



<http://dl181.zlibcdn.com/covers/books/ba/6d/fe/ba6dfec43d01f5d712813df1d0d14aa.jpg>

Business and Consumer Analytics: New Ideas

Pablo Moscato ([/g/Pablo Moscato](#)), *Natalie Jane de Vries* ([/g/Natalie Jane de Vries](#))

This two-volume handbook presents a collection of novel methodologies with applications and illustrative examples in the areas of data-driven computational social sciences. Throughout this handbook, the focus is kept specifically on business and consumer-oriented applications with interesting sections ranging from clustering and network analysis, meta-analytics, memetic algorithms, machine

Pablo Moscato • Natalie Jane de Vries
Editors

Business and Consumer Analytics: New Ideas

 Springer

Editors

Pablo Moscato
School of Electrical Engineering
and Computing
The University of Newcastle
Callaghan, NSW, Australia

Natalie Jane de Vries
School of Electrical Engineering
and Computing
The University of Newcastle
Callaghan, NSW, Australia

ISBN 978-3-030-06221-7 ISBN 978-3-030-06222-4 (eBook)
<https://doi.org/10.1007/978-3-030-06222-4>

Library of Congress Control Number: 2019934369

© Springer Nature Switzerland AG 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

Preface

It is hard to find something else to say about data science today that has not already been said. Perhaps what has not yet been discussed is that the progress in this area is profoundly challenging the way concepts have to be presented to newcomers. Data science brings together concepts from disciplines like computer science, statistics and applied mathematics, and the applications reach all possible aspects of life and economics. Consequently, universities around the world are having difficulties in addressing the need for a reformulation of their courses. There is a clear need for new ways to present the material, with an emphasis in understanding the key concepts, the novel applications and the impact of the techniques. This is, undoubtedly, a huge endeavour as there is no universally established curriculum for data science. In fact, we feel there is a need for students and practitioners who have been trained in one particular field to find a “shortcut” that would allow them to understand other areas.

This book can be seen as a first step in that direction. We aim at “bridging the gap” between some core new ideas in data science with the application in business and consumer analytics.

There are several reasons to choose this intersection as our first target. Advances in data science, data analytics and data mining methodologies are bringing many novel contributions to business and marketing applications. On the other side, the scale of e-commerce activities and the possibility of reaching a novel understanding of consumer behaviour are a driving force that pushes and challenges the field of data science. It is clear that the trend is here to stay. Conducting business and speaking out to consumers will be forever paired with data analytics. We have gained an incredible capability for collecting large amounts of widely varied data, and providing business insights from these data sources has become an important and continuous task of many researchers and business professionals.

At the time of writing this work, all the techniques included are considered novel in the area of business and consumer analytics. We are sure that more work needs to be done for many of them to reach the scalability necessary to deal with datasets of millions of consumers and products. That is a necessity of daily operations of many companies; we feel that this need for scalability will be met by the natural

algorithmic progress provided by computer science. This said, we are interested in the new ideas currently at the crossroads between developments in data science, optimization, network analytics, computational complexity, artificial intelligence and machine learning, evolutionary algorithms and their application to business scenarios. We are confident that many of these techniques will soon flourish and become more widely adopted by businesses.

There are several lessons that are normally learnt when you finish a book. For the two editors, some of them were early ones at the beginning of the work, and they, in turn, helped to mould the introductory section to address them. For instance, the overzealous preoccupation of computer scientists with the provision of highly predictive analytic learning systems often clashes with the interest of business professionals. The latter tend to prefer models with perhaps less variables, at a cost of having a reduced predictive capability, that nevertheless contain the necessary levers that can lead to improved decision-making, reducing cost and maximizing profit. For both sides of the intersection we are looking at (i.e., data scientists and business professionals), even the word “problem” conveys a different meaning, and the quest of “efficiency” in computer science may be misinterpreted by a business marketer. We also aimed at presenting some fundamentals on marketing and consumer behaviour to enlighten the “hardcore computer scientists” about some of the needs and wants of marketing and business professionals. We quote: *“Data has been king for well over a decade by now, but the way we use it is undergoing some serious change. Gone are the days of awe at pretty charts and heat maps. Gone, too, is any patience for analytics unaligned to action”*.¹ These differences in the use of language and purpose are discussed in a tutorial way to help engage both communities towards a common goal. Once again, although we are certain that more needs to be done, our intention was clear; we tried to fill this gap.

Apart from the two introductory chapters, the rest of the book is organized as follows. Each section is centred in one specific area of fast development, which is either methodological or application based. The clustering and pattern mining section contains recent developments that are important for customer segmentation and targeting. They go from an introductory tone (Chaps. 3 and 4 are of that type), while the other two chapters relate to more advanced methods currently under development.

The network section contains a review chapter, which again is of an introductory nature, followed by an introduction to the area of centrality analysis (which is a main topic for areas of computational social science and product analyses). The remaining four chapters provide a variety of methods and applications related to network analytics including novel applications in survey analyses and co-purchasing network analytics.

In the 1980s, a new word was coined by Fred Glover called “metaheuristics”, which are generic techniques aimed at guiding heuristics for a problem at hand

¹<https://www.information-management.com/opinion/data-science-underlies-everything-the-enterprise-now-does>.

which can be posed as an optimization task. Analogously, he is proposing the use of the word “meta-analytics” for identifying high-level techniques aiming at guiding a multitude of methods working together to address a data analytic problem. These ideas are introduced, and then several chapters are dedicated to specific implementations of these methodologies: two with a focus in ensemble learning and one in a classification problem.

Memetic algorithms are a paradigm that was championed by one of the editors, and during the last years, it has become a highly active field of research following the publication of the *Handbook of Memetic Algorithms* by Springer in 2011. While applications in many fields of science and technology exist, there is an enormous potential for the area of business and consumer analytics. The chapters included in this work are just the tip of the iceberg of the activity that currently exists. It includes applications in product and customer visualization, personalized recommendation systems, facility location and vehicle routing, and orienting problems. We also note that one of the contributions in the network section includes a memetic algorithm application addressing overlapping community identification in co-purchasing graphs from the Amazon group.

The final section includes a more application-oriented major theme around travel and fashion analytics with three contributions in tourism and one in recommendation systems for a fashion e-commerce service. While the major theme is on personalization of services, the chapters are self-contained and methodologically address the questions using different techniques (fuzzy clustering, mathematical programming and ranking-based techniques).

We have tried, whenever possible, to use the same dataset in some of the different chapters. This allows readers to understand the possibilities that the different techniques gave individually and to concentrate the descriptions of all the datasets in a single Appendix chapter at the end of this book.

Callaghan, NSW, Australia

Pablo Moscato
Natalie Jane de Vries

Acknowledgements

The editors would like to thank a multitude of people who helped create this interdisciplinary and collaborative piece of work. Firstly, our thanks go to Luke Mathieson, Sergio Gomez, Carlos Cotta and Fred Glover who helped us to put together the various sections in this book by inviting fellow collaborators and sharing co-edition. We would like to thank all of the reviewers and proofreaders of the many contributions in this volume, those who did so anonymously and those who did so for fellow colleagues. Further, we would like to thank Ademir Cristiano Gabardo and Mohammad Nazmul Haque for their help in a range of things from formatting to technical issues and editing figures.

Finally, the editors would like to thank all of the *data parasites*^{2,3} around the world who, like us, strive for a transparent and open research environment where collaboration, not competition, is the main motivation for researchers sharing data and where we strive towards the common goal of the advancement of science and human knowledge. We humbly dedicate our “Open Access Data” Appendix chapter to you!

Natalie de Vries would like to thank her family, friends and partner Simon Faint for their personal support of all her professional and academic endeavours. She would also like to thank fellow editor Pablo Moscato for the great experience and memories gained from creating this book together.

Pablo Moscato thanks Regina and Anna Clara for their patience and the support through the many stages of this project. He also thanks his fellow co-editor,

²*Data parasites* is a term that came to be popular after a rather controversial editorial published in *The New England Journal of Medicine* (NEJM) in January 2016 (<http://www.nejm.org/doi/full/10.1056/NEJMe1516564>) criticized the sharing of data. This editorial received a huge backlash from the research community who are in support of open access trends and data sharing activities. The satirical “Data Parasite Award” has since also been established handing out two awards: “the first recognizes an outstanding contribution from a junior parasite (postdoctoral, graduate, or undergraduate trainee), and the second recognizes an individual for a sustained period of exemplary research parasitism”.

³<http://researchparasite.com/>.

Ms. de Vries, for being essential in their “largest catch” together. He also considers that, while a few cats, but no sharks, slowed the team to get the fish, not even “the great DiMaggio who does all things perfectly” could have been a better partner in this boat.

Looking back, and measuring all the obstacles to their projects created by some people over the last years, the editors say: “*Eat that galanos. And make a dream you’ve killed us*’. This book is also dedicated to you.”

Contents

Part I Introduction

- 1 **Marketing Meets Data Science: Bridging the Gap**..... 3
Pablo Moscato and Natalie Jane de Vries
- 2 **Consumer Behaviour and Marketing Fundamentals
for Business Data Analytics** 119
Natalie Jane de Vries and Pablo Moscato

Part II Segmentation, Clustering and Pattern Mining

- 3 **Introducing Clustering with a Focus in Marketing
and Consumer Analysis** 165
Natalie Jane de Vries, Łukasz P. Olech, and Pablo Moscato
- 4 **An Introduction to Proximity Graphs**..... 213
Luke Mathieson and Pablo Moscato
- 5 **Clustering Consumers and Cluster-Specific Behavioural Models**..... 235
Natalie Jane de Vries, Jamie Carlson and Pablo Moscato
- 6 **Frequent Itemset Mining** 269
Massimo Cafaro and Marco Pulimeno

Part III Networks and Community Detection

- 7 **Business Network Analytics: From Graphs to Supernetworks**..... 307
Pablo Moscato
- 8 **Centrality in Networks: Finding the Most Important Nodes** 401
Sergio Gómez
- 9 **Overlapping Communities in Co-purchasing and Social
Interaction Graphs: A Memetic Approach** 435
Ademir Gabardo, Regina Berretta and Pablo Moscato

10	Taming a Graph Hairball: Local Exploration in a Global Context ...	467
	James Abello, Daniel Mawhirter, and Kevin Sun	
11	Network-Based Models for Social Recommender Systems	491
	Antonia Godoy-Lorite, Roger Guimerà, and Marta Sales-Pardo	
12	Using Network Alignment to Identify Conserved Consumer Behaviour Modelling Constructs	513
	Luke Mathieson, Natalie Jane de Vries, and Pablo Moscato	
Part IV Memetic Algorithms		
13	Memetic Algorithms for Business Analytics and Data Science: A Brief Survey	545
	Pablo Moscato and Luke Mathieson	
14	A Memetic Algorithm for the Team Orienteering Problem	609
	Dimitra Trachanatzi, Eleftherios Tsakirakis, Magdalene Marinaki, Yannis Marinakis, and Nikolaos Matsatsinis	
15	A Memetic Algorithm for Competitive Facility Location Problems ..	637
	Benjamin Biesinger, Bin Hu, and Günther R. Raidl	
16	Visualizing Products and Consumers: A Gestalt Theory Inspired Method	661
	Claudio Sanhueza Lobos, Natalie Jane de Vries, Mario Inostroza-Ponta, Regina Berretta, and Pablo Moscato	
Part V Meta-Analytics		
17	An Overview of Meta-Analytics: The Promise of Unifying Metaheuristics and Analytics	693
	Fred Glover and Carlos Cotta	
18	From Ensemble Learning to Meta-Analytics: A Review on Trends in Business Applications	703
	Mohammad Nazmul Haque and Pablo Moscato	
19	Metaheuristics and Classifier Ensembles	733
	Ringolf Thomschke, Stefan Voß, and Stefan Lessmann	
20	A Multi-objective Meta-Analytic Method for Customer Churn Prediction	781
	Mohammad Nazmul Haque, Natalie Jane de Vries, and Pablo Moscato	
21	Hotel Classification Using Meta-Analytics: A Case Study with Cohesive Clustering	815
	Buyang Cao, Cesar Rego, and Fred Glover	

Part VI Data Science Applications in Travel and Fashion Analytics

22 Fuzzy Clustering in Travel and Tourism Analytics	839
Pierpaolo D’Urso, Marta Disegna, and Riccardo Massari	
23 Towards Personalized Data-Driven Bundle Design with QoS Constraint	865
Mustafa Mısıř and Hoong Chuin Lau	
24 A Fuzzy Evaluation of Tourism Sustainability	911
Joseph Andria, Giacomo di Tollo, and Raffaele Pesenti	
25 New Ideas in Ranking for Personalized Fashion Recommender Systems	933
Heri Ramampiaro, Helge Langseth, Thomas Almenningen, Herman Schistad, Martin Havig, and Hai Thanh Nguyen	
Part VII Appendix	
26 Datasets for Business and Consumer Analytics	965
Natalie Jane de Vries and Pablo Moscato	
Index	989

Contributors

James Abello DIMACS and Computer Science Department, Rutgers University, New Brunswick, NJ, USA

Thomas Almenningen Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway

Joseph Andria Dipartimento di Scienze Economiche, Aziendali e Statistiche, University of Palermo, Palermo, Italy

Regina Berretta School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW, Australia

Benjamin Biesinger Institute of Computer Graphics and Algorithms, TU Wien, Vienna, Austria

Massimo Cafaro University of Salento, Lecce, Italy

Buyang Cao China Intelligent Urbanization Co-Creation Center for High Density Region, School of Software Engineering, Tongji University, Shanghai, China

Jamie Carlson Newcastle University Business School, The University of Newcastle, Callaghan, NSW, Australia

Carlos Cotta ETSI Informática, Universidad De Málaga, Malaga, Spain

Natalie Jane de Vries School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW, Australia

Giacomo di Tollo Dipartimento di Economia, Università Ca' Foscari, Venezia, Italy

Marta Disegna Accounting, Finance & Economics Department, Faculty of Management, Bournemouth University, Bournemouth, UK

Pierpaolo D'Urso Department of Social Sciences and Economics, Sapienza University of Roma, Roma, Italy

Ademir Gabardo School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW, Australia

Fred Glover Leeds School of Business and the College of Engineering & Applied Science, University of Colorado, Boulder, CO, USA

Antonia Godoy-Lorite Department of Mathematics, Imperial College London, London, UK

Sergio Gómez Universitat Rovira i Virgili, Tarragona, Spain

Roger Guimerà Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain

Departament d'Enginyeria, Química, Universitat Rovira i Virgili, Tarragona, Catalonia, Spain

Mohammad Nazmul Haque School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW, Australia

Martin Havig Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway

Bin Hu AIT Austrian Institute of Technology, Mobility Department - Dynamic Transportation Systems, Vienna, Austria

Mario Inostroza-Ponta Universidad de Santiago de Chile, Santiago, Chile

Helge Langseth Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway

Hoong Chuin Lau Singapore Management University, School of Information Systems, Singapore, Singapore

Stefan Lessmann Humboldt-Universität zu Berlin, School of Business and Economics, Berlin, Germany

Magdalene Marinaki Technical University of Crete, School of Production Engineering and Management, Chania, Greece

Yannis Marinakis Technical University of Crete, School of Production Engineering and Management, Chania, Greece

Riccardo Massari Department of Social Sciences and Economics, Sapienza University of Roma, Roma, Italy

Luke Mathieson School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW, Australia

School of Software, University of Technology Sydney, Ultimo, NSW, Australia

Nikolaos Matsatsinis Technical University of Crete, School of Production Engineering and Management, Chania, Greece

Daniel Mawhirter Colorado School of Mines, Golden, CO, USA

Mustafa Mısıř Nanjing University of Aeronautics and Astronautics, College of Computer Science and Technology, Nanjing, Jiangsu, China

Pablo Moscato School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW, Australia

Hai Thanh Nguyen Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway

Telenor Research, Trondheim, Norway

Lukasz P. Olech Department of Computational Intelligence, Faculty of Computer Science and Management, Wrocław University of Science and Technology, Wrocław, Poland

Raffaele Pesenti Dipartimento di Management, Università Ca' Foscari, Venezia, Italy

Marco Pulimeno University of Salento, Lecce, Italy

Günther R. Raidl Institute of Computer Graphics and Algorithms, TU Wien, Vienna, Austria

Heri Ramampiaro Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway

Cesar Rego School of Business Administration, University of Mississippi, Oxford, MS, USA

Marta Sales-Pardo Departament d'Enginyeria Química, Universitat Rovira i Virgili, Tarragona, Spain

Claudio Sanhueza Lobos School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW, Australia

Herman Schistad Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway

Kevin Sun Rutgers University, New Brunswick, NJ, USA

Ringolf Thomschke Humboldt-Universität zu Berlin, School of Business and Economics, Berlin, Germany

Dimitra Trachanatzi Technical University of Crete, School of Production Engineering and Management, Chania, Greece

Eleftherios Tsakirakis Technical University of Crete, School of Production Engineering and Management, Chania, Greece

Stefan Voß University of Hamburg, Institute of Information Systems, Hamburg, Germany