Contents – Part I

Classification

Joint Classification with Heterogeneous Labels Using Random Walk with Dynamic Label Propagation ........................................... 3
Yongxin Liao, Shenxi Yuan, Jian Chen, Qingyao Wu, and Bin Li

Hybrid Sampling with Bagging for Class Imbalance Learning ............. 14
Yang Lu, Yiu-ming Cheung, and Yuan Yan Tang

Sparse Adaptive Multi-hyperplane Machine .................................. 27
Khanh Nguyen, Trung Le, Vu Nguyen, and Dinh Phung

Exploring Heterogeneous Product Networks for Discovering Collective Marketing Hyping Behavior ................................. 40
Qinzhe Zhang, Qin Zhang, Guodong Long, Peng Zhang, and Chengqi Zhang

Optimal Training and Efficient Model Selection for Parameterized Large Margin Learning ......................................................... 52
Yuxun Zhou, Jae Yeon Baek, Dan Li, and Costas J. Spanos

Locally Weighted Ensemble Learning for Regression .......................... 65
Man Yu, Zongxia Xie, Hong Shi, and Qinghua Hu

Reliable Confidence Predictions Using Conformal Prediction ............ 77
Henrik Linusson, Ulf Johansson, Henrik Boström, and Tuve Löfström

Grade Prediction with Course and Student Specific Models .................. 89
Agoritsa Polyzou and George Karypis

Flexible Transfer Learning Framework for Bayesian Optimisation ........ 102
Tinu Theckel Joy, Santu Rana, Sunil Kumar Gupta, and Svetla Venkatesh

A Simple Unlearning Framework for Online Learning Under Concept Drifts ................................................................. 115
Sheng-Chi You and Hsuan-Tien Lin

User-Guided Large Attributed Graph Clustering with Multiple Sparse Annotations .......................................................... 127
Jianping Cao, Senzhang Wang, Fengcai Qiao, Hui Wang, Feiyue Wang, and Philip S. Yu
<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early-Stage Event Prediction for Longitudinal Data</td>
<td>139</td>
</tr>
<tr>
<td>Mahtab J. Fard, Sanjay Chawla, and Chandan K. Reddy</td>
<td></td>
</tr>
<tr>
<td>Toxicity Prediction in Cancer Using Multiple Instance Learning in a Multi-task Framework</td>
<td>152</td>
</tr>
<tr>
<td>Cheng Li, Sunil Gupta, Santu Rana, Wei Luo, Svetha Venkatesh, David Ashely, and Dinh Phung</td>
<td></td>
</tr>
<tr>
<td>Shot Boundary Detection Using Multi-instance Incremental and Decremental One-Class Support Vector Machine</td>
<td>165</td>
</tr>
<tr>
<td>Hanhe Lin, Jeremiah D. Deng, and Brendon J. Woodford</td>
<td></td>
</tr>
<tr>
<td>Will I Win Your Favor? Predicting the Success of Altruistic Requests</td>
<td>177</td>
</tr>
<tr>
<td>Hsun-Ping Hsieh, Rui Yan, and Cheng-Te Li</td>
<td></td>
</tr>
</tbody>
</table>

**Feature Extraction and Pattern Mining**

<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsupervised and Semi-supervised Dimensionality Reduction with Self-Organizing Incremental Neural Network and Graph Similarity Constraints</td>
<td>191</td>
</tr>
<tr>
<td>Zhiyang Xiang, Zhu Xiao, Yourong Huang, Dong Wang, Bin Fu, and Wenjie Chen</td>
<td></td>
</tr>
<tr>
<td>Cross-View Feature Hashing for Image Retrieval</td>
<td>203</td>
</tr>
<tr>
<td>Wei Wu, Bin Li, Ling Chen, and Chengqi Zhang</td>
<td></td>
</tr>
<tr>
<td>Towards Automatic Generation of Metafeatures</td>
<td>215</td>
</tr>
<tr>
<td>Fábio Pinto, Carlos Soares, and João Mendes-Moreira</td>
<td></td>
</tr>
<tr>
<td>Hash Learning with Convolutional Neural Networks for Semantic Based Image Retrieval</td>
<td>227</td>
</tr>
<tr>
<td>Jinma Guo, Shifeng Zhang, and Jianmin Li</td>
<td></td>
</tr>
<tr>
<td>Bayesian Group Feature Selection for Support Vector Learning Machines</td>
<td>239</td>
</tr>
<tr>
<td>Changde Du, Changying Du, Shandian Zhe, Ali Luo, Qing He, and Guoping Long</td>
<td></td>
</tr>
<tr>
<td>Active Distance-Based Clustering Using K-Medoids</td>
<td>253</td>
</tr>
<tr>
<td>Amin Aghaeae, Mehrdad Ghadiri, and Mahdieh Soleymani Baghshah</td>
<td></td>
</tr>
<tr>
<td>Analyzing Similarities of Datasets Using a Pattern Set Kernel</td>
<td>265</td>
</tr>
<tr>
<td>A. Ibrahim, P.S. Sastry, and Shivakumar Sastry</td>
<td></td>
</tr>
<tr>
<td>Significant Pattern Mining with Confounding Variables</td>
<td>277</td>
</tr>
<tr>
<td>Aika Terada, David duVerle, and Koji Tsuda</td>
<td></td>
</tr>
</tbody>
</table>
Building Compact Lexicons for Cross-Domain SMT by Mining Near-Optimal Pattern Sets .......................................................... 290  
  Pankaj Singh, Ashish Kulkarni, Himanshu Ojha, Vishwajeet Kumar,  
  and Ganesh Ramakrishnan

Forest CERN: A New Decision Forest Building Technique ............... 304  
  Md. Nasim Adnan and Md. Zahidul Islam

Sparse Logistic Regression with Logical Features ............................ 316  
  Yuan Zou and Teemu Roos

A Nonlinear Label Compression and Transformation Method for Multi-label Classification Using Autoencoders ......................... 328  
  Jörg Wicker, Andrey Tyukin, and Stefan Kramer

Preconditioning an Artificial Neural Network Using Naive Bayes ....... 341  
  Nayyar A. Zaidi, François Petitjean, and Geoffrey I. Webb

OCEAN: Fast Discovery of High Utility Occupancy Itemsets ............ 354  
  Bilong Shen, Zhaoduo Wen, Ying Zhao, Dongliang Zhou,  
  and Weimin Zheng

Graph and Network Data

Leveraging Emotional Consistency for Semi-supervised Sentiment Classification ................................................................. 369  
  Minh Luan Nguyen

The Effect on Accuracy of Tweet Sample Size for Hashtag Segmentation Dictionary Construction .................................................. 382  
  Laurence A.F. Park and Glenn Stone

Social Identity Link Across Incomplete Social Information Sources  
Using Anchor Link Expansion ........................................................ 395  
  Yuxiang Zhang, Lulu Wang, Xiaoli Li, and Chunjing Xiao

Discovering the Network Backbone from Traffic Activity Data .......... 409  
  Sanjay Chawla, Kiran Garimella, Aristides Gionis, and Dominic Tsang

A Fast and Complete Enumeration of Pseudo-Cliques for Large Graphs . 423  
  Hongjie Zhai, Makoto Haraguchi, Yoshiaki Okubo, and Etsuji Tomita

Incorporating Heterogeneous Information for Mashup Discovery  
with Consistent Regularization ..................................................... 436  
  Yao Wan, Liang Chen, Qi Yu, Tingting Liang, and Jian Wu

Link Prediction in Schema-Rich Heterogeneous Information Network . 449  
  Xiaohuan Cao, Yuyan Zheng, Chuan Shi, Jingzhi Li, and Bin Wu
FastStep: Scalable Boolean Matrix Decomposition
Miguel Araujo, Pedro Ribeiro, and Christos Faloutsos

Applications

An Expert-in-the-loop Paradigm for Learning Medical Image Grouping
Xuan Guo, Qi Yu, Rui Li, Cecilia Ovesdotter Alm, Cara Calvelli, Pengcheng Shi, and Anne Haake

Predicting Post-operative Visual Acuity for LASIK Surgeries
Manish Gupta, Prashant Gupta, Pravin K. Vaddavalli, and Asra Fatima

LBMF: Log-Bilinear Matrix Factorization for Recommender Systems
Yunhui Guo, Xin Wang, and Congfu Xu

An Empirical Study on Hybrid Recommender System with Implicit Feedback
Sunhwan Lee, Anca Chandra, and Divyesh Jadav

Who Will Be Affected by Supermarket Health Programs? Tracking Customer Behavior Changes via Preference Modeling
Ling Luo, Bin Li, Shlomo Berkovsky, Irena Koprinska, and Fang Chen

TrafficWatch: Real-Time Traffic Incident Detection and Monitoring Using Social Media
Hoang Nguyen, Wei Liu, Paul Rivera, and Fang Chen

Automated Setting of Bus Schedule Coverage Using Unsupervised Machine Learning
Jihed Khiari, Luis Moreira-Matias, Vitor Cerqueira, and Oded Cats

Effective Local Metric Learning for Water Pipe Assessment
Moijgan Ghanavati, Raymond K. Wong, Fang Chen, Yang Wang, and Simon Fong

Classification with Quantification for Air Quality Monitoring
Sanad Al-Maskari, Eve Bélisle, Xue Li, Sébastien Le Digabel, Amin Nawahda, and Jiang Zhong

Predicting Unknown Interactions Between Known Drugs and Targets via Matrix Completion
Qing Liao, Naiyang Guan, Chengkun Wu, and Qian Zhang

Author Index
Contents – Part II

Spatiotemporal and Image Data

Denoising Time Series by Way of a Flexible Model for Phase Space Reconstruction ................................................................. 3
Minhazul Islam Sk and Arunava Banerjee

Distributed Sequential Pattern Mining in Large Scale Uncertain Databases . . . 17
Jiaqi Ge and Yuni Xia

DeepCare: A Deep Dynamic Memory Model for Predictive Medicine . . . . . 30
Trang Pham, Truyen Tran, Dinh Phung, and Svetha Venkatesh

Indoor Positioning System for Smart Homes Based on Decision Trees
and Passive RFID ........................................................................ 42
Frédéric Bergeron, Kevin Bouchard, Sébastien Gaboury, Sylvain Giroux, and Bruno Bouchard

Deep Feature Extraction from Trajectories for Transportation Mode Estimation ................................................................. 54
Yuki Endo, Hiroyuki Toda, Kyosuke Nishida, and Akihisa Kawanobe

Online Learning for Accurate Real-Time Map Matching ......................... 67
Biwei Liang, Tengjiao Wang, Shun Li, Wei Chen, Hongyan Li, and Kai Lei

Multi-hypergraph Incidence Consistent Sparse Coding for Image Data Clustering ................................................................. 79
Xiaodong Feng, Sen Wu, Wenjun Zhou, and Zhiwei Tang

Robust Multi-view Manifold Ranking for Image Retrieval ......................... 92
Jun Wu, Jianbo Yuan, and Jiebo Luo

Image Representation Optimization Based on Locally Aggregated
Descriptors ............................................................................. 104
Shijiang Chen, Guiguang Ding, Chenxiao Li, and Yuchen Guo

Reusing Extracted Knowledge in Genetic Programming to Solve Complex
Texture Image Classification Problems ....................................... 117
Muhammad Iqbal, Bing Xue, and Mengjie Zhang
Personal Credit Profiling via Latent User Behavior Dimensions on Social Media ................................................................. 130

Guangming Guo, Feida Zhu, Enhong Chen, Le Wu, Qi Liu, Yingling Liu, and Minghui Qiu

Linear Upper Confidence Bound Algorithm for Contextual Bandit Problem with Piled Rewards ........................................... 143

Kuan-Hao Huang and Hsuan-Tien Lin

Incremental Hierarchical Clustering of Stochastic Pattern-Based Symbolic Data ................................................................. 156

Xin Xu, Jiaheng Lu, and Wei Wang

Computing Hierarchical Summary of the Data Streams .................... 168

Zubair Shah, Abdun Naser Mahmood, and Michael Barlow

**Anomaly Detection and Clustering**

Unsupervised Parameter Estimation for One-Class Support Vector Machines ................................................................. 183

Zahra Ghafoori, Sutharshan Rajasegarar, Sarah M. Erfani, Shanika Karunasekera, and Christopher A. Leckie

Frequent Pattern Outlier Detection Without Exhaustive Mining ............. 196

Arnaud Giacometti and Arnaud Soulet

Ensembles of Interesting Subgroups for Discovering High Potential Employees ................................................................. 208

Girish Keshav Palshikar, Kuleshwar Sahu, and Rajiv Srivastava

Dynamic Grouped Mixture Models for Intermittent Multivariate Sensor Data ................................................................. 221

Naoya Takeishi, Takehisa Yairi, Naoki Nishimura, Yuta Nakajima, and Noboru Takata

Parallel Discord Discovery ................................................................. 233

Tian Huang, Yongxin Zhu, Yishu Mao, Xinyang Li, Mengyun Liu, Yafei Wu, Yajun Ha, and Gillian Dobbie

Dboost: A Fast Algorithm for DBSCAN-Based Clustering on High Dimensional Data .......................................................... 245

Yuxiao Zhang, Xiaorong Wang, Bingyang Li, Wei Chen, Tengjiao Wang, and Kai Lei

A Precise and Robust Clustering Approach Using Homophilic Degrees of Graph Kernel ................................................... 257

Haolin Yang, Deli Zhao, Lele Cao, and Fuchun Sun
Constraint Based Subspace Clustering for High Dimensional Uncertain Data ....................................................... 271
   Xianchao Zhang, Lu Gao, and Hong Yu

A Clustering-Based Framework for Incrementally Repairing Entity Resolution ............................................... 283
   Qing Wang, Jingyi Gao, and Peter Christen

Adaptive Seeding for Gaussian Mixture Models ...................... 296
   Johannes Blömer and Kathrin Bujna

A Greedy Algorithm to Construct L1 Graph with Ranked Dictionary ......... 309
   Shuchu Han and Hong Qin

Novel Models and Algorithms

A Rule Based Open Information Extraction Method Using Cascaded Finite-State Transducer .................................. 325
   Hailun Lin, Yuanzhuo Wang, Peng Zhang, Weiping Wang, Yinliang Yue, and Zheng Lin

Active Learning Based Entity Resolution Using Markov Logic .................. 338
   Jeffrey Fisher, Peter Christen, and Qing Wang

Modeling Adversarial Learning as Nested Stackelberg Games. ............ 350
   Yan Zhou and Murat Kantarcioglu

Fast and Semantic Measurements on Collaborative Tagging Quality ........ 363
   Yuqing Sun, Haiqi Sun, and Reynold Cheng

Matrices, Compression, Learning Curves: Formulation, and the GROUPNTEACH Algorithms .......................... 376
   Bryan Hooi, Hyun Ah Song, Evangelos Papalexakis, Rakesh Agrawal, and Christos Faloutsos

Privacy Aware K-Means Clustering with High Utility ..................... 388
   Thanh Dai Nguyen, Sunil Gupta, Santu Rana, and Svetla Venkatesh

Secure k-NN Query on Encrypted Cloud Data with Limited Key-Disclosure and Offline Data Owner .................. 401
   Youwen Zhu, Zhikuan Wang, and Yue Zhang

Hashing-Based Distributed Multi-party Blocking for Privacy-Preserving Record Linkage .................................. 415
   Thilina Ranbaduge, Dinusha Vatsalan, Peter Christen, and Vassilios Verykios
Text Mining and Recommender Systems

Enabling Hierarchical Dirichlet Processes to Work Better for Short Texts at Large Scale. .......................................................... 431
   Khai Mai, Sang Mai, Anh Nguyen, Ngo Van Linh, and Khoat Than

Query-Focused Multi-document Summarization Based on Concept Importance ................................................................. 443
   Hai-Tao Zheng, Ji-Min Guo, Yong Jiang, and Shu-Tao Xia

Mirror on the Wall: Finding Similar Questions with Deep Structured Topic Modeling ............................................................ 454
   Arpita Das, Manish Shrivastava, and Manoj Chinnakotla

An Efficient Dynamic Programming Algorithm for STR-IC-STR-IC-LCS Problem ........................................................................ 466
   Daxin Zhu, Yingjie Wu, and Xiaodong Wang

Efficient Page-Level Data Extraction via Schema Induction and Verification ........................................................................... 478
   Chia-Hui Chang, Tian-Sheng Chen, Ming-Chuan Chen, and Jhung-Li Ding

Transfer-Learning Based Model for Reciprocal Recommendation ................................................................. 491
   Chia-Hsin Ting, Hung-Yi Lo, and Shou-De Lin

Enhanced SVD for Collaborative Filtering ..................................................................................................................... 503
   Xin Guan, Chang-Tsun Li, and Yu Guan

Social Group Based Video Recommendation Addressing the Cold-Start Problem ............................................. 515
   Chunfeng Yang, Yipeng Zhou, Liang Chen, Xiaopeng Zhang, and Dah Ming Chiu

FeRoSA: A Faceted Recommendation System for Scientific Articles. ................................................................. 528
   Tanmoy Chakraborty, Amrith Krishna, Mayank Singh, Niloy Ganguly, Pawan Goyal, and Animesh Mukherjee

Dual Similarity Regularization for Recommendation ..................................................................................... 542
   Jing Zheng, Jian Liu, Chuan Shi, Fuzhen Zhuang, Jingzhi Li, and Bin Wu

Collaborative Deep Ranking: A Hybrid Pair-Wise Recommendation Algorithm with Implicit Feedback ............................... 555
   Haochao Ying, Liang Chen, Yuwen Xiong, and Jian Wu

Author Index ..................................................................................................................................................... 569
Advances in Knowledge Discovery and Data Mining
2016, XXIV, 608 p. 184 illus., Softcover
ISBN: 978-3-319-31752-6