

Fridolin Wild

Learning Analytics in R

with SNA, LSA,
and MPIA

 Springer

Learning Analytics in R with SNA, LSA, and MPIA

Fridolin Wild

Learning Analytics in R with SNA, LSA, and MPIA

 Springer

Fridolin Wild
Performance Augmentation Lab
Department of Computing and
Communication Technologies
Oxford Brookes University
Oxford
UK

The present work has been accepted as a doctoral thesis at the University of Regensburg, Faculty for Languages, Literature, and Cultures under the title “Learning from Meaningful, Purposive Interaction: Representing and Analysing Competence Development with Network Analysis and Natural Language Processing”.

Cover photo by Sonia Bernac (<http://www.bernac.org>)

ISBN 978-3-319-28789-8 ISBN 978-3-319-28791-1 (eBook)
DOI 10.1007/978-3-319-28791-1

Library of Congress Control Number: 2016935495

© Springer International Publishing Switzerland 2016

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.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

*To my father Kurt Wild (*1943–†2010).*

Acknowledgement

This work would not have been possible without the support of my family, my son Maximilian, my brothers Benjamin, Johannes, and Sebastian, and my parents Brunhilde and Kurt, only one of whom survived its finalisation (Dad, you are missed!).

I am indebted to Prof. Dr. Peter Scott, at the time Director of the Knowledge Media Institute of The Open University and Chair for Knowledge Media, now with the University of Technology in Sydney as Assistant Deputy Vice-Chancellor, and Prof. Dr. Christian Wolff, Chair for Media Informatics of the University of Regensburg (in alphabetical order), who both helped improve this book and saved me from my worst mistakes. Without you and our engaging discussions, this book would simply not have happened.

There are many people who contributed to this work, some of them unknowingly. I'd like to thank the coordinators of the EU research projects that led to this work, also in substitution for the countless researchers and developers with whom I had the pleasure to work in these projects over the years: Wolfgang Nejd for Prolearn, Barbara Kieslinger for iCamp, Ralf Klamma and Martin Wolpers for ROLE, Peter van Rosmalen and Wolfgang Greller for LTfLL, Peter Scott for STELLAR, Vana Kamtsiou and Lampros Stergioulas for TELmap, and Paul Lefrere, Claudia Guglielmina, Eva Coscia, and Sergio Gusmeroli for TELLME.

I'd also like to thank Prof. Dr. Gustaf Neumann, Chair for Information Systems and New Media of the Vienna University of Economics and Business (WU), for whom I worked in the years 2005–2009, before I joined Peter Scott's team at the Knowledge Media Institute of The Open University. Special thanks go to Prof. Dr. Kurt Hornik of the Vienna University of Economics and Business, who first got me hooked on R (and always had an open ear for my tricky questions). I'd also like to express my gratitude to Dr. Ambjoern Naeve of the KTH Royal Institute of Technology of Sweden for awakening my passion for mathemagics more than a decade ago and believing in me over years that I have it in me.

I'd like to particularly thank Debra Haley and Katja Buelow for the work in the LTfLL project, particularly the data collection for the experiment further analysed in Sects. 10.3.2 and 10.3.3.

I owe my thanks to my current and former colleagues at the Open University and at the Vienna University of Economics and Business, only some of which I can explicitly mention here: Thomas Ullmann, Giuseppe Scavo, Anna De Liddo, Debra Haley, Kevin Quick, Katja Buelow, Chris Valentine, Lucas Anastasiou, Simon Buckingham Shum, Rebecca Ferguson, Paul Hogan, David Meyer, Ingo Feinerer, Stefan Sobernig, Christina Stahl, Felix Moedritscher, Steinn Sigurdarson, and Robert Koblischke.

Not least, I'd like to stress that I am glad and proud to be a citizen in the European Union that has brought the inspiring research framework programmes and Horizon 2020 on the way and provided funding to those projects listed above.

Fridolin Wild

The original version of the book was revised: Acknowledgement text “Cover photo by Sonia Bernac (<http://www.bernac.org>)” has been included in the copyright page. The Erratum to the book is available at DOI 10.1007/978-3-319-28791-1_12

Contents

1	Introduction	1
1.1	The Learning Software Market	2
1.2	Unresolved Challenges: Problem Statement	3
1.3	The Research Area of Technology-Enhanced Learning	5
1.4	Epistemic Roots	7
1.5	Algorithmic Roots	9
1.6	Application Area: Learning Analytics	10
1.7	Research Objectives and Organisation of This Book	14
	References	17
2	Learning Theory and Algorithmic Quality Characteristics	23
2.1	Learning from Purposive, Meaningful Interaction	25
2.2	Introducing Information	27
2.3	Foundations of Culturalist Information Theory	28
2.4	Introducing Competence	30
2.5	On Competence and Performance Demonstrations	32
2.6	Competence Development as Information Purpose of Learning	32
2.7	Performance Collections as Purposive Sets	34
2.8	Filtering by Purpose	35
2.9	Expertise Clusters	35
2.10	Assessment by Equivalence	36
2.11	About Disambiguation	37
2.12	A Note on Texts, Sets, and Vector Spaces	38
2.13	About Proximity as a Supplement for Equivalence	39
2.14	Feature Analysis as Introspection	39
2.15	Algorithmic Quality Characteristics	40
2.16	A Note on Educational Practice	41
2.17	Limitations	41
2.18	Summary	42
	References	43

- 3 Representing and Analysing Purposiveness with SNA 45**
 - 3.1 A Brief History and Standard Use Cases 45
 - 3.2 A Foundational Example 48
 - 3.3 Extended Social Network Analysis Example 56
 - 3.4 Limitations 68
 - References 69

- 4 Representing and Analysing Meaning with LSA 71**
 - 4.1 Mathematical Foundations 73
 - 4.2 Analysis Workflow with the R Package ‘lsa’ 77
 - 4.3 Foundational Example 80
 - 4.4 State of the Art in the Application of LSA for TEL 91
 - 4.5 Extended Application Example: Automated Essay Scoring 95
 - 4.6 Limitations of Latent Semantic Analysis 101
 - References 101

- 5 Meaningful, Purposive Interaction Analysis 107**
 - 5.1 Fundamental Matrix Theorem on Orthogonality 108
 - 5.2 Solving the Eigenvalue Problem 112
 - 5.3 Example with Incidence Matrices 114
 - 5.4 Singular Value Decomposition 117
 - 5.5 Stretch-Dependent Truncation 122
 - 5.6 Updating Using Ex Post Projection 125
 - 5.7 Proximity and Identity 126
 - 5.8 A Note on Compositionality: The Difference of Point,
Centroid, and Pathway 128
 - 5.9 Performance Collections and Expertise Clusters 129
 - 5.10 Summary 130
 - References 131

- 6 Visual Analytics Using Vector Maps as Projection Surfaces 133**
 - 6.1 Proximity-Driven Link Erosion 135
 - 6.2 Planar Projection With Monotonic Convergence 136
 - 6.3 Kernel Smoothing 141
 - 6.4 Spline Tile Colouring With Hypsometric Tints 142
 - 6.5 Location, Position, and Pathway Revisited 145
 - 6.6 Summary 146
 - References 147

- 7 Calibrating for Specific Domains 149**
 - 7.1 Sampling Model 150
 - 7.2 Investigation 152
 - 7.3 Results 155
 - 7.4 Discussion 158
 - 7.5 Summary 162
 - References 163

8	Implementation: The MPIA Package	165
8.1	Use Cases for the Analyst	165
8.2	Analysis Workflow	167
8.3	Implementation: Classes of the <i>mpia</i> Package	170
8.3.1	The <i>DomainManager</i>	174
8.3.2	The <i>Domain</i>	175
8.3.3	The <i>Visualiser</i>	175
8.3.4	The <i>HumanResourceManager</i>	176
8.3.5	The <i>Person</i>	176
8.3.6	The <i>Performance</i>	177
8.3.7	The Generic Functions	177
8.4	Summary	181
	References	181
9	MPIA in Action: Example Learning Analytics	183
9.1	Brief Review of the State of the Art in Learning Analytics	184
9.2	The Foundational SNA and LSA Examples Revisited	185
9.3	Revisiting Automated Essay Scoring: Positioning	196
9.4	Learner Trajectories in an Essay Space	212
9.5	Summary	221
	References	222
10	Evaluation	223
10.1	Insights from Earlier Prototypes	225
10.2	Verification	230
10.3	Validation	231
10.3.1	Scoring Accuracy	232
10.3.2	Structural Integrity of Spaces	234
10.3.3	Annotation Accuracy	238
10.3.4	Visual (in-)Accuracy	240
10.3.5	Performance Gains	241
10.4	Summary and Limitations	242
	References	243
11	Conclusion and Outlook	247
11.1	Achievement of Research Objectives	248
11.1.1	Objective 1: Represent Learning	249
11.1.2	Objective 2: Provide Instruments for Analysis	250
11.1.3	Objective 3: Re-represent to the User	252
11.1.4	Summary	256
11.2	Open Points for Future Research	258
11.3	Connections to Other Areas in TEL Research	260
11.4	Concluding Remarks	263
	References	263
	Erratum to: Learning Analytics in R with SNA, LSA, and MPIA	E1
	Annex A: Classes and Methods of the <i>mpia</i> Package	265

