Studies in Big Data

Volume 19

Series editor

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

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.

More information about this series at http://www.springer.com/series/11970

Dominik Ryżko · Piotr Gawrysiak
Marzena Kryszkiewicz · Henryk Rybiński
Editors

Machine Intelligence and Big Data in Industry

Springer

Editors
Dominik Ryżko
Institute of Computer Science
Warsaw University of Technology
Warsaw
Poland

Marzena Kryszkiewicz
Institute of Computer Science
Warsaw University of Technology
Warsaw
Poland

Piotr Gawrysiak
Institute of Computer Science
Warsaw University of Technology
Warsaw
Poland

Henryk Rybiński
Institute of Computer Science
Warsaw University of Technology
Warsaw
Poland

ISSN 2197-6503 ISSN 2197-6511 (electronic)
Studies in Big Data
ISBN 978-3-319-30314-7 ISBN 978-3-319-30315-4 (eBook)
DOI 10.1007/978-3-319-30315-4

Library of Congress Control Number: 2016932358

Printed on acid-free paper

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

Preface

This book presents valuable contributions devoted to practical and, in many cases, industrial applications of Machine Intelligence and Big Data in various branches of the industry. All the contributions are extended versions of presentations delivered at the Industrial Session of the 6th International Conference on Pattern Recognition and Machine Intelligence (PREMI 2015) in Warsaw, Poland, which passed through a rigorous reviewing process. Each paper was reviewed by at least two referees.

Part I is focused on practical applications of text processing. This part demonstrates the usefulness of text mining approaches in solving practical problems. In particular, Sobkowicz addressed the problem of automatic sentiment analysis for Polish language. Kowalski presented a process of translating legal English and Polish phrases, being a part of a bilingual university repository. Roziewski et al. studied creation of *n*-gram collection from a large-scale corpus of Polish Internet based on Common Crawl Corpus. Kozłowski applied clustering of documents containing Polish national qualifications frameworks in order to analyze study fields. Various approaches to the semantic textual similarity are compared in the work by Kazuła and Kozłowski.

Part II is devoted to data mining. El-Baz et al. proposed a solution of the problem of identification of diabetes disease by means of committees of neural network-based classifiers. Sharif et al. proposed enzyme function classification based on Borda count ranking aggregation method. The problem of mining of frequent action rules is addressed by Dardzinska and Romaniuk.

Text and multimedia processing is the subject of Part III. Protaziuk et al. proposed an automatic machine translation method for translating multi-word labels from lexical layers of domain ontologies. In the area of automated speech recognition, Madhavi et al. addressed vocal tract length normalization using different warping functions for template matching. A comparative study on music genre classification algorithms was presented by Stokowiec.

Issues of software platforms are studied in Part IV. Blachnik and Kordos described a RapidMiner Library for information selection and data compression.

Wróblewska et al. showed how to cluster offers in an e-commerce marketplace in order to improve performance of recommendations and other services. An application of machine learning algorithms to Bitcoin automated trading is described by Żbikowski.

Part V combines papers on complex systems, the Internet of Things, and agent systems. Kopczynski et al. presented a design for hardware cuts generating module for Field Programmable Gate Arrays (FPGAs). A big data solution for smart grids and smart meters was presented by Konopko. Weclawski and Jankowski presented an intelligent system of limited resource allocation for large-scale agent systems. Yadav et al. studied the problem of finding logical patterns in multi-sensor data from the industrial Internet.

We thank all the authors for their contributions to the book and we express our appreciation for the work of the reviewers. We express our gratitude to the industrial partners: mBank, Allegro, and Samsung for their financial support to the PReMI 2015 conference and to this publication.

November 2015

Dominik Ryżko
Piotr Gawrysiak
Marzena Kryszkiewicz
Henryk Rybiński

Contents