

2014 International Conference on Data Science and Advanced Analytics

(DSAA 2014)

**Shanghai, China
30 October - 1 November 2014**



**IEEE Catalog Number: CFP14DSB-POD
ISBN: 978-1-4799-6982-1**

Proceedings of the 2014 International Conference on Data Science and Advanced Analytics (DSAA 2014)

TABLE OF CONTENTS

Welcome from DSAA 2014 Chairs	iii
DSAA/BESC 2014 Organizing Committee	v
DSAA/BESC 2014 Program Committee	vi
Keynote Speeches	ix
Panel Discussion	xvi
Special Sessions	xvii
Invited talks	xxiv
Tutorials	xxix
Technical Papers	xxx

DSAA – S1: Machine Learning

Session Chair: Geoff Webb

10:20 – 12:20, Oct. 30, 2014 Room: Regal Hall 1, 1/F

Assessing the Longevity of Online Videos: A New Insight of a Video's Quality	1
Qingbo Hu, Guan Wang, Philip S. Yu	
Minimizing Expected Loss for Risk-Avoiding Reinforcement Learning	11
Jung-Jung Yeh, Tsung-Ting Kuo, William Chen and Shou-De Lin	
Large-Scale Factorization of Type-Constrained Multi-Relational Data	18
Denis Krompass, Maximilian Nickel and Volker Tresp	
Pseudo Labels for Imbalanced Multi-Label Learning	25
Wenrong Zeng, Xuewen Chen and Hong Cheng	
Proactive Learning with Multiple Class-Sensitive Labels	32
Seungwhan Moon and Jaime Carbonel	
Itemset Approximation Using Constrained Binary Matrix Factorization	39
Seyed Hamid Mirisaei, Eric Gaussier and Alexandre Termier	
Community Detection in Social Networks: the Power of Ensemble Methods	46
Rushed Kanawati	

DSAA – S2: Retrieval, Query & Search

Session Chair: Wei Wang

10:20 – 12:20, Oct. 30, 2014 Room: Regal Hall 2, 1/F

Semi-randomized Hashing for Large Scale Data Retrieval	53
Haichuan Yang, Xiao Bai, Jun Zhou, Peng Ren, Jian Cheng and Lu Bai	
Storage Efficient Graph Search by Composite Dynamic-and-Static Indexing of a Single Network	59
Yan Xie and Philip Yu	
Indexing and Retrieval of Human Motion Data Based on a Growing Self-Organizing Map	66

Da-Cheng Yu and Wei-Guang Teng An Improved Dynamic Adaptive Multi-tree Search Anti-collision Algorithm Based on RFID	72
Min Shao, Xiao-Fang Jin and Li-Biao Jin Exploiting Mobility for Location Promotion in Location-based Social Networks	76
Wen-Yuan Zhu, Wen-Chih Peng and Ling-Jyh Chen Investigating Sample Selection Bias in the Relevance Feedback Algorithm of the Vector Space Model for Information Retrieval	83
Massimo Melucci	
DSAA – S3: Analytics Foundations	
Session Chair: Huan Liu	
14:40 – 16:40, Oct. 30, 2014 Room: Regal Hall 1, 1/F	
Learning a Proximity Measure to Complete a Community	90
Maximilien Danisch, Jean-Loup Guillaume and Benedicte Le Grand A Confidence-based Entity Resolution Approach with Incomplete Information	97
Qi Gu, Yan Zhang, Jian Cao and Guandong Xu and Alfredo Cuzzocrea Matrix Completion Based on Feature Vector and Function Approximation	104
Shiwei Ye, Yuan Sun and Yi Sun An Accuracy Enhancement Algorithm for Fingerprinting Method	110
Yuntian Brian Bai, Mani Williams, Falin Wu, Allison Kealy and Kefei Zhang	
DSAA – S4: Recommendation & Services	
Session Chair: Herve Martin	
14:40 – 16:40, Oct. 30, 2014 Room: Regal Hall 2, 1/F	
Probabilistic Category-based Location Recommendation Utilizing Temporal Influence and Geographical Influence	115
Dequan Zhou and Xin Wang Recommending Funding Collaborators with Scholar Social Networks	122
Juan Zhao, Kejun Dong and Jianjun Yu An Incremental Scheme for Large-scale Social-based Recommender Systems	128
Chia-Ling Hsiao, Zih-Syuan Wang and Wei-Guang Teng Diversification in News Recommendation for Privacy Concerned Users	135
Maunendra Sankar Desarkar and Neha Shinde Similarity Analysis of Service Descriptions for Efficient Web Service Discovery	142
Sowmya Kamath S and Ananthanarayana V.S	
DSAA – S5: Classification & Clustering	
Session Chair: Osmar Zaiane	
9:50 – 11:50, Oct. 31, 2014 Room: Regal Hall 1, 1/F	
Active Learning for Text Classification Using the LSI Subspace Signature Model	149
Weizhong Zhu and Robert B. Allen A Semisupervised Associative Classification Method for POS Tagging	156
Pratibha Rani, Vikram Pudi and Dipti Sharma	
Optimizing Specificity under Perfect Sensitivity for Medical Data Classification	163
Cho-Yi Hsiao, Hung-Yi Lo, Tu-Chun Yin and Shou-De Lin Interactive Correlation Clustering	170
Floris Geerts and Reuben Ndindi Rough Possibilistic Meta-Clustering of Retail Datasets	177

Asma Ammar, Zied Elouedi and Pawan Lingras

DSAA – S6: Infrastructure, Management & Privacy

Session Chair: Vincent Tseng

9:50 – 11:50, Oct. 31, 2014 Room: Regal Hall 2, 1/F

<u>Detecting Hidden Propagation Structure and Its Application to Analyzing Phishing</u>	184
Yang Liu and Mingyan Liu	
<u>Account Level Demand Estimation and Intelligence Framework</u>	191
Pranjal Mallick, Vikash Kumar Sharma, Parikshit Bhide and Mutha Reddy Mandapati	
<u>A New Pre-Pushing VoD Scheme in Hierarchical Network Environment</u>	196
Fei Long and Xingjun Wang	
<u>Usage Signatures Analysis - An Alternative Method for Preventing Fraud in e-Commerce Applications</u>	203
Gabriel Mota, Joana Fernandes and Orlando Belo	
<u>Exploring New Privacy Approaches in a Scalable Classification Framework</u>	209
M Saravanan, Mohamed Thoufeeq, S Akshaya and V.L Jayasre Manchari	

DSAA – S7: Influence Analysis

Session Chair: Huan Liu

15:30 – 17:30, Oct. 31, 2014 Room: Regal Hall 1, 1/F

<u>Efficient Analysis of Node Influence Based on SIR Model over Huge Complex Networks</u>	216
Masahiro Kimura, Kazumi Saito, Kouzou Ohara and Hiroshi Motoda	
<u>Social Influence-Aware Reverse Nearest Neighbor Search</u>	223
Hui-Ju Hung, De-Nian Yang and Wang-Chien Lee	
<u>Influence Maximization in a Social Network in the Presence of Multiple Influences and Acceptances</u>	230
Jun-Li Lu, Ling-Yin Wei and Mi-Yen Yeh	
<u>Diversified Ranking on Graphs from the Influence Maximization Viewpoint</u>	237
Li-Yen Kuo and Ming-Syan Chen	
<u>Mining Influence in Evolving Entities: A Study on Stock Market</u>	244
Chang Liao, Yinfei Huang, Xibin Shi and Xin Jin	

DSAA – S8: Data Science Applications

Session Chair: Xiang Li

15:30 – 17:30, Oct. 31, 2014 Room: Regal Hall 2, 1/F

<u>Sentiment Detection and Visualization of Chinese Micro-blog</u>	251
Zhitao Wang, Zhiwen Yu, Liming Chen and Bin Guo	
<u>Critical Class Sensitive Active Learning Method for Classification of Remote Sensing Imagery</u>	258
Lian-Zhi Huo, Zheng Zhang and Liang Tang	
<u>Content Specific Coverage Patterns for Banner Advertisement Placement</u>	263
Venkata Trinath Atmakuri, Gowtham Srinivas Parupalli and Krishna Reddy Polepalli	
<u>Appliance and State Recognition Using Hidden Markov Models</u>	270
Antonio Ridi, Christophe Gisler and Jean Hennebert	
<u>Exploring Technological Trends for Patent Evaluation</u>	277
Shuting Wang, Wang-Chien Lee, Zhen Lei, Xianliang Zhang and Yu-Hsuan Kuo	
<u>Crowdsourced Data Analytics: A Case Study of Predictive Modeling Competition</u>	284
Yukino Baba, Nozomi Nori, Shigeru Saito and Hisashi Kashima	

DSAA – S9: Complex Data Analysis

Session Chair: Osmar Zaiane

10:20 – 12:20, Nov. 1, 2014 Room: Regal Hall 1, 1/F

<u>A Model-Selection Framework for Concept-Drifting Data Streams</u>	290
--------------------------------------------------------------------------------------------	-----

Bo-Heng Chen and Kun-Ta Chuang A Probabilistic Condensed Representation of Data for Stream Mining	297
Michael Geilke, Andreas Karwath and Stefan Kramer Incrementally Mining Temporal Patterns in Interval-based Databases	304
Yi-Cheng Chen and Suh-Yin Lee The Purpose of Motion: Learning Activities from Individual Mobility Networks	312
Salvatore Rinzivillo, Lorenzo Gabrielli, Mirco Nanni, Luca Pappalardo, Fosca Giannotti and Dino Pedreschi On Selecting Feature-Value Pairs on Smart Phones for Activity Inferences	319
Gunarto Sindoro Njoo, Yu-Hsiang Peng, Wen-Chih Peng and Kuo-Wei Hsu	
DSAA – S10: Community and Network Analysis	
Session Chair: Xue-wen Chen	
10:20 – 12:20, Nov. 1, 2014 Room: Regal Hall 2, 1/F	
Learning Sparse and Scale-free Networks	326
Melih Aslan, Xuewen Chen and Hong Cheng Overlapping Community Detection in Social Network Based on Microblog User Model	333
Yajun Gu, Bofeng Zhang, Guobing Zou, Mingqing Huang and Keyuan Jiang Information Diffusion among Users on Facebook Fan Pages over Time: Its Impact on Movie Box Office	340
Wan-Hsin Tang, Mi-Yen Yeh and Anthony J.T. Lee Inferring Potential Users in Mobile Social Networks	347
Tsung-Hao Hsu, Chien-Cheng Chen, Meng-Fen Chiang, Kuo-Wei Hsu and Wen-Chih Peng	
DSAA – S11: Cloud Computing & Parallel Computing	
Session Chair: Wei Wang	
13:20 – 15:20, Nov. 1, 2014 Room: Dynasty Hall, 1/F	
A Token Authentication Solution for Hadoop Based on Kerberos Pre-Authentication	354
Kai Zheng and Weihua Jiang Hadoop Based Deep Packet Inspection System for Traffic Analysis of E-Business Websites	361
Jiangtao Luo, Yan Liang, Wei Gao and Junchao Yang A Data Reusing Strategy Based On Hive	367
Heng Xie, Mei Wang and Jiajin Le	
Dehazing Algorithm's High Performance and Parallel Computing for GF-1 Satellite Images	374
Changmiao Hu, Xiaojun Shan and Zheng Zhang General In-Situ Matrix Transposition Algorithm for Massively Parallel Environments	379
Marcin Gorawski and Michal Lorek	

DSAA Special Sessions

DSAA – SS1: Statistical and Mathematical Tools for Data Mining - SMTDM (1)	
Session Chair: Clausel Marianne	
10:20 – 12:20, Oct. 30, 2014 Room: North Hall 1, B/F	
Efficient Learning of General Bayesian Network Classifier by Local and Adaptive Search	385
Sein Minn, Shunkai Fu and Michel C. Desmarais A Naive Bayesian Classifier in Categorical Uncertain Data Streams	392
Jiaqi Ge, Yuni Xia and Jian Wang A New Set of Random Forests with Varying Dynamic Data Reduction and Voting Techniques	399

Hussein Mohsen, Hasan Kurban, Mark Jenne and Mehmet Dalkilic Centroid Training to Achieve Effective Text Classification	406
Libiao Zhang, Yuefeng Li, Yue Xu, Dian Tjondronegoro and Chao Sun Representing Sentence with Unfolding Recursive Autoencoders and Dynamic Average Pooling	413
Yin Hang, Zhang Chunhong, Zhu Yunkai and Ji Yang	
DSAA – SS2: Statistical and Mathematical Tools for Data Mining - SMTDM (2)	
Session Chair: Clausel Marianne	
14:40 – 16:40, Oct. 30, 2014 Room: North Hall 1, B/F	
A Novel Context-based Implicit Feature Extracting Method	420
Li Sun, Sheng Li, Jiyun Li and Jutao Lv	
Local Feature Based Dynamic Time Warping	425
Zheng Zhang, Liang Tang and Ping Tang	
Mobile User Stability Prediction with Random Forest Model	430
Danqin Wang and Xiaolong Zhang	
Detecting Stock Market Manipulation Using Supervised Learning Algorithms	435
Koosha Golmohammadi, Osmar Zaiane and David Diaz	
Recommending Missing Citations for Newly Granted Patents	442
Sooyoung Oh, Zhen Lei, Wang-Chien Lee and John Yen	
DSAA – SS3: Warehousing and Intelligent Analysis of Complex Network Big Data - WIACNBD (1)	
Session Chair: Elio Masciari	
9:50 – 11:50, Oct. 31, 2014 Room: North Hall 1, B/F	
FCA for Common Interest Communities discovering	449
Soumaya Guesmi, Chiraz Trabelsi and Chiraz Latiri	
Diversification Recommendation of Popular Articles in Micro-blog Scenario	456
Jianxing Zheng, Bofeng Zhang, Guobing Zou and Xiaodong Yue	
Analysis of Circadian Rhythms from Online Communities of Individuals with Affective Disorders	463
Bo Dao, Thin Nguyen, Dinh Phung and Svetha Venkatesh	
Link Prediction and Threads in Email Networks	470
Qinna Wang	
Mining Approximate Multi-Relational Patterns	477
Eirini Spyropoulou and Tijn De Bie	
A Detecting Community Method in Complex Networks with Fuzzy Clustering	484
Xiao Feng Wang, Gong Shen Liu and Jian Hua Li	
DSAA – SS4: Warehousing and Intelligent Analysis of Complex Network Big Data - WIACNBD (2)	
Session Chair: Elio Masciari	
15:30 – 17:30, Oct. 31, 2014 Room: North Hall 1, B/F	
Neural Network-Based Approaches for Predicting Query Response Times	491
Elif Ezgi Yusufoglu, Murat Ayyildiz and Ensar Gul	
User Preference Space Partition and Product Filters for Reverse Top-k Queries	498
Zong-Hua Yang and Hung-Yu Kao	
Finding Top-k Semantically Related Terms in Relational Keyword Search	505
Xiangfu Meng and Jingyu Shao	
Design Process for Big Data Warehouses	512
Francesco Di Tria, Ezio Lefons and Filippo Tangorra	
23-bit Metaknowledge Template Towards Big Data Knowledge Discovery and Management	519

Nima Bari, Roman Vichr, Kamran Kowsari and Simon Berkovich

DSAA – SS5: Environmental and Geo-spatial Data Analytics - EGSDA

Session Chair: Mathieu Roche

10:20 – 12:20, Nov. 1, 2014 Room: North Hall 1, B/F

<u>Location Semantics Prediction for Living Analytics by Mining Smartphone Data</u>	527
Chi-Min Huang, Jia-Ching Ying, Vincent S. Tseng and Zhi-Hua Zhou	
<u>Management of Complex Data Objects in Ship Designing Process</u>	534
Ruihan Bao and Hongming Cai	
<u>SAR Target Recognition Based on Deep Learning</u>	541
Sizhe Chen and Haipeng Wang	
<u>Mining Frequent Time Interval-based Event with Duration Patterns from Temporal Database</u>	548
Kuan-Ying Chen, Bijay Prasad Jaysawal, Jen-Wei Huang and Yong-Bin Wu	

DSAA – SS6: Bioinformatics, Biomedical, Health & Medical Analytics - BHMA

Session Chair: Vincent Tseng

13:20 – 15:20, Nov. 1, 2014 Room: North Hall 1, B/F

<u>Sharing Sensitive Medical Data Sets for Research Purposes - A Case Study</u>	555
Kalpana Singh, Jia Rong and Lynn Batten	
<u>Mobile-based Food Classification for Type-2 Diabetes Using Nutrient and Textual Features</u>	563
Yan Luo, Charles Ling and Shuang Ao	
<u>Individualized Arrhythmia Detection with ECG Signals from Wearable Devices</u>	570
Thanh-Binh Nguyen, Wei Luo, Terry Caelli, Svetha Venkatesh and Dinh Phung	

DSAA – SS7: Exploratory Computing - EC

Session Chair: Gabriella Pasi

15:40 – 17:40, Nov. 1, 2014 Room: North Hall 1, B/F

<u>Exploratory Computing: A Draft Manifesto</u>	577
Nicoletta Di Blas, Mirjana Mazuran, Paolo Paolini, Elisa Quintarelli and Letizia Tanca	
<u>Exploratory Portals. The Need for a New Generation</u>	581
Paolo Paolini and Nicoletta Di Blas	
<u>Exploring Emotions over Time within the Blogosphere</u>	587
Patrick Hennig, Philipp Berger, Lukas Pirl, Lukas Schulze and Christoph Meinel	
<u>StatsReduce in the Cloud for Approximate Analytics</u>	593
Michel De Rougemont	
<u>An Implementation of the Efficient Huge Amount of Pseudo-random Unique Numbers Generator and the Acceleration Analysis of Parallelization</u>	600
Yun-Te Lin, Yung-Hsiang Huang, Yi-Hao Hsiao, Yu-Jung Cheng, Jih-Sheng Chang, Sheng-Wen Wang, Fang-Pang Lin, and Chung-Ming Wang	
<u>Toward Robust Classification Using the Open Directory Project</u>	607
Jongwoo Ha, Jung-Hyun Lee, Won-Jun Jang, Yong-Ku Lee and SangKeun Lee	
<u>Generating, Communicating, Accessing and Analyzing Data in a Context-Aware Perspective</u>	613
Gabriella Pasi, Romano Fantacci	