko
Laboratory L3i, La Rochelle University, La Rochelle, France

Xavier Dolques
ICube UMR 7537, Universite de Strasbourg, CNRS, ENGEES, Strasbourg, France

Dominik Dürrschnabel
Knowledge & Data Engineering Group, University of Kassel, Kassel, Germany
Interdisciplinary Research Center for Information System Design, University of Kassel, Kassel, Germany

Dmitry Egurnov
National Research University Higher School of Economics, Moscow, Russia

Manuel Enciso
Universidad de Málaga, Málaga, Spain

Cyril Faucher
Laboratory L3i, La Rochelle University, La Rochelle, France

Sébastien Ferré
Univ Rennes, CNRS, IRISA, Rennes, France

Alain Gély
Université de Lorraine, CNRS, LORIA, Metz, France

Alain Gutierrez
LIRMM, Univ Montpellier, CNRS, Montpellier, France

Tom Hanika
Knowledge & Data Engineering Group, University of Kassel, Kassel, Germany
Interdisciplinary Research Center for Information System Design, University of Kassel, Kassel, Germany

Marianne Huchard
LIRMM, Univ Montpellier, CNRS, Montpellier, France

Dmitry I. Ignatov
National Research University Higher School of Economics, Moscow, Russia

Mehdi Kaytoue
Infologic R&D, Bourg-Lès-Valence, France

Priscilla Keip
CIRAD, UPR AIDA, Montpellier, France
AIDA, Univ Montpellier, CIRAD, Montpellier, France

Léonard Kwuida
Business School, Bern University of Applied Sciences, Bern, Switzerland

Florence Le Ber
ICube UMR 7537, Universite de Strasbourg, CNRS, ENGEES, Strasbourg, France

Domingo López-Rodríguez
Universidad de Malaga, Departamento Matematica Aplicada, Malaga, Spain

Pierre Martin
CIRAD, UPR AIDA, Montpellier, France
AIDA, Univ Montpellier, CIRAD, Montpellier, France

Rokia Missaoui
University of Quebec in Outaouais, Gatineau, QC, Canada

Ángel Mora
Universidad de Málaga, Málaga, Spain

Emilio Muñoz-Velasco
Universidad de Malaga, Departamento Matematica Aplicada, Malaga, Spain

Amedeo Napoli
Université de Lorraine, CNRS, Inria, LORIA, Nancy, France

Cristina Nica
Nicolae Titulescu University of Bucharest, Bucharest, Romania

Manuel Ojeda-Aciego
Universidad de Malaga, Departamento Matematica Aplicada, Malaga, Spain

Tim Pattison
Defence Science and Technology Group, Adelaide, SA, Australia

Mohamed Quafafou
Aix-Marseille Université, Marseille, France

Jérémy Richard
Laboratory L3i, La Rochelle University, La Rochelle, France

Pierre Silvie
IRD, UMR IPME, Montpellier, France
CIRAD, UPR AIDA, Montpellier, France
AIDA, Univ Montpellier, CIRAD, Montpellier, France

Maximilian Stubbemann
Knowledge & Data Engineering Group, University of Kassel, Kassel, Germany
Interdisciplinary Research Center for Information System Design, University of Kassel, Kassel, Germany
L3S Research Center, Leibniz University Hannover, Hannover, Germany

Dmitry Tochilkin
National Research University Higher School of Economics, Moscow, Russia

Derek Weber
Defence Science and Technology Group, Adelaide, SA, Australia

Acronyms

ACN	Abstract Conceptual Navigation
FCA	Formal Concept Analysis
FD	Functional Dependency
IR	Information Retrieval
KG	Knowledge Graph
NLP	Natural Language Processing
OA	Object-Attribute
OAC	Object-Attribute-Condition
PAC	Probably Approximately Correct
PGP	Projected Graph Pattern
RCA	Relational Concept Analysis
TCA	Triadic Concept Analysis