Intelligent Systems Reference Library

Volume 159

The aim of this series is to publish a Reference Library, including novel advances and developments in all aspects of Intelligent Systems in an easily accessible and well structured form. The series includes reference works, handbooks, compendia, textbooks, well-structured monographs, dictionaries, and encyclopedias. It contains well integrated knowledge and current information in the field of Intelligent Systems. The series covers the theory, applications, and design methods of Intelligent Systems. Virtually all disciplines such as engineering, computer science, avionics, business, e-commerce, environment, healthcare, physics and life science are included. The list of topics spans all the areas of modern intelligent systems such as: Ambient intelligence, Computational intelligence, Social intelligence, Computational neuroscience, Artificial life, Virtual society, Cognitive systems, DNA and immunity-based systems, e-Learning and teaching, Human-centred computing and Machine ethics, Intelligent control, Intelligent data analysis, Knowledge-based paradigms, Knowledge management, Intelligent agents, Intelligent decision making, Intelligent network security, Interactive entertainment, Learning paradigms, Recommender systems, Robotics and Mechatronics including human-machine teaming, Self-organizing and adaptive systems, Soft computing including Neural systems, Fuzzy systems, Evolutionary computing and the Fusion of these paradigms, Perception and Vision, Web intelligence and Multimedia.

** Indexing: The books of this series are submitted to ISI Web of Science, SCOPUS, DBLP and Springerlink.

More information about this series at http://www.springer.com/series/8578

Anna Esposito · Antonietta M. Esposito ·
Lakhmi C. Jain
Editors

Innovations in Big Data Mining and Embedded Knowledge

Springer

Editors
Anna Esposito
Dipartimento di Psicologia and International Institute for Advanced Scientific Studies (IIASS)
Università degli Studi della Campania "Luigi Vanvitelli"
Caserta, Italy

Antonietta M. Esposito
Sezione di Napoli, Osservatorio Vesuviano
Istituto Nazionale di Geofisica e Vulcanologia
Napoli, Italy

Lakhmi C. Jain
University of Technology Sydney
Sydney, Australia

University of Canberra
Canberra, ACT, Australia

KES International
Shoreham-by-Sea, UK

Liverpool Hope University
Liverpool, UK

ISSN 1868-4394 ISSN 1868-4408 (electronic)
Intelligent Systems Reference Library
ISBN 978-3-030-15938-2 ISBN 978-3-030-15939-9 (eBook)
https://doi.org/10.1007/978-3-030-15939-9

Library of Congress Control Number: 2019935537

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

Preface

Data Mining or Knowledge Discovery in Databases (KDD) had received a great interest among researchers and practitioners due to its focus on processes transforming big amount of data into novel, valid, useful, and structured knowledge by detecting concealed patterns and relationships in data.

The book reports new directions in knowledge discovery using data mining and knowledge embedding through models. Contributors have reported a number of schemes to explain how data mining or embedding knowledge can be beneficial to social organizations, domestic spheres, and ICT market.

The concept of knowledge is broad and speculative and had produced epistemological debates in western philosophies. The intensified interest in knowledge management and data mining stems from the difficulty in identifying computational models able to approximate to a certain degree human behaviors and abilities in resolving organizational, social, and physical problems. Current ICT interfaces are not adequate to be introduced into the domestic spheres in order to support and simulate abilities of medical doctors, teachers, assistants, housekeepers, and so on. The domestic world has been denied to the machines because, differently from industrial contexts where abilities are routinely applied, the domestic world is continuously changing and unpredictable. The questions posited in this field are very challenging. Does knowledge locked in conventions, rules of conducts, common senses, ethics, emotions, laws, cultures, and experiences be mined from data? It will suffice for socially believable and emotionally behaving automatic systems, to rule complex interactions only through the mining of big amount of data?

The themes afforded by this book are multidisciplinary in nature and suggest that computational models able to approximate human behaviors and abilities in resolving organizational, social, and physical problems.

Innovations in Big Data Mining and Embedded Knowledge will prove useful to:

a. The academic research community
b. The ICT market
c. Ph.D. Students and Early Stage Researchers

d. Schools, Hospitals, Rehabilitation, and Assisted-Living Centers
e. Representatives from multimedia industries and standardization bodies

The editors would like to thank the contributors for their rigorous and invaluable scientific contributions, dedication, and priceless selection process. We wish to acknowledge the time and expertise of the reviewers for their contributions. Thanks are also due to the Springer-Verlag for their excellent support during the developmental phase of this research book.

Caserta, Italy — Anna Esposito
Napoli, Italy — Antonietta M. Esposito
Canberra, Australia — Lakhmi C. Jain

Sponsoring Organizations

- Università della Campania “Luigi Vanvitelli”, Dipartimento di Psicologia
- International Institute for Advanced Scientific Studies “E.R. Caianiello” (IIASS, www.iiassvietri.it/), Italy
- Società Italiana Reti Neuroniche (SIREN)

Contents

1 More Than Data Mining 1
Anna Esposito, Antonietta M. Esposito and Lakhmi C. Jain
1.1 Introduction 1
1.2 Discussion of Content 2
1.3 Conclusions 8
References 9

2 Designing a Recommender System for Touristic Activities in a Big Data as a Service Platform 13
Valerio Bellandi, Paolo Ceravolo, Ernesto Damiani and Eugenio Tacchini
2.1 Introduction 13
2.2 The Big Data Era 15
2.2.1 Dimensional Model of the Big Data 16
2.2.2 Tourism in the Future 16
2.3 Recommender System 17
2.3.1 Basic Techniques 18
2.3.2 Evaluation Metrics 19
2.3.3 Tourism Recommender Systems 19
2.4 Architecture 20
2.4.1 Lambda Architecture 20
2.4.2 A Multi-agent Tourism Recommender System 25
2.4.3 RS Architecture 28
2.4.4 Multi-agent Approach 30
2.4.5 Future Work 31
References 32

3 EML: A Scalable, Transparent Meta-Learning Paradigm for Big Data Applications 35
Uday Kamath, Carlotta Domeniconi, Amarda Shehu and Kenneth De Jong
3.1 Introduction 35
3.2 Background 37
3.2.1 Statistical Learning 37
3.2.2 Feature-Based Learning 38
3.2.3 EAs for Exploration of Vast, Complex Features Spaces 39
3.2.4 Big Data Machine Learning 39
3.2.5 Summary 41
3.3 Methodology 41
3.3.1 Feature Construction via EFG 42
3.3.2 The PSBML Algorithm 45
3.3.3 EML Framework 48
3.4 Experimental Validation 50
3.4.1 Evaluation Setup: Genome-Wide Sequence Classification 51
3.4.2 Experimental Setup: Hardware and Software 51
3.4.3 Datasets 52
3.4.4 Comparative Evaluation on the Splice Site Recognition Problem 52
3.5 Conclusion 54
References 54

4 Towards Addressing the Limitations of Educational Policy Based on International Large-Scale Assessment Data with Castoriadean Magmas 61
Evangelos Kapros
4.1 Introduction 61
4.1.1 Previous Work on ILSA and Culture 62
4.2 Data Analysis 63
4.3 Findings 67
4.4 Discussion 71
4.4.1 Culture, Education, and Assessment 71
4.4.2 Castoriadean Magmas 73
4.4.3 Principles for Designing Culturally-Appropriate EdAT 75
4.5 Conclusion 78
4.5.1 ILSA and Culture 78
4.5.2 Design for (Cultural) Magmas 79
4.5.3 Future Work 79
References 80

5 What Do Prospective Students Want? An Observational Study of Preferences About Subject of Study in Higher Education 83
Alessandro Vinciarelli, Walter Riviera, Francesca Dalmasso, Stefan Raue and Chamila Abeyratna
5.1 Introduction 83
5.2 Data Collection 85
5.3 Preference Analysis 86
5.4 Gender Segregation and Choice of Higher Education 89
5.5 Social Status and Access to Higher Education 92
5.6 Conclusions 95
References 96

6 Speech Pause Patterns in Collaborative Dialogs 99
Maria Koutsombogera and Carl Vogel
6.1 Introduction 99
6.1.1 Challenges in Multimodal Analytics 100
6.1.2 From Big Data to Sample Data 101
6.1.3 Multimodal Group Dialog: Turn-Taking and Pauses . . . 102
6.2 Speech Pauses: Background 103
6.3 Data Description 104
6.4 Data Analysis 105
6.4.1 Manual and Automatic Extraction of Pauses 106
6.4.2 Speech Pause Frequency 107
6.4.3 Silence Intervals and Dialog Topics 108
6.4.4 Local Context 110
6.4.5 Language Effects on Pause Duration 111
6.5 Conclusion 112
References 113

7 Discovering Knowledge Embedded in Bio-medical Databases: Experiences in Food Characterization and in Medical Process Mining 117
Giorgio Leonardi, Stefania Montani, Luigi Portinale, Silvana Quaglini and Manuel Striani
7.1 Introduction 117
7.2 Counterfeit and Adulteration Detection in Agrifood Products 119
7.2.1 Unsupervised Analysis: Clustering 121
7.2.2 Supervised Analysis: Classification 122
7.3 Medical Process Mining: A Semantic Approach 124
7.3.1 Knowledge-Based Trace Abstraction 125
7.3.2 Interface to the ProM Tool 128
7.3.3 Experimental Results 128
7.4 Related Works 131
7.5 Conclusions 133
References 134

8 A Paradigm for Democratizing Artificial Intelligence Research . . . 137
Erwan Moreau, Carl Vogel and Marguerite Barry
8.1 Introduction . . . 137
8.2 Motivations . . . 139
8.2.1 Familiarizing the General Public with AI . . . 139
8.2.2 Boosting NLP Research by Enhancing NLP Tools . . . 141
8.3 Approach . . . 146
8.3.1 Key Ideas . . . 146
8.3.2 Strategy . . . 152
8.4 Impact . . . 159
8.4.1 Long-Term Benefits of Involving Society with AI . . . 160
8.4.2 Indirect Benefits to Research and Industry . . . 162
8.5 Final Remarks . . . 162
References . . . 163

9 Big Data and Multimodal Communication: A Perspective View . . . 167
Costanza Navarretta and Lucretia Oemig
9.1 Introduction . . . 167
9.2 Studies on Multimodal Communication . . . 169
9.3 Big Data and Multimodality . . . 172
9.3.1 Multimodal Corpora . . . 172
9.3.2 Processing Big Data . . . 173
9.4 Using Big Data "in the Wild" Lawfully . . . 174
9.5 Discussion . . . 178
References . . . 180

10 A Web Application for Characterizing Spontaneous Emotions Using Long EEG Recording Sessions . . . 185
Giuseppe Placidi, Luigi Cinque and Matteo Polsinelli
10.1 Introduction . . . 186
10.2 Actors and Requirements . . . 187
10.3 System Design . . . 189
10.3.1 Signal Database . . . 190
10.3.2 Metadata Database . . . 192
10.3.3 Web Application . . . 193
10.3.4 Computational Module . . . 196
10.4 Conclusion . . . 197
References . . . 198

11 Anticipating the User: Acoustic Disposition Recognition in Intelligent Interactions . . . 203
Ronald Böck, Olga Egorow, Juliane Höbel-Müller, Alicia Flores Requardt, Ingo Siegert and Andreas Wendemuth
11.1 Introduction . . . 204
11.2 State of the Art of Corpora Collection and Recognition Techniques . . . 206

11.3 Generating Enriched, Naturalistic Interaction Data 209
11.3.1 The LAST MINUTE Corpus 209
11.3.2 The Integrated Health and Fitness Corpus 212
11.4 Using Socio-Demographic Information to Improve Disposition Modelling 213
11.4.1 The Influence of Age and Sex Information 213
11.4.2 The Influence of Personality Traits in Human-Computer Interaction 214
11.5 Speech Events Indicating User States and Changes 215
11.5.1 Special Speech Events in Interactions 215
11.5.2 Detection of User Dispositions Changes 218
11.6 Recent Approaches for Disposition Recognition from Speech 219
11.6.1 Corpus-Independent Analysis of Features 220
11.6.2 Transfer Learning and Active Learning in Disposition Recognition 221
11.6.3 Selected Neural Methods for Speech-Based Disposition Classification 221
11.7 Conclusion and Future Trends 223
References 225

12 Humans Inside: Cooperative Big Multimedia Data Mining 235
Shahin Amiriparian, Maximilian Schmitt, Simone Hantke, Vedhas Pandit and Björn Schuller
12.1 Introduction 235
12.2 System Structure 236
12.3 Complex Network Analyser 237
12.4 Intelligent Crowdsourcing for Data Annotation 238
12.4.1 Intelligent Audio Analysis 239
12.4.2 Active Learning Algorithms 239
12.4.3 Annotator Trustability Score and Annotation Reduction 240
12.5 Databases 242
12.5.1 Speech and Vocalisation Database 242
12.5.2 Vehicles Database 242
12.5.3 Ball Sports Database 242
12.6 Machine Learning Approaches 243
12.6.1 Recurrent Sequence-to-Sequence Autoencoders 244
12.6.2 Bag-of-Deep-Features 245
12.6.3 Crossmodal-Bag-of-Words 250
12.7 Results 251
12.8 Closing Remarks 253
References 254

13 Conversational Agents and Negative Lessons from Behaviourism 259
Milan Gnjatović
13.1 Introduction 259
13.2 Lack of External Criteria for Dialogue Corpora Production 261
13.3 Criticism of the Hidden Markov Model-Based Approach 263
13.4 Criticism of the Neural Network-Based Approach 267
13.5 Discussion and Conclusion 271
References 273

Author Index 275

About the Editors

Anna Esposito received her "Laurea Degree" summa cum laude in Information Technology and Computer Science from the Università di Salerno in 1989 with a thesis on: The Behavior and Learning of a Deterministic Neural Net (published on Complex System, 6(6), 507–517, 1992). She received her Ph.D. in Applied Mathematics and Computer Science from Università di Napoli "Federico II" in 1995. Her Ph.D. thesis on: Vowel Height and Consonantal Voicing Effects: Data from Italian (published on Phonetica, 59 (4), 197–231, 2002) was developed at Massachusetts Institute of Technology (MIT), Research Laboratory of Electronics (RLE), under the supervision of professor Kenneth N. Stevens.

She has been Post Doc at the International Institute for Advanced Scientific Studies (IIASS), and Assistant Professor at Università di Salerno (Italy), where she taught Cybernetics, Neural Networks, and Speech Processing (1996–2000). She had a position as Research Professor (2000–2002) at the Department of Computer Science and Engineering at WSU, Dayton, OH, USA, where she is currently Research Affiliate. She has been the Italian MCM of COST ACTIONS: COST 277: www.cost.esf.org/domains_actions/ict/Actions/277 (2001–2005); COST IC1002: www.cost.esf.org/domains_actions/ict/Actions/IC1002 (2010–2014); COST TIMELY: www.timely-cost.eu (2010–2014). She has been the proposer and chair of COST 2102: www.cost.esf.org/domains_actions/ict/Actions/2102 (2006–2010). Currently, she is the Italian MCM of COST CA15218: www.cost.eu/COST_Actions/ca/

CA15218?management, and the Italian MCM Substitute of COST Action IS1406: www.cost.eu/COST_Actions/isch/IS1406?management. Since 2006, she is a Member of the EuCognition, www.eucognition.org.

Her research interests are on contextual and interactional cross-modal analysis of typical and disordered (e.g., depressed or impaired individuals) human–machine interactional exchanges through speech, gestures, vocal, and facial emotional expressions in order to assess user's requirements and expectations to appropriately inform the implementation of emotional and social believable Human–Computer Interaction (HCI).

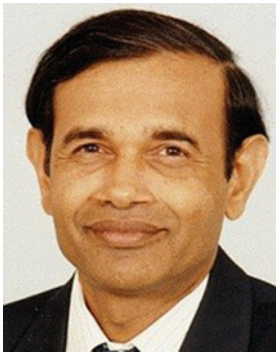
Antonietta M. Esposito was born and educated in Italy. She got the "Laurea" (Degree) from the University of Salerno (Italy) in 2001, with a Laurea thesis on the processing of speech signals (published in the IEEE Proceedings of the IEEE International Workshop on Circuits and Systems, R. L. Ewing et al. (Eds.), 2, 516–519, 2001).

From November 2001 to 2003, she participated at the Master in "Advanced Technologies of Information and Communication", supported by a 2-year fellowship of the "Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR)". From November 2003 to April 2005, her research activity was on speech recognition, at the CIRTE SpA (a telecommunication company in Salerno), and jointly on the analysis of seismic signals, at the Department of Physics at Salerno University and the IIASS.

She joined the Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Napoli Osservatorio Vesuviano, Napoli, Italy in May 2005 where she currently works as senior researcher for the monitoring and surveillance of Campanian Volcanoes (Vesuvio, Campi Flegrei and Ischia, Italy).

Her research interests are on seismic signal processing, speech, nonverbal communication features, neural networks, and mining of large amount of data with intelligent learning algorithms.

She is author of several contributes in peer-reviewed journals, books, and conferences and co-editors of three international books.



Dr. Lakhmi C. Jain, Ph.D., M.E., B.E.(Hons) Fellow (Engineers Australia) is with the University of Technology Sydney, Australia; University of Canberra, Australia and Liverpool Hope University, UK.

Professor Jain founded the KES International for providing a professional community, opportunities for publications, knowledge exchange, cooperation, and teaming. Involving around 5,000 researchers drawn from universities and companies world-wide, KES facilitates international cooperation and generate synergy in teaching and research. KES regularly provides networking opportunities for professional community through one of the largest conferences of its kind in the area of KES. www.kesinternational.org.

His interests focus on the artificial intelligence paradigms and their applications in complex systems, security, e-education, e-healthcare, unmanned air vehicles, and intelligent agents.

Keywords

Data mining
Embedded knowledge
Domestic spheres
Social agents and robots
Models
Assistive technologies
Complex autonomous systems
Socially believable interfaces
Computational complexity