# Intelligent Systems Reference Library

Volume 138

**Series editors**

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

Lakhmi C. Jain, University of Canberra, Canberra, Australia;
Bournemouth University, UK;
KES International, UK
e-mail: jainlc2002@yahoo.co.uk; jainlakhmi@gmail.com
URL: http://www.kesinternational.org/organisation.php

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.

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

Urszula Stańczyk · Beata Zielosko
Lakhmi C. Jain
Editors

# Advances in Feature Selection for Data and Pattern Recognition

Springer

*Editors*
Urszula Stańczyk
Silesian University of Technology
Gliwice
Poland

Beata Zielosko
University of Silesia in Katowice
Sosnowiec
Poland

Lakhmi C. Jain
University of Canberra
Canberra
Australia

and

Bournemouth University
Poole
UK

and

KES International
Shoreham-by-Sea
UK

ISSN 1868-4394 ISSN 1868-4408 (electronic)
Intelligent Systems Reference Library
ISBN 978-3-319-67587-9 ISBN 978-3-319-67588-6 (eBook)
https://doi.org/10.1007/978-3-319-67588-6

Library of Congress Control Number: 2017952019

© Springer International Publishing AG 2018
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.

Printed on acid-free paper

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

# Preface

This research book provides the reader with a selection of high-quality texts dedicated to the recent advances, developments, and research trends in the area of feature selection for data and pattern recognition.

The book can be treated as a continuation of our previous multi-authored volume as below:

**Urszula Stańczyk, Lakhmi C. Jain (Eds.)**
**Feature Selection for Data and Pattern Recognition**
**Studies in Computational Intelligence vol. 584**
**Springer-Verlag, Germany, 2015**

In particular, this second volume points to a number of advances topically subdivided into four parts:

- nature and representation of data;
- ranking and exploration of features;
- image, shape, motion, and audio detection and recognition;
- decision support systems.

The volume presents one introductory and 14 reviewed research papers, reflecting the work by 33 researchers from nine countries, namely Australia, Brazil, Canada, Germany, Hungary, Poland, Romania, Turkey, and USA.

Preparation and compilation of this monograph has been made possible by a number of people. Our warm thanks go to the commendable efforts of many institutions, teams, groups, and all individuals who have supported their laudable work. We wish to express our sincere gratitude to the contributing authors and all

who helped us in the review process of the submitted manuscripts. In addition, we extend an expression of gratitude to the members of staff at Springer, for their support in making this volume possible.

Gliwice, Poland — Urszula Stańczyk
Sosnowiec, Poland — Beata Zielosko
Canberra, Australia — Lakhmi C. Jain
August 2017

# Contents


**1 Advances in Feature Selection for Data and Pattern Recognition: An Introduction** . . . . . 1
Urszula Stańczyk, Beata Zielosko and Lakhmi C. Jain
1.1 Introduction . . . . . 1
1.2 Chapters of the Book . . . . . 3
1.3 Concluding Remarks . . . . . 7
References . . . . . 7

**Part I Nature and Representation of Data**

**2 Attribute Selection Based on Reduction of Numerical Attributes During Discretization** . . . . . 13
Jerzy W. Grzymała-Busse and Teresa Mroczek
2.1 Introduction . . . . . 13
2.2 Dominant Attribute Discretization . . . . . 14
2.3 Multiple Scanning Discretization . . . . . 19
2.4 Experiments . . . . . 19
2.5 Conclusions . . . . . 23
References . . . . . 23

**3 Improving Bagging Ensembles for Class Imbalanced Data by Active Learning** . . . . . 25
Jerzy Błaszczyński and Jerzy Stefanowski
3.1 Introduction . . . . . 25
3.2 Improving Classifiers Learned from Imbalanced Data . . . . . 28
3.2.1 Nature of Imbalanced Data . . . . . 28
3.2.2 Evaluation of Classifiers on Imbalanced Data . . . . . 29
3.2.3 Main Approaches to Improve Classifiers for Imbalanced Data . . . . . 31
3.3 Active Learning . . . . . 32

3.4 Ensembles Specialized for Imbalanced Data . . . 33
3.5 Active Selection of Examples in Under-Sampling Bagging . . . 35
3.6 Experimental Evaluation . . . 39
3.6.1 Experimental Setup . . . 39
3.6.2 Results of Experiments . . . 40
3.7 Conclusions . . . 48
References . . . 50

**4 Attribute-Based Decision Graphs and Their Roles in Machine Learning Related Tasks** . . . 53
João Roberto Bertini Junior and Maria do Carmo Nicoletti
4.1 Introduction . . . 54
4.2 The Attribute-Based Decision Graph Structure (AbDG) for Representing Training Data . . . 56
4.2.1 Constructing an AbDG . . . 56
4.2.2 Assigning Weights to Vertices and Edges . . . 58
4.2.3 Computational Complexity for Building an AbDG . . . 61
4.3 Using the AbDG Structure for Classification Tasks . . . 62
4.4 Using an AbDG for Classification Purposes - A Case Study . . . 63
4.5 Using the AbDG Structure for Imputation Tasks . . . 64
4.6 Searching for Refined AbDG Structures via Genetic Algorithms . . . 66
4.7 Conclusions . . . 69
References . . . 70

**5 Optimization of Decision Rules Relative to Length Based on Modified Dynamic Programming Approach** . . . 73
Beata Zielosko and Krzysztof Żabiński
5.1 Introduction . . . 73
5.2 Background . . . 75
5.3 Main Notions . . . 76
5.4 Modifed Algorithm for Directed Acyclic Graph Construction $\Delta^*_\beta(T)$ . . . 78
5.5 Procedure of Optimization Relative to Length . . . 81
5.6 Experimental Results . . . 83
5.6.1 Attributes' Values Selection and Size of the Graph . . . 83
5.6.2 Comparison of Length of $\beta$-Decision Rules . . . 85
5.6.3 Classifier Based on Rules Optimized Relative to Length . . . 90
5.7 Conclusions . . . 91
References . . . 91

**Part II Ranking and Exploration of Features**

**6 Generational Feature Elimination and Some Other Ranking Feature Selection Methods** . . . . . 97
Wiesław Paja, Krzysztof Pancerz and Piotr Grochowalski
6.1 Introduction . . . . . 98
6.2 Selected Methods and Algorithms . . . . . 99
6.2.1 Rough Set Based Feature Selection . . . . . 100
6.2.2 Random Forest Based Feature Selection . . . . . 102
6.2.3 Generational Feature Elimination Algorithm . . . . . 103
6.3 Feature Selection Experiments . . . . . 105
6.4 Results and Conclusions . . . . . 107
References . . . . . 110

**7 Ranking-Based Rule Classifier Optimisation** . . . . . 113
Urszula Stańczyk
7.1 Introduction . . . . . 113
7.2 Background . . . . . 114
7.2.1 Attribute Ranking . . . . . 115
7.2.2 Rule Classifiers . . . . . 116
7.2.3 Filtering Rules . . . . . 117
7.3 Research Framework . . . . . 118
7.3.1 Preparation of the Input Datasets . . . . . 118
7.3.2 Rankings of Features . . . . . 119
7.3.3 DRSA Decision Rules . . . . . 120
7.3.4 Weighting Rules . . . . . 122
7.4 Experimental Results . . . . . 123
7.4.1 Filtering Rules by Attributes . . . . . 123
7.4.2 Filtering Rules by Weights . . . . . 126
7.4.3 Summary of Results . . . . . 127
7.5 Conclusions . . . . . 129
References . . . . . 130

**8 Attribute Selection in a Dispersed Decision-Making System** . . . . . 133
Małgorzata Przybyła-Kasperek
8.1 Introduction . . . . . 133
8.2 Related Works . . . . . 134
8.3 Basics of the Rough Set Theory . . . . . 135
8.4 An Overview of Dispersed Systems . . . . . 136
8.4.1 Basic Definitions . . . . . 137
8.4.2 Static Structure . . . . . 138
8.4.3 Dynamic Structure with Disjoint Clusters . . . . . 139
8.4.4 Dynamic Structure with Inseparable Clusters . . . . . 141
8.4.5 Dynamic Structure with Negotiations . . . . . 141

8.5 Description of the Experiments . . . 143
8.5.1 Data Sets . . . 143
8.5.2 Attribute Selection . . . 144
8.5.3 Evaluation Measures and Parameters Optimization . . . 145
8.6 Experiments and Discussion . . . 148
8.6.1 Results for Static Structure . . . 148
8.6.2 Results for Dynamic Structure with Disjoint Clusters . . . 157
8.6.3 Results for Dynamic Structure with Inseparable Clusters . . . 157
8.6.4 Results for Dynamic Structure with Negotiations . . . 157
8.6.5 Comparison of All Methods . . . 158
8.6.6 Comparison for Data Sets . . . 160
8.7 Conclusion . . . 160
References . . . 161

**9 Feature Selection Approach for Rule-Based Knowledge Bases** . . . 163
Agnieszka Nowak-Brzezińska
9.1 Introduction . . . 163
9.2 Feature Selection Methods for Rule Mining Processes . . . 164
9.2.1 Significance of Rules Mining Process . . . 166
9.2.2 Feature Selection in the Rules and Their Clusters . . . 167
9.2.3 Related Works . . . 168
9.2.4 Clustering the Rules Based on the Similarity Approach . . . 168
9.2.5 Clustering the Rules . . . 170
9.2.6 Cluster's Representative . . . 172
9.3 Rough Set Approach in Creating Rules' Clusters Representatives . . . 173
9.3.1 Lower and Upper Approximation Approach . . . 174
9.3.2 KbExplorer - A Tool for Knowledge Base Exploration . . . 175
9.4 Experiments . . . 176
9.4.1 From Original Data to the Rules . . . 176
9.4.2 The Results of the Experiments . . . 176
9.4.3 The Summary of the Experiments . . . 180
9.5 Conclusion . . . 180
References . . . 181

**Part III Image, Shape, Motion, and Audio Detection and Recognition**

**10 Feature Selection with a Genetic Algorithm for Classification of Brain Imaging Data** . . . 185
Annamária Szenkovits, Regina Meszlényi, Krisztian Buza, Noémi Gaskó, Rodica Ioana Lung and Mihai Suciu

10.1 Introduction . . . 186
10.2 Materials and Methods . . . 189
10.2.1 Data and Preprocessing . . . 189
10.2.2 Least Absolute Shrinkage and Selection Operator . . . 189
10.2.3 Minimalist Genetic Algorithm for Feature Selection . . . 191
10.2.4 Experimental Set-Up . . . 194
10.3 Results . . . 195
10.4 Discussion . . . 198
10.5 Conclusions . . . 199
References . . . 200

**11 Shape Descriptions and Classes of Shapes. A Proximal Physical Geometry Approach** . . . 203
James Francis Peters and Sheela Ramanna
11.1 Introduction . . . 204
11.2 Introduction to Simplicial Complexes . . . 205
11.2.1 Examples: Detecting Shapes from Simplicial Complexes . . . 206
11.3 Preliminaries of Proximal Physical Geometry . . . 209
11.3.1 Image Object Shape Geometry . . . 209
11.3.2 Descriptions and Proximities . . . 210
11.3.3 Descriptive Proximities . . . 212
11.3.4 Edelsbrunner–Harer Nerve Simplexes . . . 215
11.4 Features of Image Object Shapes . . . 218
11.5 Concluding Remarks . . . 223
References . . . 223

**12 Comparison of Classification Methods for EEG Signals of Real and Imaginary Motion** . . . 227
Piotr Szczuko, Michał Lech and Andrzej Czyżewski
12.1 Introduction . . . 228
12.2 EEG Signal Parameterisation . . . 229
12.3 Data Classification Method . . . 232
12.4 Classification Results . . . 234
12.5 Conclusions . . . 236
References . . . 237

**13 Application of Tolerance Near Sets to Audio Signal Classification** . . . 241
Ashmeet Singh and Sheela Ramanna
13.1 Introduction . . . 241
13.2 Related Work . . . 243
13.3 Automatic Detection of Video Blocks . . . 245
13.4 Audio and Video Features . . . 246
13.4.1 Audio Features . . . 246
13.4.2 Video Features . . . 249

13.5 Theoretical Framework: Near Sets and Tolerance Near Sets . . . 249
13.5.1 Preliminaries . . . 250
13.5.2 Tolerance Near Sets . . . 252
13.6 Tolerance Class Learner - TCL . . . 254
13.6.1 Algorithms . . . 255
13.6.2 Phase II: Classification . . . 257
13.7 Experiments . . . 257
13.7.1 Speech and Non-speech Dataset . . . 257
13.7.2 Discussion . . . 259
13.7.3 Music and Music + Speech Dataset . . . 260
13.7.4 Detection of Commercial Blocks in News Data . . . 261
13.8 Conclusion and Future Work . . . 263
References . . . 263

**Part IV Decision Support Systems**

**14 Visual Analysis of Relevant Features in Customer Loyalty Improvement Recommendation** . . . 269
Katarzyna A. Tarnowska, Zbigniew W. Raś, Lynn Daniel and Doug Fowler
14.1 Introduction . . . 270
14.2 Related Applications . . . 270
14.3 Dataset and Application . . . 271
14.3.1 Dataset . . . 271
14.4 Decision Problem . . . 272
14.4.1 Attribute Analysis . . . 272
14.4.2 Attribute Reduction . . . 273
14.4.3 Customer Satisfaction Analysis and Recognition . . . 273
14.4.4 Providing Recommendations . . . 274
14.5 Proposed Approach . . . 274
14.5.1 Machine Learning Techniques . . . 274
14.5.2 Visualization Techniques . . . 282
14.6 Evaluation Results . . . 286
14.6.1 Single Client Data (Local) Analysis . . . 287
14.6.2 Global Customer Sentiment Analysis and Prediction . . . 288
14.6.3 Recommendations for a Single Client . . . 289
14.7 Conclusions . . . 292
References . . . 292

**15 Evolutionary and Aggressive Sampling for Pattern Revelation and Precognition in Building Energy Managing System with Nature-Based Methods for Energy Optimization** . . . 295
Jarosław Utracki and Mariusz Boryczka
15.1 Introduction . . . 296

15.2 Ant Colony Optimization . . . . . 296
15.2.1 Ant System . . . . . 296
15.2.2 Ant Colony Optimization Metaheuristics . . . . . 298
15.2.3 Ant Colony System . . . . . 299
15.2.4 Type of Issues That Are Able To Be Solved Using *ACO* . . . . . 302
15.3 Building's Installation as 3D Matrix-Like Grids . . . . . 303
15.4 Classical *ACO* Approach in Revealing a Pattern . . . . . 306
15.4.1 Capability of Revealing with *ACO* . . . . . 306
15.5 Aggressive *ACO* Sampling in Revealing a Pattern . . . . . 308
15.5.1 Ant Decomposition and Specialization . . . . . 308
15.5.2 Precognition Paths . . . . . 309
15.5.3 Ants Spreading . . . . . 311
15.6 Algorithm and Procedures . . . . . 312
15.7 Conclusion . . . . . 317
References . . . . . 318

**Index** . . . . . 321

# Editors and Contributors

## About the Editors

**Urszula Stańczyk** received the M.Sc. degree in Computer Science, and Ph.D. degree (with honours) in technical sciences with specialisation in informatics from the Silesian University of Technology (SUT), Gliwice, Poland. She is with the Institute of Informatics, SUT. From 2003 to 2010 Editor in Chief of the "Activity Report" for the Institute, Dr. Stańczyk is a member of KES International (www.kesinternational.org), ADAA Group (http://adaa.polsl.pl/), MIR Labs (http://www.mirlabs.org/), and International Rough Set Society (http://www.roughsets.org/), a member of Program Committees for several scientific conferences, and one of the key persons responsible for establishing a series of *International Conferences on Man-Machine Interactions (ICMMI)*. She is a member of the Editorial Board of Intelligent Decision Technologies: An International Journal (http://www.iospress.nl/journal/intelligent-decision-technologies/).

Her research interests include artificial intelligence, pattern recognition and classification, neural and rough processing, feature extraction and selection, induction of decision rules, rule quality measures, stylometric processing of texts, data mining. She co-edited conference proceedings and a multi-authored monograph on feature selection, authored and co-authored a two-volume monograph on synthesis and analysis of logic circuits, academic textbooks on arithmetic of digital systems, chapters, conference papers, and journal articles focused on applications of computational intelligence techniques to stylometry.

**Beata Zielosko** works as an Associate Professor in the Institute of Computer Science, University of Silesia (US), in Katowice. She earned the M.Sc. degree in 2002 and received a Ph.D. in Computer Science in 2008 from this University. In January 2017, she received a habilitation from the Institute of Computer Science of the Polish Academy of Sciences. From 2009 to 2016, Dr. Zielosko worked as an Assistant Professor in the Institute Computer Science, US. From 2011 to 2013, she worked as a Senior Research Scientist in the Mathematical and Computer Sciences and Engineering Division at King Abdullah University of Science and Technology, Saudi Arabia. Dr. Zielosko is a member of International Rough Set Society (www.roughsets.org), KES International (www.kesinternational.org) and a member of Program Committees of several international scientific conferences. She is also a co-editor of international conference proceedings IJCRS 2017.

Her main areas of research are data mining, optimization of decision and association rules, greedy algorithms for decision and association rule induction, feature selection, and machine learning. Dr. Zielosko is a co-author of three research monographs published by Springer.

**Dr. Lakhmi C. Jain** is with the Faculty of Education, Science, Technology and Mathematics at the University of Canberra, Australia, and Bournemouth University, UK. He is a Fellow of the Institution of Engineers Australia.

Professor Jain founded the KES International for providing a professional community the opportunities for publications, knowledge exchange, cooperation, and teaming. Involving around 5000 researchers drawn from universities and companies worldwide, KES facilitates international cooperation and generates 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.

## Contributors

**João Roberto Bertini Junior** School of Technology, University of Campinas, Limeira, SP, Brazil

**Jerzy Błaszczyński** Institute of Computing Science, Poznań University of Technology, Poznań, Poland

**Mariusz Boryczka** Institute of Computer Science, University of Silesia in Katowice, Sosnowiec, Poland

**Krisztian Buza** Knowledge Discovery and Machine Learning, Institute für Informatik III, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany

**Andrzej Czyżewski** Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Gdańsk, Poland

**Lynn Daniel** The Daniel Group, Charlotte, NC, USA

**Doug Fowler** The Daniel Group, Charlotte, NC, USA

**Noémi Gaskó** Centre for the Study of Complexity, Babeş-Bolyai University, Cluj-Napoca, Romania

**Piotr Grochowalski** Faculty of Mathematics and Natural Sciences, Department of Computer Science, University of Rzeszów, Rzeszów, Poland

**Jerzy W. Grzymała-Busse** Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA; Department of Expert Systems and Artificial Intelligence, University of Information Technology and Management, Rzeszów, Poland

**Lakhmi C. Jain** University of Canberra, Canberra, Australia

**Michał Lech** Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Gdańsk, Poland

**Rodica Ioana Lung** Centre for the Study of Complexity, Babeş-Bolyai University, Cluj-Napoca, Romania

**Regina Meszlényi** Department of Cognitive Science, Budapest University of Technology and Economics, Budapest, Hungary; Brain Imaging Centre, Research Centre for Natural Sciences, Hungarian Academy of Science, Budapest, Hungary

**Teresa Mroczek** Department of Expert Systems and Artificial Intelligence, University of Information Technology and Management, Rzeszów, Poland

**Maria do Carmo Nicoletti** Campo Limpo Paulista School, Campo Limpo Paulista, SP, Brazil; Computer Science Department, Federal University of São Carlos, São Carlos, SP, Brazil

**Agnieszka Nowak-Brzezińska** Institute of Computer Science, University of Silesia in Katowice, Sosnowiec, Poland

**Wiesław Paja** Faculty of Mathematics and Natural Sciences, Department of Computer Science, University of Rzeszów, Rzeszów, Poland

**Krzysztof Pancerz** Faculty of Mathematics and Natural Sciences, Department of Computer Science, University of Rzeszów, Rzeszów, Poland

**James Francis Peters** Computational Intelligence Laboratory, ECE Department, University of Manitoba, Winnipeg, MB, Canada; Faculty of Arts and Sciences, Department of Mathematics, Adiyaman University, Adiyaman, Turkey

**Małgorzata Przybyła-Kasperek** Institute of Computer Science, University of Silesia in Katowice, Sosnowiec, Poland

**Sheela Ramanna** Department of Applied Computer Science, University of Winnipeg, Winnipeg, MB, Canada; Computational Intelligence Laboratory, ECE Department, University of Manitoba, Winnipeg, MB, Canada

**Zbigniew W. Raś** University of North Carolina, Charlotte, NC, USA; Polish-Japanese Academy of IT, Warsaw University of Technology, Warsaw, Poland

**Ashmeet Singh** Department of Applied Computer Science, University of Winnipeg, Winnipeg, MB, Canada

**Urszula Stańczyk** Institute of Informatics, Silesian University of Technology, Gliwice, Poland

**Jerzy Stefanowski** Institute of Computing Science, Poznań University of Technology, Poznań, Poland

**Mihai Suciu** Centre for the Study of Complexity, Babeş-Bolyai University, Cluj-Napoca, Romania

**Piotr Szczuko** Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Gdańsk, Poland

**Annamária Szenkovits** Centre for the Study of Complexity, Babeş-Bolyai University, Cluj-Napoca, Romania

**Katarzyna A. Tarnowska** University of North Carolina, Charlotte, NC, USA

**Jarosław Utracki** Institute of Computer Science, University of Silesia in Katowice, Sosnowiec, Poland

**Beata Zielosko** Institute of Computer Science, University of Silesia in Katowice, Sosnowiec, Poland

**Krzysztof Żabiński** Institute of Computer Science, University of Silesia in Katowice, Sosnowiec, Poland