Contents

New Generation Data Warehouses Design

Evaluation of Data Warehouse Design Methodologies in the Context of Big Data .......................................................... 3
Francesco Di Tria, Ezio Lefons, and Filippo Tangorra

Optimal Task Ordering in Chain Data Flows: Exploring the Practicality of Non-scalable Solutions ......................................... 19
Georgia Kougka and Anastasios Gounaris

Exploiting Mathematical Structures of Statistical Measures for Comparison of RDF Data Cubes ........................................... 33
Claudia Diamantini, Domenico Potena, and Emanuele Storti

S2D: Shared Distributed Datasets, Storing Shared Data for Multiple and Massive Queries Optimization in a Distributed Data Warehouse ........ 42
Rado Ratsimbazafy, Omar Boussaid, and Fadila Bentayeb

Cloud and NoSQL Databases

Enforcing Privacy in Cloud Databases ............................................. 53
Somayeh Sobati Moghadam, Jérôme Darmont, and Gérald Gavin

TARDIS: Optimal Execution of Scientific Workflows in Apache Spark .... 74
Daniel Gaspar, Fabio Porto, Reza Akbarinia, and Esther Pacitti

MDA-Based Approach for NoSQL Databases Modelling ................... 88
Fatma Abdelhedi, Amal Ait Brahim, Faten Atigui, and Gilles Zurfluh

Advanced Programming Paradigms

MiSeRe-Hadoop: A Large-Scale Robust Sequential Classification Rules Mining Framework .................................................... 105
Elias Egho, Dominique Gay, Romain Trinquart, Marc Boullé, Nicolas Voisine, and Fabrice Clérot

An Efficient Map-Reduce Framework to Mine Periodic Frequent Patterns .... 120
Alampally Anirudh, R. Uday Kiran, P. Krishna Reddy, M. Toyoda, and Masaru Kitsuregawa
MapReduce-Based Complex Big Data Analytics over Uncertain and Imprecise Social Networks .................................................. 130
  Peter Braun, Alfredo Cuzzocrea, Fan Jiang, Carson Kai-Sang Leung, and Adam G.M. Pazdor

Non-functional Requirements Satisfaction

A Case for Abstract Cost Models for Distributed Execution of Analytics Operators ............................................................... 149
  Rundong Li, Ningfang Mi, Mirek Riedewald, Yizhou Sun, and Yi Yao

Pre-processing and Indexing Techniques for Constellation Queries in Big Data .......................................................... 164
  Amir Khatibi, Fabio Porto, Joao Guilherme Rittmeyer, Eduardo Ogasawara, Patrick Valduriez, and Dennis Shasha

A Lightweight Elastic Queue Middleware for Distributed Streaming Pipeline ............................................................. 173
  Weiping Qu and Stefan Dessloch

Modeling Data Flow Execution in a Parallel Environment ......................... 183
  Georgia Kougka, Anastasios Gounaris, and Ulf Leser

Machine Learning

Accelerating K-Means by Grouping Points Automatically .................. 199
  Qiao Yu and Bi-Ru Dai

A Machine Learning Trainable Model to Assess the Accuracy of Probabilistic Record Linkage ........................................... 214
  Robespierre Pita, Everton Mendonça, Sandra Reis, Marcos Barreto, and Spiros Denaxas

An Efficient Approach for Instance Selection ........................................ 228
  Joel Luís Carbonera

Search Result Personalization in Twitter Using Neural Word Embeddings ... 244
  Sameendra Samarawickrama, Shanika Karunasekera, Aaron Harwood, and Ramamohanarao Kotagiri

Diverse Selection of Feature Subsets for Ensemble Regression ............ 259
  Arvind Kumar Shekar, Patricia Iglesias Sánchez, and Emmanuel Müller

K-Means Clustering Using Homomorphic Encryption and an Updatable Distance Matrix: Secure Third Party Data Clustering with Limited Data Owner Interaction ........................................ 274
  Nawal Almutairi, Frans Coenen, and Keith Dures
Reweighting Forest for Extreme Multi-label Classification ........................................... 286
Zhun-Zheng Lin and Bi-Ru Dai

Social Media and Twitter Analysis

A Relativistic Opinion Mining Approach to Detect Factual or Opinionated News Sources ................................................................. 303
Erhan Sezerer and Selma Tekir

A Reliability-Based Approach for Influence Maximization Using the Evidence Theory ............................................................... 313
Siwar Jendoubi and Arnaud Martin

Sentiment Analysis on Twitter to Improve Time Series Contextual Anomaly Detection for Detecting Stock Market Manipulation ....................... 327
Koosha Golmohammadi and Osmar R. Zaiane

Automatic Segmentation of Big Data of Patent Texts ........................................... 343
Mustafa Sofean

Sentiment Analysis and User Influence

Tag Me a Label with Multi-arm: Active Learning for Telugu Sentiment Analysis ................................................................. 355
Sandeep Sricharan Mukku, Subba Reddy Oota, and Radhika Mamidi

Belief Temporal Analysis of Expert Users: Case Study Stack Overflow ............. 368
Dorra Attiaoui, Arnaud Martin, and Boutheina Ben Yaghlane

Leveraging Hierarchy and Community Structure for Determining Influencers in Networks ............................................................ 383
Sharanjit Kaur, Rakhi Saxena, and Vasudha Bhatnagar

Using Social Media for Word-of-Mouth Marketing ........................................... 391
Nagendra Kumar, Yash Chandarana, Konjengbam Anand, and Manish Singh

Knowledge Discovery

Knowledge Discovery of Complex Data Using Gaussian Mixture Models ......... 409
Linfei Zhou, Wei Ye, Claudia Plant, and Christian Böhm

Optimized Mining of Potential Positive and Negative Association Rules .......... 424
Parfait Bemarisika and André Totohasina
Big Data Analytics and Knowledge Discovery
Bellatreche, L.; Chakravarthy, S. (Eds.)
2017, XIV, 488 p. 137 illus., Softcover
ISBN: 978-3-319-64282-6