Contents

Similarity Measure and Clustering

Determining Query Readiness for Structured Data ............................... 3
   Farid Alborzi, Rada Chirkova, Jon Doyle, and Yahya Fathi

Efficient Cluster Detection by Ordered Neighborhoods ......................... 15
   Emin Aksehirli, Bart Goethals, and Emmanuel Müller

Unsupervised Semantic and Syntactic Based Classification of Scientific Citations ............................................ 28
   Mohammad Abdullatif, Yun Sing Koh, and Gillian Dobbie

Data Mining

HI-Tree: Mining High Influence Patterns Using External and Internal Utility Values ........................................... 43
   Yun Sing Koh and Russel Pears

Balancing Tree Size and Accuracy in Fast Mining of Uncertain Frequent Patterns ........................................... 57
   Carson Kai-Sang Leung and Richard Kyle MacKinnon

Secure Outsourced Frequent Pattern Mining by Fully Homomorphic Encryption ........................................... 70
   Junqiang Liu, Jiuyong Li, Shijian Xu, and Benjamin C.M. Fung

Supervised Evaluation of Top-k Itemset Mining Algorithms .................... 82
   Claudio Lucchese, Salvatore Orlando, and Raffaele Perego

Finding Banded Patterns in Data: The Banded Pattern Mining Algorithm ....... 95
   Fatimah B. Abdullahi, Frans Coenen, and Russell Martin

Discrimination-Aware Association Rule Mining for Unbiased Data Analytics ........................................... 108
   Ling Luo, Wei Liu, Irena Koprinska, and Fang Chen

Social Computing

Big Data Analytics of Social Networks for the Discovery of “Following” Patterns ........................................... 123
   Carson Kai-Sang Leung and Fan Jiang
Sentiment Extraction from Tweets: Multilingual Challenges .......................... 136
  Nantia Makrynioti and Vasilis Vassalos

TiDE: Template-Independent Discourse Data Extraction. ......................... 149
  Jayendra Barua, Dhaval Patel, and Vikram Goyal

Heterogeneous Networks and Data

A New Relevance Measure for Heterogeneous Networks ........................... 165
  Mukul Gupta, Pradeep Kumar, and Bharat Bhasker

UFOMQ: An Algorithm for Querying for Similar Individuals in Heterogeneous
  Ontologies ...................................................................................... 178
  Yinuo Zhang, Anand Panangadan, and Viktor K. Prasanna

Semantics-Based Multidimensional Query Over Sparse Data Marts .............. 190
  Claudia Diamantini, Domenico Potena, and Emanuele Storti

Data Warehouses

Automatically Tailoring Semantics-Enabled Dimensions for Movement
  Data Warehouses ........................................................................... 205
  Juarez A.P. Sacenti, Fabio Salvini, Renato Fileto, Alessandra Raffaetà,
  and Alessandro Roncato

Real-Time Snapshot Maintenance with Incremental ETL Pipelines in Data
  Warehouses. .................................................................................. 217
  Weiping Qu, Vinanthi Basavaraj, Sahana Shankar, and Stefan Dessloch

Eco-Processing of OLAP Complex Queries ............................................ 229
  Amine Roukh and Ladjel Bellatreche

Materializing Baseline Views for Deviation Detection Exploratory OLAP ... 243
  Pedro Furtado, Sergi Nadal, Veronika Peralta, Mahfoud Djedaini,
  Nicolas Labroche, and Patrick Marcel

Stream Processing

Binary Shapelet Transform for Multiclass Time Series Classification ........ 257
  Aaron Bostrom and Anthony Bagnall

StreamXM: An Adaptive Partitional Clustering Solution for Evolving Data
  Streams ......................................................................................... 270
  Robert Anderson and Yun Sing Koh
Data Stream Mining with Limited Validation Opportunity: Towards Instrument Failure Prediction

Katie Atkinson, Frans Coenen, Phil Goddard, Terry Payne, and Luke Riley

Distributed Classification of Data Streams: An Adaptive Technique

Alfredo Cuzzocrea, Mohamed Medhat Gaber, and Ary Mazharuddin Shiddiqi

New Word Detection and Tagging on Chinese Twitter Stream

Yuzhi Liang, Pengcheng Yin, and S.M. Yiu

Applications of Big Data Analysis

Text Categorization for Deriving the Application Quality in Enterprises

Thomas Zinner, Florian Lemmerich, Susanna Schwarzmann, Matthias Hirth, Peter Karg, and Andreas Hotho

MultiSpot: Spotting Sentiments with Semantic Aware Multilevel Cascaded Analysis

Despoina Chatzakou, Nikolaos Passalis, and Athena Vakali

Online Urban Mobility Detection Based on Velocity Features

Fernando Terroso-Saenz, Mercedes Valdes-Vela, and Antonio F. Skarmeta-Gomez

Big Data

Partition and Conquer: Map/Reduce Way of Substructure Discovery

Soumyava Das and Sharma Chakravarthy

Implementation of Multidimensional Databases with Document-Oriented NoSQL

M. Chevalier, M. El Malki, A. Kopliku, O. Teste, and R. Tournier

A Graph-Based Concept Discovery Method for n-Ary Relations

Nazmiye Ceren Abay, Alev Mutlu, and Pinar Karagoz

Exact Detection of Information Leakage in Database Access Control

Farid Alborzi, Rada Chirkova, and Ting Yu

Author Index
Big Data Analytics and Knowledge Discovery
17th International Conference, DaWaK 2015, Valencia, Spain, September 1-4, 2015, Proceedings
Madria, S.; Hara, T. (Eds.)
2015, XIII, 418 p. 116 illus., Softcover
ISBN: 978-3-319-22728-3