

Founding Editors

Gerhard Goos

Karlsruhe Institute of Technology, Karlsruhe, Germany

Juris Hartmanis

Cornell University, Ithaca, NY, USA

Editorial Board Members

Elisa Bertino

Purdue University, West Lafayette, IN, USA

Wen Gao

Peking University, Beijing, China

Bernhard Steffen 

TU Dortmund University, Dortmund, Germany

Gerhard Woeginger 

RWTH Aachen, Aachen, Germany

Moti Yung

Columbia University, New York, NY, USA

More information about this subseries at <http://www.springer.com/series/7409>

Ladjel Bellatreche · Vikram Goyal ·
Hamido Fujita · Anirban Mondal ·
P. Krishna Reddy (Eds.)

Big Data Analytics

8th International Conference, BDA 2020
Sonepat, India, December 15–18, 2020
Proceedings

Editors

Ladjet Bellatreche
ISAE-ENSMA
Chasseneuil, France

Hamido Fujita
Iwate Prefectural University
Takizawa, Japan

P. Krishna Reddy
IIIT Hyderabad
Hyderabad, India

Vikram Goyal
Indraprastha Institute of Information
Technology
New Delhi, India

Anirban Mondal
Ashoka University
Sonapat, India

ISSN 0302-9743

ISSN 1611-3349 (electronic)

Lecture Notes in Computer Science

ISBN 978-3-030-66664-4

ISBN 978-3-030-66665-1 (eBook)

<https://doi.org/10.1007/978-3-030-66665-1>

LNCS Sublibrary: SL3 – Information Systems and Applications, incl. Internet/Web, and HCI

© Springer Nature Switzerland AG 2020

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, expressed 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 Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The amount of data stored in computer systems has greatly increased in recent years due in part to advances in networking technologies, storage systems, the adoption of mobile and cloud computing and the wide deployment of sensors for data collection. To make sense of this big data to support decision-making, the field of Big Data Analytics has emerged as a key research and application area for industry and other organizations. Numerous applications of big data analytics are found in several important and diverse fields such as e-commerce, finance, healthcare, education, e-governance, media and entertainment, security and surveillance, smart cities, telecommunications, agriculture, astronomy, and transportation.

Analysis of big data raises several challenges such as how to handle massive volumes of data, process data in real-time, and deal with complex, uncertain, heterogeneous and streaming data, which are often stored in multiple remote repositories. To address these challenges, innovative big data analysis solutions must be designed by drawing expertise from recent advances in several fields such as big data processing, data mining, database systems, statistics, machine learning and artificial intelligence. There is also an important need to build data analysis systems for emerging applications such as vehicular networks and social media analysis, and to facilitate the deployment of big data analysis techniques.

The Eighth International Conference on Big Data Analytics (BDA) was held from 15th December to 18th December 2020 at Ashoka University, Sonapat, Haryana, India. This proceedings book includes 15 peer-reviewed research papers and contributions by keynote speakers, invited speakers and tutorial speakers. This year's program covered a wide range of topics related to big data analytics on themes such as: big data analytics, gesture detection, networking, social media, search, information extraction, image processing and analysis, spatial, text, mobile and graph data analysis, machine learning, and healthcare.

It is expected that research papers, keynote speeches, invited talks and tutorials presented at the conference will encourage research on big data analytics and stimulate the development of innovative solutions and their adoption in industry.

The conference received 48 submissions. The Program Committee (PC) consisted of researchers from both academia and industry from 13 countries or territories, namely Australia, Canada, China, Estonia, France, India, Japan, New Caledonia, New Zealand, Norway, Spain, Taiwan and the USA. Each submission was reviewed by at least two, and most by three Program Committee members, and was discussed by PC chairs before taking the decision. Based on the above review process, the Program Committee selected 15 full papers. The overall acceptance rate was about 31%.

We would like to extend our sincere thanks to the members of the Program Committee and external reviewers for their time, energy and expertise in providing support to BDA 2020.

Additionally, we would like to thank all the authors who considered BDA 2020 as the forum to publish their research contributions. The Steering Committee and the Organizing Committee deserve praise for the support they provided. A number of individuals contributed to the success of the conference. We thank Prof. H. V. Jagadish, Prof. M. Kitsuregawa, Prof. Divyakant Agrawal, Prof. Amit Sheth and Prof. Ouri Wolfson for their insightful keynote talks. We also thank all the invited speakers and tutorial speakers. We would like to thank the sponsoring organizations including National Institute of Technology Delhi (NITD), India; Indraprastha Institute of Information Technology Delhi (IIITD), India; International Institute of Information Technology Hyderabad (IIITH), India; The University of Aizu, Japan; University of Delhi, India; Ahmedabad University, India and the Department of Computer Science at Ashoka University, as they deserve praise for the support they provided.

The conference received invaluable support from the management of Ashoka University in hosting and organizing the conference. Moreover, thanks are also extended to the faculty, staff members and student volunteers of the Department of Computer Science at Ashoka University for their constant cooperation and support.

Ladjet Bellatreche
Vikram Goyal
Hamido Fujita
Anirban Mondal
P. Krishna Reddy

Organization

BDA 2020 was organized by Ashoka University, Sonapat, Haryana 131029, India.

Honorary Chairs

Malabika Sarkar (Vice-chancellor)	Ashoka University, India
Praveen Kumar (Director)	NIT Delhi, India

General Chair

Anirban Mondal	Ashoka University, India
----------------	--------------------------

Steering Committee Chair

P. Krishna Reddy	IIIT Hyderabad, India
------------------	-----------------------

Steering Committee

S. K. Gupta	IIT Delhi, India
Srinath Srinivasa	IIIT Bangalore, India
Krithi Ramamritham	IIT Bombay, India
Sanjay Kumar Madria	Missouri University of Science and Technology, USA
Masaru Kitsuregawa	University of Tokyo, Japan
Raj K. Bhatnagar	University of Cincinnati, USA
Vasudha Bhatnagar	University of Delhi, India
Mukesh Mohania	IBM Research, Australia
H. V. Jagadish	University of Michigan, USA
Ramesh Kumar Agrawal	Jawaharlal Nehru University, India
Divyakant Agrawal	University of California, Santa Barbara, USA
Arun Agarwal	University of Hyderabad, India
Subhash Bhalla	The University of Aizu, Japan
Jaideep Srivastava	University of Minnesota, USA
Anirban Mondal	Ashoka University, India
Sharma Chakravorthy	The University of Texas at Arlington, USA
Sanjay Chaudhary	Ahmedabad University, India

Program Committee Chairs

Ladjel Bellatreche	ENSMA, France
Hamido Fujita	Iwate Prefectural University, Japan
Vikram Goyal	IIIT Delhi, India

Organizing Chair

Shelly Sachdeva NIT Delhi, India

Publication Chair

Subhash Bhalla The University of Aizu, Japan

Workshop Chairs

Sanjay Chaudhary Ahmedabad University, India
Vasudha Bhatnagar University of Delhi, India
Satish Narayana Srirama University of Tartu, Estonia

Tutorial Chairs

Sanjay Kumar Madria Missouri University of Science and Technology, USA
Punam Bedi University of Delhi, India

Publicity Chairs

Manu Awasthi Ashoka University, India
Sonali Agarwal IIIT Allahabad, India
Rajiv Ratn Shah IIIT Delhi, India

Panel Chair

Sharma Chakravarthy The University of Texas at Arlington, USA

Website Chair

Samant Saurabh IIM Bodh Gaya, India

Website Administrator

Raghav Mittal Ashoka University, India

Program Committee

Satish Narayana University of Tartu, Estonia
Naresh Manwani IIIT Hyderabad, India
Jun Sasaki Iwate Prefectural University, Japan
Praveen Rao University of Missouri, USA
Nazha Selmaoui-Folcher University of New Caledonia, France
Morteza Zihayat Ryerson University, Canada
Tin Truong Chi Dalat University, Vietnam

Ji Zhang	University of Southern Queensland, Australia
Himanshu Gupta	IBM Research, India
Jose Maria Luna	University of Córdoba, Spain
Chun Wei	Western Norway University of Applied Sciences, Norway
Srinath Srinavasa	IIIT Bangalore, India
Engelbert Mephu Nguifo	Université Clermont Auvergne, France
Uday Kiran	The University of Aizu, Japan
Sebastián Ventura	University of Córdoba, Spain
Prem Prakash Jayaraman	Swinburne University of Technology, Australia
Santhanagopalan	IIIT Bangalore, India
Rajagopalan	
Tzung-Pei Hong	National University of Kaohsiung, Taiwan
Amin Beheshti	Macquarie University, Australia
Andres Gerardo	ESIME, Mexico
Hernandez-Matamoros	
Rinkle Aggarwal	Thapar Institute of Engineering and Technology, India
Amin Mesmoudi	LIAS-Université de Poitiers, France
Sadok Ben Yahia	Tallinn University of Technology, Estonia
Sangeeta Mittal	IIIT, India
Carlos Ordóñez	University of Houston, USA
Alok Singh	University of Hyderabad, India
Samant Saurabh	IIM Bodh Gaya, India

Sponsoring Institutions

National Institute of Technology Delhi, India
 Indraprastha Institute of Information Technology Delhi, India
 International Institute of Information Technology Hyderabad, India
 The University of Aizu, Japan
 University of Delhi, India
 Ahmedabad University, Ahmedabad, India

Contents

Data Science: Systems

A Comparison of Data Science Systems.	3
<i>Carlos Ordonez</i>	
Architectural Patterns for Integrating Data Lakes into Data Warehouse Architectures	12
<i>Olaf Herden</i>	
Study and Understanding the Significance of Multilayer-ELM Feature Space	28
<i>Rajendra Kumar Roul</i>	
Spectral Learning of Semantic Units in a Sentence Pair to Evaluate Semantic Textual Similarity	49
<i>Akanksha Mehndiratta and Krishna Asawa</i>	

Data Science: Architectures

i-Fence: A Spatio-Temporal Context-Aware Geofencing Framework for Triggering Impulse Decisions	63
<i>Jaiteg Singh, Amit Mittal, Ruchi Mittal, Karamjeet Singh, and Varun Malik</i>	
Face Mask Detection Using Transfer Learning of InceptionV3	81
<i>G. Jignesh Chowdary, Narinder Singh Pun, Sanjay Kumar Sonbhadra, and Sonali Agarwal</i>	
Analysis of GPS Trajectories Mapping on Shape Files Using Spatial Computing Approaches	91
<i>Saravjeet Singh and Jaiteg Singh</i>	

Big Data Analytics in Healthcare

A Transfer Learning Approach to Classify the Brain Age from MRI Images	103
<i>Animesh Kumar, Pramod Pathak, and Paul Styne</i>	

'Precision Health': Balancing Reactive Care and Proactive Care Through the Evidence Based Knowledge Graph Constructed from Real-World Electronic Health Records, Disease Trajectories, Diseasesome, and Patholome	113
<i>Asoke K Talukder, Julio Bonis Sanz, and Jahnavi Samajpati</i>	
Prediction for the Second Wave of COVID-19 in India	134
<i>Shweta Thakur, Dhaval K. Patel, Brijesh Soni, Mehul Raval, and Sanjay Chaudhary</i>	
Texture Feature Extraction: Impact of Variants on Performance of Machine Learning Classifiers: Study on Chest X-Ray – Pneumonia Images.	151
<i>Anamika Gupta, Anshuman Gupta, Vaishnavi Verma, Aayush Khattar, and Devansh Sharma</i>	
Computer-Aided Diagnosis of Thyroid Dysfunction: A Survey	164
<i>Bhavisha S. Parmar and Mayuri A. Mehta</i>	
Information Interchange of Web Data Resources	
Generic Key Value Extractions from Emails.	193
<i>Rajeev Gupta</i>	
The Next Generation Web: Technologies and Services.	209
<i>Asoke K. Talukder</i>	
Adversarial Collusion on the Web: State-of-the-Art and Future Directions . . .	230
<i>Hridoy Sankar Dutta and Tanmoy Chakraborty</i>	
Comparing Performance of Classifiers Applied to Disaster Detection in Twitter Tweets – Preliminary Considerations	236
<i>Maryan Plakhtiy , Maria Ganzha, and Marcin Paprzycki</i>	
Business Analytics	
Applying Machine Learning to Anomaly Detection in Car Insurance Sales. . .	257
<i>Michał Piesio , Maria Ganzha, and Marcin Paprzycki</i>	
Authorship Identification Using Stylometry and Document Fingerprinting . . .	278
<i>Shubham Yadav, Santosh Singh Rathore, and Satyendra Singh Chouhan</i>	
A Revenue-Based Product Placement Framework to Improve Diversity in Retail Businesses.	289
<i>Pooja Gaur, P. Krishna Reddy, M. Kumara Swamy, and Anirban Mondal</i>	

Recommending Question-Answers for Enriching Textbooks	308
<i>Shobhan Kumar and Arun Chauhan</i>	
OWI: Open-World Intent Identification Framework for Dialog Based System	329
<i>Jitendra Parmar, Sanskar Soni, and Satyendra Singh Chouhan</i>	
Author Index	345