

Contents

Challenges in Data Mining and Big Data

Evolutionary Computation and Big Data: Key Challenges and Future Directions	3
<i>Shi Cheng, Bin Liu, Yuhui Shi, Yaochu Jin, and Bin Li</i>	
Prospects and Challenges in Online Data Mining: Experiences of Three-Year Labour Market Monitoring Project	15
<i>Maxim Bakaev and Tatiana Avdeenko</i>	

Data Mining Algorithms

Enhance AdaBoost Algorithm by Integrating LDA Topic Model.	27
<i>Fangyu Gai, Zhiqiang Li, Xinwen Jiang, and Hongchen Guo</i>	
An Improved Algorithm for MicroRNA Profiling from Next Generation Sequencing Data	38
<i>Salim A., Amjesh R., and Vinod Chandra S.S.</i>	
Utilising the Cross Industry Standard Process for Data Mining to Reduce Uncertainty in the Measurement and Verification of Energy Savings	48
<i>Colm V. Gallagher, Ken Bruton, and Dominic T.J. O'Sullivan</i>	
Implementing Majority Voting Rule to Classify Corporate Value Based on Environmental Efforts	59
<i>Ratna Hidayati, Katsutoshi Kanamori, Ling Feng, and Hayato Ohwada</i>	
Model Proposal of Knowledge Management for Technology Based Companies	67
<i>Jorge Leonardo Puentes Morantes, Nancy Yurani Ortiz Guevara, and José Ignacio Rodríguez Molano</i>	

Frequent Itemset Mining

Oracle and Vertica for Frequent Itemset Mining	77
<i>Hristo Kyurkchiev and Kalinka Kaloyanova</i>	
Reconstructing Positive Surveys from Negative Surveys with Background Knowledge	86
<i>Dongdong Zhao, Wenjian Luo, and Lihua Yue</i>	

Spatial Data Mining

Application of the Spatial Data Mining Methodology and Gamification for the Optimisation of Solving the Transport Issues of the “Varsovian Mordor”	103
<i>Robert Olszewski and Agnieszka Turek</i>	
A Geo-Social Data Model for Moving Objects	115
<i>Hengcai Zhang, Feng Lu, and Jie Chen</i>	
Optimization on Arrangement of Precaution Areas Serving for Ships’ Routeing in the Taiwan Strait Based on Massive AIS Data	123
<i>Jinhai Chen, Feng Lu, Mingxiao Li, Pengfei Huang, Xiliang Liu, and Qiang Mei</i>	

Prediction

Bulk Price Forecasting Using Spark over NSE Data Set.	137
<i>Vijay Krishna Menon, Nithin Chekravarthi Vasireddy, Sai Aswin Jami, Viswa Teja Naveen Pedomallu, Varsha Sureshkumar, and K.P. Soman</i>	
Prediction and Survival Analysis of Patients After Liver Transplantation Using RBF Networks	147
<i>C.G. Raji and S.S. Vinod Chandra</i>	
Link Prediction by Utilizing Correlations Between Link Types and Path Types in Heterogeneous Information Networks	156
<i>Hyun Ji Jeong, Kim Taeyeon, and Myoung Ho Kim</i>	
Advanced Predictive Methods of Artificial Intelligence in Intelligent Transport Systems.	165
<i>Viliam Lendel, Lucia Pancikova, and Lukas Falat</i>	
Range Prediction Models for E-Vehicles in Urban Freight Logistics Based on Machine Learning	175
<i>Johannes Kretzschmar, Kai Gebhardt, Christoph Theiß, and Volkmar Schau</i>	

Feature Selection

Partitioning Based N-Gram Feature Selection for Malware Classification	187
<i>Weiwei Hu and Ying Tan</i>	
A Supervised Biclustering Optimization Model for Feature Selection in Biomedical Dataset Classification	196
<i>Saziye Deniz Oguz Arikan and Cem Iyigun</i>	

Term Space Partition Based Ensemble Feature Construction for Spam Detection	205
<i>Guyue Mi, Yang Gao, and Ying Tan</i>	

Information Extraction

Term Extraction from German Computer Science Textbooks	219
<i>Kevin Möhlmann and Jörn Syrbé</i>	
An FW-DTSS Based Approach for News Page Information Extraction	227
<i>Leiming Ma and Zhengyou Xia</i>	
A Linear Regression Approach to Multi-criteria Recommender System	235
<i>Tanisha Jhalani, Vibhor Kant, and Pragya Dwivedi</i>	

Classification

Classification of Power Quality Disturbances Using Forest Algorithm	247
<i>Fábbio Borges, Ivan Silva, Ricardo Fernandes, and Lucas Moraes</i>	
A Sequential k -Nearest Neighbor Classification Approach for Data-Driven Fault Diagnosis Using Distance- and Density-Based Affinity Measures	253
<i>Myeongsu Kang, Gopala Krishnan Ramaswami, Melinda Hodkiewicz, Edward Cripps, Jong-Myon Kim, and Michael Pecht</i>	
A Hybrid Model Combining SOMs with SVRs for Patent Quality Analysis and Classification	262
<i>Pei-Chann Chang, Jheng-Long Wu, Cheng-Chin Tsao, and Chin-Yuan Fan</i>	
Mining Best Strategy for Multi-view Classification	270
<i>Jing Peng and Alex J. Aved</i>	

Anomaly Pattern and Diagnosis

Detecting Variable Length Anomaly Patterns in Time Series Data.	279
<i>Ngo Duy Khanh Vy and Duong Tuan Anh</i>	
Bigger Data Is Better for Molecular Diagnosis Tests Based on Decision Trees	288
<i>Alexandru G. Floares, George A. Calin, and Florin B. Manolache</i>	
Waiting Time Screening in Diagnostic Medical Imaging – A Case-Based View.	296
<i>Marisa Esteves, Henrique Vicente, Sabino Gomes, António Abelha, M. Filipe Santos, José Machado, João Neves, and José Neves</i>	

Data Visualization Analysis

Real-Time Data Analytics: An Algorithmic Perspective	311
<i>Sarwar Jahan Morshed, Juwel Rana, and Marcelo Milrad</i>	
High-Dimensional Data Visualization Based on User Knowledge	321
<i>Qiaolian Liu, Jianfei Zhao, Naiwang Guo, Ding Xiao, and Chuan Shi</i>	
A Data Mining and Visual Analytics Perspective on Sustainability-Oriented Infrastructure Planning	330
<i>Dimitri N. Mavris, Michael Balchanos, WoongJe Sung, and Olivia J. Pinon</i>	
Visual Interactive Approach for Mining Twitter's Networks	342
<i>Youcef Abdelsadek, Kamel Chelghoum, Francine Herrmann, Imed Kacem, and Benoît Otjacques</i>	

Privacy Policy

Key Indicators for Data Sharing - In Relation with Digital Services.	353
<i>Sheak Rashed Haider Noori, Md. Kamrul Hossain, and Juwel Rana</i>	
Efficient Probabilistic Methods for Proof of Possession in Clouds	364
<i>Lukasz Krzywiecki, Krzysztof Majcher, and Wojciech Macyna</i>	
Cloud-Based Storage Model with Strong User Privacy Assurance	373
<i>Amir Rezapour, Wei Wu, and Hung-Min Sun</i>	

Social Media

The Role of Social Media in Innovation and Creativity: The Case of Chinese Social Media	381
<i>Jiwat Ram, Siqu Liu, and Andy Koronois</i>	
Malay Word Stemmer to Stem Standard and Slang Word Patterns on Social Media	391
<i>Mohamad Nizam Kassim, Mohd Aizaini Maarof, Anazida Zainal, and Amirudin Abdul Wahab</i>	
Two-Phase Computing Model for Chinese Microblog Sentimental Analysis . . .	401
<i>Jianyong Duan, Chao Wang, Mei Zhang, and Hui Liu</i>	
Local Community Detection Based on Bridges Ideas	409
<i>Xia Zhang, Zhengyou Xia, and Jiandong Wang</i>	
Environment for Data Transfer Measurement	416
<i>Sergey Khoruzhnikov, Vladimir Grudin, Oleg Sadov, Andrey Shevel, Stefanos Georgiou, and Arsen Kairkanov</i>	

Query Optimization and Processing Algorithm

Ant Colony-Based Approach for Query Optimization	425
<i>Hany A. Hanafy and Ahmed M. Gadallah</i>	
A Range Query Processing Algorithm Hiding Data Access Patterns in Outsourced Database Environment.	434
<i>Hyeong-Il Kim, Hyeong-Jin Kim, and Jae-Woo Chang</i>	

Big Data

Big Data Tools: Hadoop, MongoDB and Weka	449
<i>Paula Catalina Jaraba Navas, Yesid Camilo Guacaneme Parra, and José Ignacio Rodríguez Molano</i>	
Big Data Meaning in the Architecture of IoT for Smart Cities	457
<i>Christian David Gómez Romero, July Katherine Díaz Barriga, and José Ignacio Rodríguez Molano</i>	
Linear TV Recommender Through Big Data	466
<i>Mikhail A. Baklanov and Olga E. Baklanova</i>	
Data Models in NoSQL Databases for Big Data Contexts	475
<i>Maribel Yasmína Santos and Carlos Costa</i>	
Probabilistic Mining in Large Transaction Databases	486
<i>Hareendran S. Anand and S.S. Vinod Chandra</i>	

Computational Aspects of Pattern Recognition and Computer Vision

A First Attempt on Online Data Stream Classifier Using Context	497
<i>Michał Woźniak and Bogusław Cyganek</i>	
An Effective Semi-analytic Algorithm for Solving Helmholtz Equation in 1-D	505
<i>Chunhui Zhu, Lijun Liu, Yanhui Liu, and Zhen Yu</i>	
Ensemble of One-Dimensional Classifiers for Hyperspectral Image Analysis.	513
<i>Paweł Ksieniewicz, Bartosz Krawczyk, and Michał Woźniak</i>	
Industrial Internet of Things: An Architecture Prototype for Monitoring in Confined Spaces Using a Raspberry Pi	521
<i>José Ignacio Rodríguez Molano, Víctor Hugo Medina, and Javier Felipe Moncada Sánchez</i>	
Efficient Multidimensional Pattern Recognition in Kernel Tensor Subspaces . . .	529
<i>Bogusław Cyganek and Michał Woźniak</i>	

Fingerprint Reference Point Detection Based on High Curvature Points	538
<i>Krzysztof Wrobel, Rafal Doroz, and Piotr Porwik</i>	
Study of the Harmonic Distortion (THD) and Power Factor Oriented to the Luminaries Used in the Traffic Light System of Bogotá City, Colombia	548
<i>Carlos Andres Parra Martinez, Hugo Alejandro Serrato Vanegas, and José Ignacio Rodriguez Molano</i>	
Extraction of Dynamic Nonnegative Features from Multidimensional Nonstationary Signals	557
<i>Rafał Zdunek and Michalina Kotyla</i>	
Author Index	567

Data Mining and Big Data

First International Conference, DMBD 2016, Bali,
Indonesia, June 25-30, 2016. Proceedings

Tan, Y.; Shi, Y. (Eds.)

2016, XVI, 569 p. 141 illus., Softcover

ISBN: 978-3-319-40972-6