

# **Studies in Big Data**

Volume 8

*Series editor*

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland  
e-mail: [kacprzyk@ibspan.waw.pl](mailto:kacprzyk@ibspan.waw.pl)

For further volumes:  
<http://www.springer.com/series/11970>

### *About this Series*

The series “Studies in Big Data” (SBD) publishes new developments and advances in the various areas of Big Data-quickly and with a high quality. The intent is to cover the theory, research, development, and applications of Big Data, as embedded in the fields of engineering, computer science, physics, economics and life sciences. The books of the series refer to the analysis and understanding of large, complex, and/or distributed data sets generated from recent digital sources coming from sensors or other physical instruments as well as simulations, crowd sourcing, social networks or other internet transactions, such as emails or video click streams and other. The series contains monographs, lecture notes and edited volumes in Big Data spanning the areas of computational intelligence incl. neural networks, evolutionary computation, soft computing, fuzzy systems, as well as artificial intelligence, data mining, modern statistics and Operations research, as well as self-organizing systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

Witold Pedrycz · Shyi-Ming Chen  
Editors

# Information Granularity, Big Data, and Computational Intelligence

 Springer

*Editors*

Witold Pedrycz  
Department of Electrical and Computer  
Engineering  
University of Alberta  
Edmonton, AB  
Canada

Shyi-Ming Chen  
Department of Computer Science  
and Information Engineering  
National Taiwan University of Science  
and Technology  
Taipei  
Taiwan

ISSN 2197-6503

ISSN 2197-6511 (electronic)

ISBN 978-3-319-08253-0

ISBN 978-3-319-08254-7 (eBook)

DOI 10.1007/978-3-319-08254-7

Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014944331

© Springer International Publishing Switzerland 2015

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. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law. 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.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

The recent pursuits emerging in big data processing, interpretation, collection, and organization have emerged in numerous sectors including business, industry, and not-for-profit organizations. Data sets such as customer transactions for a mega-retailer, weather monitoring, intelligence gathering can quickly outpace the capacity of traditional techniques and tools of data analysis. We have been witnessing an emergence of new techniques and tools including NoSQL databases, MapReduce, Natural Language Processing, Machine Learning, visualization, acquisition, and serialization.

It becomes imperative to fully become aware what happens when big data grows up: how they are being applied and where they start playing a crucial role. We also need to become fully become aware of implications and requirements imposed on the existing techniques and various methods under development.

Soft Computing regarded as a plethora of technologies of fuzzy sets (or Granular Computing, in general), neurocomputing, and evolutionary optimization brings forward a number of unique features that might be instrumental to the development of concepts and algorithms to deal with big data. In particular, setting up a suitable and fully legitimate level of abstraction by forming semantically meaningful information granules is of paramount relevance. In light of their sheer volume, big data may call for distributed processing, where results of intensive data mining realized locally are afterwards reconciled leading to information granules of higher type. Neurocomputing operating at information granules leads to more tractable learning tasks. Evolutionary computing delivers an essential framework supporting global optimization.

In light of the inherent human-centric facet of Granular Computing the principles and practice of Computational Intelligence have been poised to play a vital role in the analysis, design, and interpretation of the architectures and functioning of mechanisms of big data.

Our ultimate objectives of this edited volume is to provide the reader with an updated, in-depth material on the emerging principles, conceptual underpinnings, algorithms, and practice of Computational Intelligence in the realization of concepts and implementation of big data architectures, analysis, and interpretation as well as data analytics.

An overall concise characterization of the objectives of the edited volume is expressed by highlighting several focal points:

- Systematic exposure of the concepts, design methodology, and detailed algorithms. In general, the volume adheres to the top-down strategy starting with the concepts and motivation and then proceeding with the detailed design that materializes in specific algorithms and representative applications.
- Individual chapters with clearly delineated agenda and well-defined focus and additional reading material available via carefully selected references.
- A wealth of carefully structured and organized illustrative material. The volume includes a series of brief illustrative numeric experiments, detailed schemes, and more advanced problems. They make the material more readable and appealing.
- Self-containment. Given the emerging character of the area of big data, our ultimate intent is to deliver a material that is self-contained and provides the reader with all necessary prerequisites and, if necessary, augments some parts with a step-by-step explanation of more advanced concepts supported by a significant amount of illustrative numeric material and some application scenarios to motivate the reader and make some abstract concepts more tangible.

The area of big data is highly diversified and this volume offers a quite representative view of the area. The contributions published here can be organized into three main parts. The first part, Fundamentals, which comprises chapters “[Nearest Neighbor Queries on Big Data](#)” to “[Building Fuzzy Robust Regression Model Based on Granularity and Possibility Distribution](#)” is focused on the methodological issues covering a broad spectrum of the approaches and detailed algorithmic pursuits including essential topics of forming cliques in big data, exploiting robust regression and its variants, constructing and optimizing rule-based models, Latent Semantic Indexing, information granulation, and Nearest Neighbor Querying. Part II entitled Architectures consisting of chapters “[The Role of Cloud Computing Architectures in Big Data](#)” to “[The Web Know ARR Framework: Orchestrating Computational Intelligence with Graph Databases](#)” is aimed at looking at the dedicated computing architectures such as cloud computing and the use of data storage techniques. Part III (case studies) includes chapters “[Customer Relationship Management and Big Data Mining](#)” to “[Application of Computational Intelligence on Analysis of Air Quality Monitoring Big Data](#)” which offer a suite of studies serving as a testimony to a wealth of promising applications including among others Customer Relationship management, market movements, weather forecasting, and air quality monitoring.

Given the theme of this project, this book is aimed at a broad audience of researchers and practitioners. Owing to the nature of the material being covered and the way it is organized, one can project with high confidence that it will appeal to the well-established communities including those active in various disciplines in which big data, their analysis, and optimization are of genuine relevance. Those involved in data mining, data analysis, management, various branches of engineering, and economics will benefit from the exposure to the subject matter.

Considering a way in which the edited volume is structured, this book could serve as a highly useful reference material for graduate students and senior undergraduate students in courses such as those on intelligent system, data mining, pattern recognition, decision-making, Internet engineering, Computational Intelligence, management, operations research, and knowledge-based systems.

We would like to take this opportunity to express our sincere thanks to the authors for sharing the results of their innovative research and delivering their insights into the area. The reviewers deserve our thanks for their constructive and timely input. We greatly appreciate a continuous support and encouragement coming from the Editor-in-Chief, Prof. Janusz Kacprzyk whose leadership and vision makes this book series a unique vehicle to disseminate the most recent, highly relevant and far-reaching publications in the domain of Computational Intelligence and its various applications.

We hope that the readers will find this volume of genuine interest and the research reported here will help foster further progress in research, education, and numerous practical endeavors.

Witold Pedrycz  
Shyi-Ming Chen

# Contents

## Part I Fundamentals

<b>Nearest Neighbor Queries on Big Data</b> . . . . .	3
Georgios Chatzimilioudis, Andreas Konstantinidis and Demetrios Zeinalipour-Yazti	
<b>Information Mining for Big Information</b> . . . . .	23
Yuichi Goto	
<b>Information Granules Problem: An Efficient Solution of Real-Time Fuzzy Regression Analysis</b> . . . . .	39
Azizul Azhar Ramli, Junzo Watada and Witold Pedrycz	
<b>How to Understand Connections Based on Big Data: From Cliques to Flexible Granules</b> . . . . .	63
Ali Jalal-Kamali, M. Shahriar Hossain and Vladik Kreinovich	
<b>Graph-Based Framework for Evaluating the Feasibility of Transition to Maintainomics</b> . . . . .	89
Bo Xing	
<b>Incrementally Mining Frequent Patterns from Large Database</b> . . . . .	121
Yue-Shi Lee and Show-Jane Yen	
<b>Improved Latent Semantic Indexing-Based Data Mining Methods and an Application to Big Data Analysis of CRM</b> . . . . .	141
Jianxiong Yang and Junzo Watada	
<b>The Property of Different Granule and Granular Methods Based on Quotient Space</b> . . . . .	171
Yan-ping Zhang, Ling Zhang and Chenchu Xu	
<b>Towards an Optimal Task-Driven Information Granulation</b> . . . . .	191
Alexander Ryjov	



**Unified Framework for Construction of Rule Based Classification Systems . . . . . 209**  
 Han Liu, Alexander Gegov and Frederic Stahl

**Multi-granular Evaluation Model Through Fuzzy Random Regression to Improve Information Granularity. . . . . 231**  
 Nureize Arbaiy and Junzo Watada

**Building Fuzzy Robust Regression Model Based on Granularity and Possibility Distribution . . . . . 247**  
 Yoshiyuki Yabuuchi and Junzo Watada

**Part II Architectures**

**The Role of Cloud Computing Architecture in Big Data. . . . . 275**  
 Mehdi Bahrami and Mukesh Singhal

**Big Data Storage Techniques for Spatial Databases: Implications of Big Data Architecture on Spatial Query Processing . . . . . 297**  
 Roger Frye and Mark McKenney

**The Web KnowARR Framework: Orchestrating Computational Intelligence with Graph Databases. . . . . 325**  
 Edy Portmann and Patrick Kaltenrieder

**Part III Case Studies**

**Customer Relationship Management and Big Data Mining. . . . . 349**  
 Yi Hui Liang

**Performance Competition for ISCIFCM and DPEI Models Under Uncontrolled Circumstances . . . . . 361**  
 Jui Fang Chang

**Rough Set Model Based Knowledge Acquisition of Market Movements from Economic Data . . . . . 375**  
 Yoshiyuki Matsumoto and Junzo Watada

**Deep Neural Network Modeling for Big Data Weather Forecasting. . . 389**  
 James N. K. Liu, Yanxing Hu, Yulin He, Pak Wai Chan and Lucas Lai

**Current Knowledge and Future Challenge for Visibility  
Forecasting by Computational Intelligence . . . . . 409**  
Wang-Kun Chen and Chung-Shin Yuan

**Application of Computational Intelligence on Analysis  
of Air Quality Monitoring Big Data . . . . . 427**  
Tzu-Yi Pai, Moo-Been Chang and Shyh-Wei Chen

**Index . . . . . 443**