

SpringerBriefs in Computer Science

Series editors

Stan Zdonik, Brown University, Providence, RI, USA

Shashi Shekhar, University of Minnesota, Minneapolis, MN, USA

Xindong Wu, University of Vermont, Burlington, VT, USA

Lakhmi C. Jain, University of South Australia, Adelaide, SA, Australia

David Padua, University of Illinois Urbana-Champaign, Urbana, IL, USA

Xuemin Sherman Shen, University of Waterloo, Waterloo, ON, Canada

Borko Furht, Florida Atlantic University, Boca Raton, FL, USA

V. S. Subrahmanian, Department of Computer Science, University of Maryland, College Park, MD, USA

Martial Hebert, Carnegie Mellon University, Pittsburgh, PA, USA

Katsushi Ikeuchi, Meguro-ku, University of Tokyo, Tokyo, Japan

Bruno Siciliano, Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione, Università di Napoli Federico II, Napoli, Italy

Sushil Jajodia, George Mason University, Fairfax, VA, USA

Newton Lee, Institute for Education, Research and Scholarships, Los Angeles, CA, USA

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic.

Typical topics might include:

- A timely report of state-of-the art analytical techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- An in-depth case study or clinical example
- A presentation of core concepts that students must understand in order to make independent contributions

Briefs allow authors to present their ideas and readers to absorb them with minimal time investment. Briefs will be published as part of Springer's eBook collection, with millions of users worldwide. In addition, Briefs will be available for individual print and electronic purchase. Briefs are characterized by fast, global electronic dissemination, standard publishing contracts, easy-to-use manuscript preparation and formatting guidelines, and expedited production schedules. We aim for publication 8–12 weeks after acceptance. Both solicited and unsolicited manuscripts are considered for publication in this series.

More information about this series at <http://www.springer.com/series/10028>

Murad Khan · Bilal Jan · Haleem Farman

Deep Learning: Convergence to Big Data Analytics

 Springer

Murad Khan
Department of Computer Science
Sarhad University of Science
and Information Technology
Peshawar, Pakistan

Haleem Farman
Department of Computer Science
Islamia College Peshawar
Peshawar, Pakistan

Bilal Jan
Department of Computer Science
Fata University
FR Kohat, Pakistan

ISSN 2191-5768 ISSN 2191-5776 (electronic)
SpringerBriefs in Computer Science
ISBN 978-981-13-3458-0 ISBN 978-981-13-3459-7 (eBook)
<https://doi.org/10.1007/978-981-13-3459-7>

Library of Congress Control Number: 2018962781

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2019

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.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

About This Book

This book presents deep learning techniques, concepts, and algorithms to classify and analyze big data. The big data analytics using traditional techniques have various challenges and issues such as high speed analytics, accuracy, and efficient processing of big data in real time. In addition, the Internet of things is progressively increasing in various fields such as smart cities, smart homes, and e-health. This enormous number of connected devices generate a huge amount of data every day, and hence, we need sophisticated algorithms to deal, organize, and classify such huge amount of data in less processing time and space. Similarly, the existing techniques and algorithms for deep learning in big data field have several advantages because of the two main branches of deep learning, i.e., convolution and deep belief networks. This book gives an insight into all those techniques and applications based on these two types of deep learning. Also, the book helps the students, researchers, and newcomers in understanding big data analytics based on deep learning approaches. Further, the book gives an introductory-level understanding to the new programming languages and tools used to analyze big data specifically in real time such as Hadoop, Spark, and GraphX. Various machine learning techniques are discussed in concatenation with the deep learning paradigm to support high-end data processing, data classifications, and real-time data processing issues.

The classification and presentation of the book are kept quite simple to help the readers and students understand the basic concepts of various deep learning paradigms and frameworks. The book mainly focuses on theory instead of presenting the mathematical background of the deep learning concepts, because the theory behind these techniques still needs further improvements and refining to enable the students and practitioners to understand and develop their mathematical models. The book consists of five chapters beginning with an introductory explanation of the big data alongside deep learning techniques, followed by integration of big data and deep learning techniques, and finally the future directions in big data analytics using deep learning techniques.

Aims and Scope of the Book

Deep learning techniques are widely adopted in many fields of computer science and engineering for data analysis. The introduction of big data analytics has several challenges such as processing of huge amount of data efficiently and in less time. However, the conventional methods and database management systems fail to offer such functionalities of processing huge data in less and finite time. Therefore, the researchers try various techniques and algorithms from various fields to come up with a system which can process data as quick as possible. The Hadoop ecosystem is developed for such problems with a map and reduce concept. The Hadoop ecosystem can process a huge amount of data; however, it cannot offer the functionality of processing data in real time. Therefore, sophisticated techniques are need of the day to deal with such huge amount of data generated in real time with high velocity, volume, and varsity (i.e., 3V concept). In addition, the big data is always generated from heterogeneous sources; therefore, learning functionalities with conventional SQL query techniques always require high-end nodes. However, installing high-end node everywhere is a difficult and impossible task. Thus, incorporated learning technique without installing high-end node can be possible using deep learning techniques. The aim of this book is to bring the academia, research community, and industry at one platform to design and suggest deep learning techniques for big data analytics. The classification of the book is further divided into four main categories: (1) novel and sophisticated real-time deep learning analytics for cross-layer and inter-domain services, (2) state-of-the-art theories and practice for application scenarios from heterogeneous and cross-domain big data analytics, (3) survey and tutorials from world's top research community and laborites working in big data analytics, and (4) building of benchmark and test beds for analyzing big data using deep learning techniques. Further, the scope of the book consists of the following topics:

- Feature and data-pattern learning using deep learning techniques
- Large-scale data processing-based learning techniques
- Architectural design of deep learning algorithms and techniques
- Pattern recognition via deep learning techniques

- Innovative scientific methodologies for big data processing in real time
- Big data processing for urban planning, management, and sustainability
- Sustainability of business models for smart home, smart cities, e-health care
- The analysis and processing of big data generated from Internet of things
- Information privacy and data leakage prevention using big data analytics
- Novel and efficient machine and data mining on big data
- Big data sets and benchmarks generation and processing for big data analytics based on deep learning techniques.

Contents

1	Introduction	1
	Bilal Jan, Haleem Farman and Murad Khan	
1.1	Machine Learning	1
1.1.1	Supervised Learning	2
1.1.2	Unsupervised Learning	2
1.2	Deep Learning	3
1.3	Conventional Data Processing Techniques	4
1.4	Data Mining Techniques: A Big Data Analysis Approach	5
1.5	Big Data Analytics	7
1.6	Deep Learning in Big Data Analytics	10
	References	11
2	Big Data Analytics	13
	Bhagya Nathali Silva, Muhammad Diyan and Kijun Han	
2.1	Overview	14
2.2	Characteristics of Big Data	17
2.3	Big Data Processing	18
2.4	Data Analysis Problems in Big Data	21
2.5	Applications of Big Data	23
2.5.1	Healthcare	23
2.5.2	Manufacturing	23
2.5.3	Government	24
2.5.4	Internet of Things	24
2.6	Data Types of Big Data	24
2.7	Big Data Tools	26
2.8	Opportunities and Challenges	27
	References	29

3	Deep Learning Methods and Applications	31
	Jamil Ahmad, Haleem Farman and Zahoor Jan	
3.1	Background	32
3.2	Categorization of Deep Learning Networks	32
3.3	Deep Networks for Supervised Learning	33
3.4	Deep Networks for Unsupervised Learning	35
3.5	Hybrid Approach	36
3.6	Transfer Learning Techniques	36
	3.6.1 Homogenous Transfer Learning	37
	3.6.2 Heterogeneous Transfer Learning	37
3.7	Applications of Deep Learning	37
	3.7.1 Computer Vision	38
	3.7.2 Information Retrieval	38
	3.7.3 Natural Language Processing	39
	3.7.4 Multitask Learning	40
	References	41
4	Integration of Big Data and Deep Learning	43
	Muhammad Talha, Shaukat Ali, Sajid Shah, Fiaz Gul Khan and Javed Iqbal	
4.1	Machine Learning in Big Data Analytics	43
	4.1.1 Machine Learning and Big Data Applications	44
4.2	Efficient Deep Learning Algorithms in Big Data Analytics	44
4.3	From Machine to Deep Learning: A Comparative Approach	46
	4.3.1 Performance on Data Size	46
	4.3.2 Hardware Requirements	47
	4.3.3 Feature Selection	47
	4.3.4 Problem-Solving Approach	47
	4.3.5 Execution Time	47
4.4	Applications of Deep and Transfer Learning in Big Data	47
	4.4.1 Healthcare	48
	4.4.2 Finance	49
4.5	Deep Learning Challenges in Big Data	49
	4.5.1 Internet of Things (IoT) Data	49
	4.5.2 Enterprise Data	50
	4.5.3 Medical and Biomedical Data	50
	References	51
5	Future of Big Data and Deep Learning for Wireless Body Area Networks	53
	Fasee Ullah, Ihtesham Ul Islam, Abdul Hanan Abdullah and Atif Khan	
5.1	Introduction	54
5.2	Feed-Forward Network Model	56
	5.2.1 Deep Learning Frameworks	58

- 5.3 Future of Deep Learning 59
- 5.4 Introduction to Wireless Body Area Networks 59
- 5.5 Applications of Wireless Body Area Networks 62
 - 5.5.1 Future Applications of Wireless Body Area Networks 63
 - 5.5.2 Use of Biomedical Sensors in Wireless Body Area Networks 64
- 5.6 Existing Challenges in Wireless Body Area Networks 65
 - 5.6.1 Routing Protocols 65
- 5.7 MAC Protocols 67
 - 5.7.1 Superframe Structure of IEEE 802.15.4 67
 - 5.7.2 Superframe Structure of IEEE 802.15.6 69
- 5.8 Introduction to Big Data 69
- 5.9 Applications of Big Data in WBAN 70
 - 5.9.1 Monitoring of Vital Signs and Analysis 71
 - 5.9.2 Early Detection of Abnormal Conditions of Patient 71
 - 5.9.3 Daily Basis Activity Monitoring of a Patient Using BMSs 72
- 5.10 Open Issues of WBAN 72
 - 5.10.1 Resource-Constraint Architecture of BMS 72
 - 5.10.2 Hotspot Paths 72
 - 5.10.3 QoS in WBAN 73
 - 5.10.4 Path Loss in WBAN 73
 - 5.10.5 Data Protection in WBAN 73
 - 5.10.6 Step-Down in Energy Consumption 73
 - 5.10.7 Channel Access Allocation and Its Complexity 73
 - 5.10.8 Permission- and Preemption-Based Channel Assignment 73
- 5.11 Open Issues of Big Data 74
 - 5.11.1 Varieties of Data 74
 - 5.11.2 Increased Amount of Data Storage 74
 - 5.11.3 Integration of Data from Different Sources 74
 - 5.11.4 Allocation of Channel, Processing, and Management of Data 74
 - 5.11.5 Cost-Effective Business Model 74
 - 5.11.6 Delay-Aware Models for Quick Solutions 75
 - 5.11.7 Automation in Allocation of Services 75
- References 75
- Index 79**

About the Authors

Murad Khan received his BS in computer science from the University of Peshawar, Pakistan, in 2008. He completed his Ph.D. in computer science and engineering from School of Computer Science and Engineering in Kyungpook National University, Daegu, Korea. He has published over 60 international conference and journal papers along with two chapters in Springer and CRC Press. He also served as a TPC member in world-reputed conferences and as a reviewer in numerous journals such as Future Generation Systems (Elsevier) and IEEE Access. In 2016, he was awarded Qualcomm Innovation Award at Kyungpook National University for designing a smart home control system. He was also awarded Bronze Medal in ACM SAC 2015, Salamanca, Spain, on his distinguished work in multi-criteria-based handover techniques. He is a member of various communities such as ACM and IEEE, and CRC Press. His areas of expertise include ad hoc and wireless networks, architecture designing for Internet of things, communication protocols designing for smart cities and homes, big data analytics, etc.

Bilal Jan received his MS and Ph.D. from the Department of Control and Computer Engineering (DAUIN), Politecnico di Torino, Italy, in 2010 and 2015, respectively. He has published several papers in reputed journals and conferences. He is currently working as Assistant Professor and Head of the Department of Computer Science, FATA University, Darra Adam Khel, FR Kohat, Pakistan. He is a reviewer for numerous leading journals. His research interests include general-purpose programming in GPUs, high-performance computing, wireless sensor networks, Internet of things (IoT), deep learning, and big data.

Haleem Farman received his MS from the International Islamic University, Islamabad, Pakistan, in 2008. He is currently pursuing his Ph.D. in Computer Science in the Department of Computer Science, University of Peshawar, Pakistan, and working as Lecturer in the Department of Computer Science, Islamia College

Peshawar, Pakistan. He has authored/co-authored more than 20 research papers in respected journals and conferences. In addition, he serves as an invited reviewer for several journals, such as Elsevier Sustainable Cities and Society. His fields of interest include wireless sensor networks, Internet of things, big data analytics, privacy, optimization techniques, and quality of service issues in wireless networks.

List of Figures

Fig. 1.1	Three layers of big data	9
Fig. 2.1	Data growth evolution from 1970 to 2020 (predicted)	15
Fig. 2.2	Stages of big data processing to extract valuable knowledge from big data	16
Fig. 2.3	Laney’s 3V model based on fundamental characteristics of BD	19
Fig. 2.4	Stages of big data life cycle with corresponding technologies	20
Fig. 3.1	A fully connected feed-forward deep neural network	33
Fig. 3.2	AlexNet convolutional neural network architecture	34
Fig. 3.3	Architecture of a single recurrent unit in RNN	34
Fig. 3.4	Architecture of a simple autoencoder showing the encoder and decoder parts	35
Fig. 3.5	Typical framework of an IR system	39
Fig. 3.6	MTL with hard parameter sharing	40
Fig. 3.7	MTL with soft parameter sharing	41
Fig. 4.1	Challenges of big data present in the form of four Vs	51
Fig. 5.1	Performance relationship of deep networks with the increase in amount of data original slide by Professor Andrew Ng	55
Fig. 5.2	a (Di Cataldo et al. 2014; Ul-Islam 2014) Traditional machine learning with feature engineering by a domain expert versus end-to-end deep convolution learning b Deep CNN-end to end learning.	56
Fig. 5.3	a A feed-forward network (or MLP) with input layer, hidden layers, and output layer and b the computation at a single neuron with input and output activation	57
Fig. 5.4	Different activation functions are shown in a sigmoid, b tanh, c ReLU, and d leaky ReLU	57
Fig. 5.5	WHO reports on average death rates in million	60
Fig. 5.6	Typical deployment of BMSs to monitor vital signs of patient	61

Fig. 5.7 Typical deployment of BMSs 61

Fig. 5.8 Applications of wireless body area network (Applications of WBAN 2018). 62

Fig. 5.9 Future applications of wireless body area network (Future Applications of WBAN 2018) 63

Fig. 5.10 Classification of routing protocols in WBAN 66

Fig. 5.11 Superframe structure of IEEE 802.15.4 (re-drawn) 68

Fig. 5.12 Framework for WBAN big data (Du et al. 2015) 70

Fig. 5.13 WBAN with cloudlet-based data collection system (Quwaider 2014). 71