

*Commenced Publication in 1973*

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

## Editorial Board

David Hutchison

*Lancaster University, Lancaster, UK*

Takeo Kanade

*Carnegie Mellon University, Pittsburgh, PA, USA*

Josef Kittler

*University of Surrey, Guildford, UK*

Jon M. Kleinberg

*Cornell University, Ithaca, NY, USA*

Friedemann Mattern

*ETH Zurich, Zurich, Switzerland*

John C. Mitchell

*Stanford University, Stanford, CA, USA*

Moni Naor

*Weizmann Institute of Science, Rehovot, Israel*

C. Pandu Rangan

*Indian Institute of Technology, Madras, India*

Bernhard Steffen

*TU Dortmund University, Dortmund, Germany*

Demetri Terzopoulos

*University of California, Los Angeles, CA, USA*

Doug Tygar

*University of California, Berkeley, CA, USA*

Gerhard Weikum

*Max Planck Institute for Informatics, Saarbrücken, Germany*

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

Ying Tan · Hideyuki Takagi  
Yuhui Shi (Eds.)

# Data Mining and Big Data

Second International Conference, DMBD 2017  
Fukuoka, Japan, July 27 – August 1, 2017  
Proceedings

*Editors*  
Ying Tan  
Peking University  
Beijing  
China

Hideyuki Takagi  
Kyushu University  
Fukuoka  
Japan

Yuhui Shi  
Southern University of Science  
and Technology  
Shenzhen  
China

ISSN 0302-9743                      ISSN 1611-3349 (electronic)  
Lecture Notes in Computer Science  
ISBN 978-3-319-61844-9              ISBN 978-3-319-61845-6 (eBook)  
DOI 10.1007/978-3-319-61845-6

Library of Congress Control Number: 2017945275

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

© Springer International Publishing AG 2017

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 volume (LNCS vol. 10387) constitutes the proceedings of the Second International Conference on Data Mining and Big Data (DMBD 2017), which was held in conjunction with the 8th International Conference on Swarm Intelligence (ICSI 2017), from July 27 to August 1, 2017, in Fukuoka, Japan.

The Second International Conference on Data Mining and Big Data (DMBD 2017) (IEEE Conference Record 41362) served as an international forum for researchers and practitioners to exchange the latest advances in theories, technologies, and applications of data mining and big data. The theme of DMBD 2017 was “Serving Life with Data Science.” DMBD 2017 was the second conference in the series after the successful first event (DMBD 2016) at Bali Island, Indonesia, during June 25–29, 2016.

Data mining refers to the activity of going through big data sets to look for relevant or pertinent information. This type of activity is a good example of “looking for a needle in a haystack.” The idea is that businesses collect massive sets of data that may be homogeneous or automatically collected. Decision-makers need access to smaller, more specific pieces of data from these large sets. They use data mining to uncover the pieces of information that will inform leadership and help chart the course for a business. Big data contains a huge amount of data and information and is worth researching in depth. Big data, also known as massive data or mass data, refers to the amount of data involved that are too great to be interpreted by a human. Currently, the suitable technologies include data mining, crowdsourcing, data fusion and integration, machine learning, natural language processing, simulation, time series analysis, and visualization. It is important to find new methods to enhance the effectiveness of big data. With the advent of big data analysis and intelligent computing techniques we are facing new challenges to make the information transparent and understandable efficiently. DMBD 2017 provided an excellent opportunity and an academic forum for academics and practitioners to present and discuss the latest scientific results, methods, and innovative ideas and advantages in theories, technologies, and applications in data mining, big data, and intelligent computing. The technical program covered all aspects of data mining, big data, and swarm intelligence as well as intelligent computing methods applied to all fields of computer science, machine learning, data mining and knowledge discovery, robotics, big data, scheduling, parallel realization, etc.

DMBD 2017 took place in the center of the historical Fukuoka City. Fukuoka is the fifth largest city in Japan with 1.55 million inhabitants and is the seventh most liveable city in the world according to the 2016 Quality of Life Survey by *Monocle*. Fukuoka is the northern end of Kyushu Island and is the economic and cultural center of Kyushu Island. Because of its closeness to the Asian mainland, Fukuoka has been an important harbor city for many centuries. Today’s Fukuoka is the product of the fusion of two cities in the year 1889, when the port city of Hakata and the former castle town of Fukuoka were united into one city called Fukuoka. The participants of ICSI 2017 enjoyed traditional Japanese dances, the local cuisine, beautiful landscapes, and the

hospitality of the Japanese people in modern Fukuoka, whose sites are part of UNESCO's World Heritage.

DMBD 2017 received 96 submissions from about 231 authors in 34 countries and regions (Algeria, Australia, Bangladesh, Brazil, Brunei Darussalam, Bulgaria, China, Colombia, Ecuador, France, Germany, Hong Kong SAR China, India, Indonesia, Iran, Japan, Malaysia, The Netherlands, Pakistan, Poland, Portugal, Romania, Russia, Serbia, Slovakia, South Africa, South Korea, Spain, Chinese Taiwan, Thailand, Turkey, USA, UK, and Vietnam) across six continents (Asia, Europe, North America, South America, Africa, and Oceania). Each submission was reviewed by at least two reviewers, and on average 2.6 reviewers. Based on rigorous reviews by the Program Committee members and reviewers, 53 high-quality papers were selected for publication in this proceedings volume with an acceptance rate of 55.21%. The papers are organized in 13 cohesive sections covering major topics of data mining and big data.

On behalf of the Organizing Committee of DMBD 2017, we would like to express sincere thanks to the Research Center for Applied Perceptual Science of Kyushu University and the Computational Intelligence Laboratory of Peking University for their sponsorship, to the IEEE Computational Intelligence Society for its technical sponsorship, to the Japan Chapter of IEEE Systems, Man and Cybernetics Society for its technical co-sponsorship, as well as to our supporters the International Neural Network Society, World Federation on Soft Computing, IEEE Beijing Section, Beijing Xinghui Hi-Tech Co., and Springer. We would also like to thank the members of the Advisory Committee for their guidance, the members of the international Program Committee and additional reviewers for reviewing the papers, and the members of the Publications Committee for checking the accepted papers in a short period of time. We are particularly grateful to Springer for publishing the proceedings in the prestigious series *Lecture Notes in Computer Science*. Moreover, we wish to express our heartfelt appreciation to the plenary speakers, session chairs, and student helpers. In addition, there are still many more colleagues, associates, friends, and supporters who helped us in immeasurable ways; we express our sincere gratitude to them all. Last but not the least, we would like to thank all the speakers, authors, and participants for their great contributions that made DMBD 2017 successful and all the hard work worthwhile.

May 2017

Ying Tan  
Hideyuki Takagi  
Yuhui Shi

# Organization

## General Co-chairs

Ying Tan	Peking University, China
Hideyuki Takagi	Kyushu University, Japan

## Program Committee Chair

Yuhui Shi	Southern University of Science and Technology, China
-----------	--

## Advisory Committee Co-chairs

Russell C. Eberhart	IUPUI, USA
Gary G. Yen	Oklahoma State University, USA
Hisao Ishibuchi	Osaka Prefecture University, Japan

## Technical Committee Co-chairs

Kay Chen Tan	National University of Singapore, Singapore
Xiaodong Li	RMIT University, Australia
Nikola Kasabov	Auckland University of Technology, New Zealand
Ponnuthurai N. Suganthan	Nanyang Technological University, Singapore

## Plenary Session Co-chairs

Mengjie Zhang	Victoria University of Wellington, New Zealand
Andreas Engelbrecht	University of Pretoria, South Africa

## Invited Session Co-chairs

Yan Pei	University of Aizu, Japan
Chaomin Luo	University of Detroit Mercy, USA

## Special Sessions Co-chairs

Shangce Gao	University of Toyama, Japan
Ben Niu	Shenzhen University, China
Qirong Tang	Tongji University, China

## **Tutorial Co-chairs**

Milan Tuba	John Naisbitt University, Serbia
Andreas Janecek	University of Vienna, Austria

## **Publications Co-chairs**

Swagatam Das	Indian Statistical Institute, India
Xinshe Yang	Middlesex University, UK

## **Publicity Co-chairs**

Yew-Soon Ong	Nanyang Technological University, Singapore
Carlos Coello	CINVESTAV-IPN, Mexico
Yaochu Jin	University of Surrey, UK
Shi Cheng	Shanxi Normal University, China
Bin Xue	Victoria University of Wellington, New Zealand

## **Finance and Registration Co-chairs**

Chao Deng	Peking University, China
Suicheng Gu	Google Corporation, USA

## **Local Arrangements Chair**

Ryohei Funaki	Kyushu University, Japan
---------------	--------------------------

## **Conference Secretariat**

Xiangyu Liu	Peking University, China
-------------	--------------------------

## **ICSI 2017 International Program Committee**

Mohd Helmy Abd Wahab	Universiti Tun Hussein Onn Malaysia, Malaysia
Miltos Alamaniotis	Purdue University, USA
Tomasz Andrysiak	UTP Bydgoszcz, Poland
Duong Tuan Anh	HoChiMinh City University of Technology, Vietnam
Carmelo J.A. Bastos Filho	University of Pernambuco, Brazil
David Camacho	Universidad Autonoma de Madrid, Spain
Vinod Chandra S.S.	Kerala University, India
Yenming J. Chen	National Kaohsiung First University of Science and Technology, Taiwan
Shi Cheng	Shaanxi Normal University, China
Jose Alfredo Ferreira Costa	Federal University
Bogusław Cyganek	AGH University of Science and Technology, Poland
Ke Ding	Tencent Corporation



Teresa Guarda	Universidad Estatal de Peninsula de Santa Elena, Ecuador
Cem Iyigun	Middle East Technical University, Turkey
Dariusz Jankowski	Wrocław University of Technology, Poland
Mingyan Jiang	Shandong University, China
Chen Junfeng	Hohai University, China
Imed Kacem	Université de Lorraine, France
Kalinka Kaloyanova	University of Sofia, Bulgaria
Germano Lambert-Torres	PS Solutions
Wei-Po Lee	National Sun Yat-sen University, Taiwan
Bin Li	University of Science and Technology of China
Michael Li	Central Queensland University, Australia
Churn-Jung Liao	Academia Sinica at Taipei, Taiwan
Andrei Lihu	Politehnica University of Timisoara, Romania
Wenjian Luo	University of Science and Technology of China
Wojciech Macyna	Wrocław University of Technology, Poland
Guojun Mao	Central University of Finance and Economics
Vasanth Kumar Mehta	SCSVMV University
Mohamed Arezki Mellal	M'Hamed Bougara University
Sanaz Mostaghim	Institute IWS
Sheak Rashed Haider Noori	Daffodil International University, Bangladesh
Somnuk Phon-Amnuaisuk	Universiti Teknologi, Brunei
Gerald Schaefer	Loughborough University, UK
Manik Sharma	DAV University
Ivan Silva	University of São Paulo, Brazil
Pramod Kumar Singh	ABV-IIITM Gwalior, India
Hung-Min Sun	National Tsing Hua University, Taiwan
Ying Tan	Peking University, China
Paulo Trigo	ISEL
Milan Tuba	John Naisbitt University, Serbia
Agnieszka Turek	Warsaw University of Technology, Poland
Gai-Ge Wang	Jiangsu Normal University, China
Guoyin Wang	Chongqing University of Posts and Telecommunications, China
Lei Wang	Tongji University, China
Zhenzhen Wang	Jinling Institute of Technology, Nanjing, China
Ka-Chun Wong	City University of Hong Kong, SAR China
Michal Wozniak	Wrocław University of Technology, Poland
Bo Xing	University of Johannesburg, South Africa
Yingjie Yang	De Montfort University, UK
Jie Zhang	Newcastle University, UK
Qieshi Zhang	Shaanxi Normal University, China

**Additional Reviewers for ICSI 2017**

Diván, Mario José  
Fernandes, Bruno  
Guo, Yuqian  
Habib, Md. Tarek  
Junyi, Chen  
Li, Xiangtao

Lima, Emerson  
Lin, Jiecong  
Saganowski, Łukasz  
Yan, Shankai  
Zhang, Jiao

# Contents

## Association Analysis

A Process for Exploring Employees' Relationships via Social Network and Sentiment Analysis . . . . .	3
<i>Jeydels Barahona and Hung-Min Sun</i>	
Mining Relationship Between User Purposes and Product Features Towards Purpose-Oriented Recommendation . . . . .	9
<i>Sopheaktra Yong and Yasuhito Asano</i>	
Finding Top-k Fuzzy Frequent Itemsets from Databases. . . . .	22
<i>Haifeng Li, Yue Wang, Ning Zhang, and Yuejin Zhang</i>	
Association Rule Mining in Healthcare Analytics . . . . .	31
<i>S. Anand Hareendran and S.S. Vinod Chandra</i>	
Does Student's Diligence to Study Relate to His/Her Academic Performance? . . . . .	40
<i>Toshiro Minami, Yoko Ohura, and Kensuke Baba</i>	
Correlation Analysis of Diesel Engine Performance Testing Data Based on Mixed-Copula Method. . . . .	48
<i>Zha Dongye, Qin Wei, Zhang Jie, and Zhuang Zilong</i>	

## Clustering

Comparative Study of Apache Spark MLlib Clustering Algorithms . . . . .	61
<i>Sasan Harifi, Ebrahim Byagowi, and Madjid Khalilian</i>	
L-DP: A Hybrid Density Peaks Clustering Method . . . . .	74
<i>Mingjing Du and Shifei Ding</i>	

## Prediction

Incremental Adaptive Time Series Prediction for Power Demand Forecasting . . . . .	83
<i>Petra Vrabecová, Viera Rozínajová, and Anna Bou Ezzeddine</i>	
Food Sales Prediction with Meteorological Data — A Case Study of a Japanese Chain Supermarket . . . . .	93
<i>Xin Liu and Ryutaro Ichise</i>	

Cascade Spatial Autoregression for Air Pollution Prediction . . . . .	105
<i>Yangping Li, Xiaorui Wei, and Tianming Hu</i>	

Machine Learning Techniques for Prediction of Pre-fetched Objects in Handling Big Data Storage. . . . .	113
<i>Nur Syahela Hussien, Sarina Sulaiman, and Siti Mariyam Shamsuddin</i>	

**Classification**

Spectral-Spatial Mineral Classification of Lunar Surface Using Band Parameters with Active Learning. . . . .	127
<i>Sukanta Roy, Sujai Subbanna, Srinidhi Venkatesh Channagiri, Sharath R Raj, and Omkar S.N</i>	

R <sup>2</sup> CEDM: A Rete-Based RFID Complex Event Detection Method . . . . .	137
<i>Xiaoli Peng, Linjiang Zheng, and Ting Liao</i>	

Learning Analytics for SNS-Integrated Virtual Learning Environment . . . . .	148
<i>Fang-Fang Chua, Chia-Ying Khor, and Su-Cheng Haw</i>	

LBP vs. LBP Variance for Texture Classification . . . . .	156
<i>Gerald Schaefer and Niraj Doshi</i>	

A Decision Tree of Ignition Point for Simple Inflammable Chemical Compounds . . . . .	165
<i>Ryoko Hayashi</i>	

Analyzing Consumption Behaviors of Pet Owners in a Veterinary Hospital . . . . .	173
<i>Jo Ting Wei, Shih-Yen Lin, You-Zhen Yang, and Hsin-Hung Wu</i>	

**Schedule and Sequence Analysis**

Modeling Inter-country Connection from Geotagged News Reports: A Time-Series Analysis . . . . .	183
<i>Yihong Yuan</i>	

Mining Sequential Patterns of Students' Access on Learning Management System . . . . .	191
<i>Leonard K.M. Poon, Siu-Cheung Kong, Michael Y.W. Wong, and Thomas S.H. Yau</i>	

Design of a Dynamic Electricity Trade Scheduler Based on Genetic Algorithms. . . . .	199
<i>Junghoon Lee and Gyung-Leen Park</i>	

Increasing Coverage of Information Spreading in Social Networks with Supporting Seeding . . . . .	209
<i>Jarosław Jankowski and Radosław Michalski</i>	

## Big Data

B-Learning and Big Data: Use in Training an Engineering Course . . . . .	221
<i>Leonardo Emiro Contreras Bravo, Jose Ignacio Rodriguez Molano, and Giovanni Mauricio Tarazona Bermudez</i>	
Mapping Knowledge Domain Research in Big Data: From 2006 to 2016. . . .	234
<i>Li Zeng, Zili Li, Tong Wu, and Lixin Yang</i>	
Template Based Industrial Big Data Information Extraction and Query System. . . . .	247
<i>Jie Wang, Yan Peng, Yun Lin, and Kang Wang</i>	
Finding the Typical Communication Black Hole in Big Data Environment. . .	255
<i>Jinbo Zhang, Shenda Hong, Yiyong Lin, Yanchao Mou, and Hongyan Li</i>	
A Solution for Mining Big Data Based on Distributed Data Streams and Its Classifying Algorithms . . . . .	263
<i>Guojun Mao and Jiewei Qiao</i>	

## Data Analysis

Design of a Quality-Aware Data Capture System . . . . .	275
<i>R. Vasanth Kumar Mehta and Shubham Verma</i>	
Data Architecture for the Internet of Things and Industry 4.0 . . . . .	283
<i>José Ignacio Rodriguez Molano, Leonardo Emiro Contreras Bravo, and Eduyn Ramiro López Santana</i>	
Machine Learning in Data Lake for Combining Data Silos. . . . .	294
<i>Merlinda Wibowo, Sarina Sulaiman, and Siti Mariyam Shamsuddin</i>	
Development of 3D Earth Visualization for Taiwan Ocean Environment Demonstration . . . . .	307
<i>Franco Lin, Wen-Yi Chang, Whey-Fone Tsai, and Chien-Chou Shih</i>	
Exploring Potential Use of Mobile Phone Data Resource to Analyze Inter-regional Travel Patterns in Japan . . . . .	314
<i>Canh Xuan Do and Makoto Tsukai</i>	
REBUILD: Graph Embedding Based Method for User Social Role Identity on Mobile Communication Network . . . . .	326
<i>Jinbo Zhang, Yipeng Chen, Shenda Hong, and Hongyan Li</i>	

Study a Join Query Strategy Over Data Stream Based on Sliding Windows . . . .	334
<i>Yang Sun, Lin Teng, Shoulin Yin, Jie Liu, and Hang Li</i>	

**Data Mining**

L2 Learners' Proficiency Evaluation Using Statistics Based on Relationship Among CEFR Rating Scales. . . . .	345
<i>Hajime Tsubaki</i>	
Research Hotspots and Trends in Data Mining: From 1993 to 2016 . . . . .	353
<i>Zili Li and Li Zeng</i>	
Proposal for a New Reduct Method for Decision Tables and an Improved STRIM . . . . .	366
<i>Jiwei Fei, Tetsuro Saeki, and Yuichi Kato</i>	
Post-marketing Drug Safety Evaluation Using Data Mining Based on FAERS . . . . .	379
<i>Rui Duan, Xinyuan Zhang, Jingcheng Du, Jing Huang, Cui Tao, and Yong Chen</i>	
Crowd Density Estimation from Few Radio-Frequency Tracking Devices: I. A Modelling Framework. . . . .	390
<i>Yenming J. Chen and Albert Jing-Fuh Yang</i>	

**Text Mining**

Mining Textual Reviews with Hierarchical Latent Tree Analysis. . . . .	401
<i>Leonard K.M. Poon, Chun Fai Leung, and Nevin L. Zhang</i>	
Extraction of Temporal Events from Clinical Text Using Semi-supervised Conditional Random Fields . . . . .	409
<i>Gandhimathi Moharasan and Tu-Bao Ho</i>	
Characteristics of Language Usage in Inquires Asked to an Online Help Desk . . . . .	422
<i>Haruka Adachi and Mikito Toda</i>	

**Deep Learning**

A Novel Diagnosis Method for SZ by Deep Neural Networks . . . . .	433
<i>Chen Qiao, Yan Shi, Bin Li, and Tai An</i>	
A Novel Data-Driven Fault Diagnosis Method Based on Deep Learning . . . .	442
<i>Yuyan Zhang, Liang Gao, Xinyu Li, and Peigen Li</i>	

## High Performance Computing

Simulated Precipitation and Reservoir Inflow in the Chao Phraya River Basin by Multi-model Ensemble CMIP3 and CMIP5 . . . . .	455
<i>Thannob Aribarg and Seree Supratid</i>	
An IDL-Based Parallel Model for Scientific Computations on Multi-core Computers . . . . .	464
<i>Weili Kou, Lili Wei, Changxian Liang, Ning Lu, and Qiuhua Wang</i>	
Accelerating Redis with RDMA Over InfiniBand . . . . .	472
<i>Wenhui Tang, Yutong Lu, Nong Xiao, Fang Liu, and Zhiguang Chen</i>	

## Knowledge Base and its Framework

Application of Decision Trees in the Development of a Knowledge Base for a System of Support for Revitalization Processes . . . . .	487
<i>Agnieszka Turek</i>	
ABC Metaheuristic Based Optimized Adaptation Planning Logic for Decision Making Intelligent Agents in Self Adaptive Software System . . . . .	496
<i>Binu Rajan and Vinod Chandra</i>	
A Knowledge-Based Framework for Mitigating Hydro-Meteorological Disasters . . . . .	505
<i>Pg. Hj. Asmali Pg. Badarudin, Thien Wan Au, and Somnuk Phon-Amnuaisuk</i>	

## Fuzzy Control

Improved Stability and Stabilization Criteria for T-S Fuzzy Systems with Distributed Time-Delay. . . . .	517
<i>Qianqian Ma, Hongwei Xia, Guangcheng Ma, Yong Xia, and Chong Wang</i>	
Adaptive Neuro-Fuzzy Inference System: Overview, Strengths, Limitations, and Solutions . . . . .	527
<i>Mohd Najib Mohd Salleh, Noureen Talpur, and Kashif Hussain</i>	
Acquisition of Knowledge in the Form of Fuzzy Rules for Cases Classification . . . . .	536
<i>Tatiana Avdeenko and Ekaterina Makarova</i>	

<b>Author Index</b> . . . . .	545
-------------------------------	-----