Contents

Graph Mining

Greedy Graph Edit Distance ........................................... 3
  Kaspar Riesen, Miquel Ferrer, Rolf Dornberger, and Horst Bunke

Learning Heuristics to Reduce the Overestimation of Bipartite Graph Edit Distance Approximation .............................. 17
  Miquel Ferrer, Francesc Serratosa, and Kaspar Riesen

Seizure Prediction by Graph Mining, Transfer Learning, and Transformation Learning .................................................. 32
  Nimit Dhulekar, Srinivas Nambirajan, Basak Oztan, and Bülent Yener

Classification and Regression

Local and Global Genetic Fuzzy Pattern Classifiers .................. 55
  Søren Atmakuri Davidsen, E. Sreedevi, and M. Padmavathamanna

IKLTSAn: An Incremental Kernel LTSA Method ...................... 70
  Chao Tan, Jihong Guan, and Shuiqeng Zhou

Sentiment Analysis

SentiSAIL: Sentiment Analysis in English, German and Russian ... 87
  Gayane Shalunts and Gerhard Backfried

Sentiment Analysis for Government: An Optimized Approach .... 98
  Angelo Corallo, Laura Fortunato, Marco Matera, Marco Alessi,
  Alessio Camillò, Valentina Chetta, Enza Giangreco, and Davide Storelli

Data Preparation and Missing Values

A Novel Algorithm for the Integration of the Imputation of Missing Values and Clustering ........................................... 115
  Roni Ben Ishay and Maya Herman

Improving the Algorithm for Mapping of OWL to Relational Database Schema ......................................................... 130
  Chien D.C. Ta and Tuoi Phan Thi

Robust Principal Component Analysis of Data with Missing Values 140
  Tommi Kärkkäinen and Mirka Saarela
Association and Sequential Rule Mining

Efficient Mining of High-Utility Sequential Rules .......................... 157
   Souleymane Zida, Philippe Fournier-Viger, Cheng-Wei Wu,
   Jerry Chun-Wei Lin, and Vincent S. Tseng

MOGACAR: A Method for Filtering Interesting Classification
Association Rules ......................................................... 172
   Diana Benavides Prado

Support Vector Machines

Classifying Grasslands and Cultivated Pastures in the Brazilian Cerrado
Using Support Vector Machines, Multilayer Perceptrons and Autoencoders. . . 187
   Wanderson Costa, Leila Fonseca, and Thales Körting

Hybrid Approach for Inductive Semi Supervised Learning
Using Label Propagation and Support Vector Machine .......................... 199
   Aruna Govada, Pravin Joshi, Sahil Mittal, and Sanjay K. Sahay

Frequent Item Set Mining and Time Series Analysis

Optimizing the Data-Process Relationship for Fast Mining of Frequent
Itemsets in MapReduce .................................................. 217
   Saber Salah, Reza Akbarinia, and Florent Masseglia

Aggregation-Aware Compression of Probabilistic Streaming Time Series . . 232
   Reza Akbarinia and Florent Masseglia

Clustering

Applying Clustering Analysis to Heterogeneous Data Using Similarity
Matrix Fusion (SMF) ..................................................... 251
   Aalaa Mojahed, Joao H. Bettencourt-Silva, Wenjia Wang,
   and Beatriz de la Iglesia

On Bicluster Aggregation and its Benefits for Enumerative Solutions ....... 266
   Saullo Oliveira, Rosana Veroneze, and Fernando J. Von Zuben

Semi-Supervised Stream Clustering Using Labeled Data Points .............. 281
   Kritsana Treechalong, Thanawin Rakthanmanon,
   and Kitsana Waiyamai

Avalanche: A Hierarchical, Divisive Clustering Algorithm ..................... 296
   Paul K. Amalaman and Christoph F. Eick
Text Mining

Author Attribution of Email Messages Using Parse-Tree Features . . . . . . . . . . 313
  Jagadeesh Patchala, Raj Bhatnagar, and Sridharan Gopalakrishnan

Query Click and Text Similarity Graph for Query Suggestions . . . . . . . . . . . . 328
  D. Sejal, K.G. Shailesh, V. Tejaswi, Dinesh Anvekar, K.R. Venugopal,
  S.S. Iyengar, and L.M. Patnaik

Offline Writer Identification in Tamil Using Bagged Classification Trees . . . . 342
  Sudarshan Babu

Applications of Data Mining

Data Analysis for Courses Registration . . . . . . . . . . . . . . . . . . . . . . . . . . . 357
  Nada Alzahrani, Rasha Alsulim, Nourah Alaseem, and Ghada Badr

Learning the Relationship Between Corporate Governance and Company
Performance Using Data Mining . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 368
  Darie Moldovan and Simona Mutu

A Bayesian Approach to Sparse Learning-to-Rank for Search Engine
Optimization. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 382
  Olga Krasotkina and Vadim Mottl

Data Driven Geometry for Learning . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 395
  Elizabeth P. Chou

Mining Educational Data to Predict Students’ Academic Performance . . . . 403
  Mona Al-Saleem, Norah Al-Kathiry, Sara Al-Osimi, and Ghada Badr

Patient-Specific Modeling of Medical Data. . . . . . . . . . . . . . . . . . . . . . . . . . 415
  Guilherme Alberto Sousa Ribeiro, Alexandre Cesar Muniz de Oliveira,
  Antonio Luiz S. Ferreira, Shyam Visweswaran, and Gregory F. Cooper

A Bayesian Approach to Sparse Cox Regression in High-Dimentional
Survival Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 425
  Olga Krasotkina and Vadim Mottl

Data Mining in System Biology, Drug Discovery, and Medicine

Automatic Cell Tracking and Kinetic Feature Description of Cell Paths
for Image Mining . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 441
  Petra Perner

Author Index . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 453
Machine Learning and Data Mining in Pattern Recognition
11th International Conference, MLDM 2015, Hamburg, Germany, July 20–21, 2015, Proceedings
Perner, P. (Ed.)
2015, IX, 454 p. 132 illus., Softcover
ISBN: 978-3-319-21023-0