Contents

SESSION: DATA PRE-PROCESSING AND SAMPLING

SIPPA for Biometric Data Reconstruction
Bon Sy

An Effective Entity Resolution Method
Chuan Zhao, Krishnamoorthy Sivakumar

Cluster Infiltration in Fraud and Money Laundering Prediction
Felipe Cardona, Gustavo Garcia, Juan Buritica

An Approach for Learning from Small and Unbalanced Data Sets Using Gaussian Noise During Artificial Neural Network Training
Icamaan Viegas, Paulo Adeodato

SESSION: WEB AND TEXT MINING

Enhanced Named Entity Extraction via Error-Driven Aggregation
Tracy Lemmond, Nathan Perry, Joseph Guensche, John Nitao, Ronald Glaser, Paul Kidwell, William Hanley

Finding Topic-specific Strings in Text Categorization and Opinion Mining Contexts
Remi Lavalley, Chloe Clavel, Marc El-Beze, Patrice Bellot

Using and Parsing Wikipedia Articles For Text Classification
Anjum Gupta, Craig Martell

A Hybrid Filter Car Recommender System
Jayalakshmi Jagadeesan, Robert Chun

Classifying Independent Medical Examination Reports using SOM Networks
Yinghao Huang, Naeem Seliya, Yi Murphey, Roy Friedenthal

Tuning Topical Query Classification in Large Search Engines to Understand Hidden User Information Need
Abdulbaghi Ghaderzadeh, Behrouz Minaie Bidgoli

A Framework for Social Spam Detection Based on Relational Bayes Classifier
Ahmet Aycan Atak, Sule Gunduz Oguducu
Detection of MMORPG Misconducts Based on Action Frequencies, Types and Time-Intervals 78
Ruck Thawonmas, Yoshitaka Kashifuji

SESSION: REAL-WORLD DATA MINING APPLICATIONS, CHALLENGES, AND PERSPECTIVES

Preliminary Results of Ranking Political Figures Using Naive Bayes Text Classification 85
James Ryder, Sen Zhang

A BI 2.0 Application Architecture for Healthcare Data Mining Services in the Cloud 92
Joseph M Woodside

Performance Evaluation of Intrusion Detection System Using Optimal Decision Tree SVM 98
Manju Bala, Namita Aggrawal, R. K. Agrawal

Learner Interaction Monitoring System (LiMS): Capturing the Behaviors of Online Learners and Evaluating Online Training Courses 106
Peter Sorenson, Leah P. Macfadyen

Elaboration of Applied Technologies for Modelling Connected Geomechanical, Geofiltration and Geodynamic Processes in Rock's Massif 112
Michael Zhuravkov, Aleg Kanavalau

A New Technique For Feature Selection And Cluster Center Initialization 119
Dharmveer Singh Rajoot, Pramod K. Singh, Mahua Bhattacharya

Examining Class Dropping with Data Mining 126
Kevin Daimi, Lu Wang

Predicting Protein Secondary Structure: A Recurrent Neural Network Approach 133
Mahmoud Abou-Nasr

Using Data Mining to Analyze Patient Discharge Data for an Urban Hospital 139
Xiaochun Jiang, Xiuli Qu, Lauren Davis

Application of Data Mining for Breast Cancer Survivability 145
Fatemeh Hosseinkhah, Hassan Ashktorab, Mohammad Owrang O.

Winning Baseball Through Data Mining 151
Gregory Swan, Anthony Scime

An Application of Clustering Techniques to Urban Studies 158
Zahra Ferdowsi, Raffaella Settimi, Daniela Raicu
Peeling Decision Tree Based on Fisher's Linear Discriminant Analysis
Jie Ouyang, Nilesh Patel

Hybrid PSO Algorithms for Dynamic Clustering
Andrea Villagra, Daniel Pandolfi, Guillermo Leguizamon

Revenue Generation in Hospital Foundations: Neural Network versus Regression Model Recommendations
Mary Malliaris, Maria Pappas

Real-time Automatic Building Identification
Robert Woodley, Warren Noll, Joseph Barker

Association Analysis of Semi-structured Data for Discrimination Discovery in Business
Binh Luong Thanh, Franco Turini

WIBE: Mining Frequent Closed Patterns Without Candidate Maintenance in Microarray Dataset
Miao Wang, Xuequn Shang, Jingni Diao, Zhanhuai Li

s-Tuple Inclusion - A New Method for Privacy Preserving Publication of Datasets
Ram Prasad Reddy S, Valli Kumari V, VSVN Raju K

Knowledge Discovery in the Virtual Social Network Due to Common Knowledge of Proverbs
Armando Mendes, Matthias Funk, Luis Cavique

Music Databases and Data Mining Approaches
Satyasaivani Bommakanti, Shashi M, Nagalakshmi V, Vikram D

A Student-Oriented University Ontology
Anemie Vorstermans, Pavol Tanuska, Werner Verschelde, Michal Kebisek

SESSION: PREDICTIVE MODELLING

SMO-Style Algorithms for Learning Using Privileged Information
Dmitry Pechyony, Rauf Izmailov, Akshay Vashist, Vladimir Vapnik

Learning Ensemble Models on Categorized Datasets
Xiong Deng, Yike Guo, Moustafa Ghanem

Solving Classification Problems with the NWEA-evolved ANNs
Kristina Davoian, Wolfram-M. Lippe
Bi-Directional Subspace Decomposition in Classification
Iryna Skrypnyk

Simple Method for Interpretation of High-Dimensional Nonlinear SVM Classification Models
Vladimir Cherkassky, Sauptik Dhar

Inference for Neural Network Predictive Models with Impulse Interventions
Nikolaos Kourentzes, Sven F. Crone

Ensembles of Locally Linear Models: Application to Bankruptcy Prediction
Laura Kainulainen, Qi Yu, Yoan Miche, Emil Eirola, Eric Séverin, Amaury Lendasse

An Approach for Time Series Forecasting by Simulating Stochastic Processes Through Time Lagged Feed-forward Neural Network
Cristian Rodriguez Rivero, Julian Pucheta, Josef Baumgartner, Hector D. Patino, B. Kuchen

Feature Selection Using a Novel Swarm Intelligence Algorithm with Rough Sets
Noorhaniza Wahid, Yuk Ying Chung, Wei-Chang Yeh, Guang Liu

A Border-Based Approach for Hiding Fuzzy Weighted Sensitive Itemsets
Durga Toshniwal, Mridula Verma

High Dimensional Data Classification by the PoSOM Algorithm
Lu Xu, T. W. S. Chow

Data Mining and Electronic Health Records: Selecting Optimal Clinical Treatments in Practice
Casey Bennett, Thomas Doub

Fat Tailed Distribution of Neural Networks Forecasting
Domingos S. P. Salazar, Maira de O. Santos, Adrian L. Arnaud, Paulo J. L. Adeodato

SESSION: SEGMENTATION, CLUSTERING, ASSOCIATION

Clustering Scatter Plots Using Data Depth Measures
Zhanpan Zhang, Xinping Cui, Daniel Jeske, Xiaoxiao Li, Jonathan Braun, James Borneman

Novel Significance Weighting Schemes for Collaborative Filtering: Generating Improved Recommendations in Sparse Environments
Mustansar Ali Ghazanfar, Adam Prugel-Bennett

SP&EPPF: A Novel Algorithm for Sequential Interval-based Pattern Mining
Jieh-Shan Yeh, Chia-Hsien Ting
Interactive Clustering of Proteomic Data - a Comparison between Self-Organizing Map and Neural Gas
*Vemund Jakobsen, Terje Kristensen*

Extracting Interesting Patterns of Hard CSPs
*Chendong Li, Jiayin Wang, Yichen Liu*

Application of Clustering in a Hybrid Global Optimization Algorithm
*Ting-Yu Chen, Jyun Hao Huang*

Extension of Hierarchical Information Acquisition by Association Rule Mining from Semi-Structured Data
*Ryosuke Saga, Chihiro Sugaya, Kodai Kitami*

Discovering Latent Healthy Nutritional Dietary Patterns with Association Rule Mining and Constraint Handling Rules
*Chendong Li, Yichen Liu*

A Novel Architecture for Hiding Sensitive Association Rules
*Muhammad Naeem, Sohail Asghar*

Finding Sequential Patterns from Temporal Datasets
*Fokrul Alom Mazarbhuia, Anjana Kakoti Mahanta, Hemanta K. Baruah*

Using Unsupervised Machine Learning Methods in High-Throughput Screening
*Chérif Mballo, Vladimir Makarenkov*

A Dimension-wise Algorithm for Multi-dimensional Data
*Yong Shi, Ali Tavakoli*

**SESSION: STOCHASTIC RESONANCE, CLASSIFICATION, AND INDEXING**

Stochastic Resonance Noise Benefits in Markov Chain Monte Carlo Density Estimation
*Brandon Franzke, Bart Kosko*

Document Indexing by Latent Dirichlet Allocation
*In-Chan Choi, Jae-Sung Lee*

Novel Hybrid Method for Sentiment Classification of Movie Reviews
*Zied Kechaou, Ali Wall, Mohamed Benammar, Adel M. Alimi*