

# **2009 IEEE Symposium on Computational Intelligence and Data Mining**

**(CIDM)**

**Nashville, Tennessee, USA  
30 March – 2 April 2009**



**IEEE Catalog Number: CFP09IDM-PRT  
ISBN: 978-1-4244-2765-9**

# TABLE OF CONTENTS

## **Session CIDM-1: CI/Probabilistic/Statistical and other methods I**

*Monday, March 30, 8:30AM-10:30AM, Room: Hermitage C, Chair: Barry Chen, USA*

### **Building Ultra-Low False Alarm Rate Support Vector Classifier Ensembles Using Random Subspaces [#9004] ..... 1**

Barry Chen, Tracy Lemmond and William Hanley  
Lawrence Livermore National Laboratory, United States

### **Collaborative Filtering with Fine-grained Trust Metric [#9017] ..... 9**

Su Chen, Tiejian Luo, Wei Liu and Yanxiang Xu  
Graduate University of Chinese Academy of Sciences, China

### **HybridSOM: A Generic Rule Extraction Framework for Self-Organizing Feature Maps [#9021] ..... 17**

Willem S. van Heerden and Andries P. Engelbrecht  
University of Pretoria, South Africa

### **Assessing the Influence Probability between Objects: A Random Walker Approach [#9022] ..... 25**

Pei Li, Zhixu Li, Hongyan Liu, Jun He and Xiaoyong Du  
Renmin University of China, China; Tsinghua University, China

## **Session CIDM-2: CI/Probabilistic/Statistical and other methods II**

*Monday, March 30, 11:00AM-1:00PM, Room: Hermitage C, Chair: Oliver Schulte, Simon Fraser University, USA*

### **Intelligent Feature Extraction and Knowledge Mining by Multivariate Analyses [#9023] ..... 33**

Yisong Chen and Hong Cui  
Key Laboratory of Machine Perception (Ministry of Education), Peking University, China; Unis Nast Developing Co. Ltd, Tshinghua Unisplendour Group, China

### **Efficient Model Selection for Support Vector Machine with Gaussian Kernel Function [#9027] ..... 40**

Yaohua Tang, Weimin Guo and Jinghuai Gao  
Henan Electric Power Research Institute, China; School of Electronic and Information Engineering, Xi'an Jiaotong University, China

**A Fault Tolerant Peer-to-Peer Distributed EM Algorithm [#9030] .....46**

Behrooz Safarinejadian, Mohammad Bagher Menhaj and Mehdi Karrari  
Amirkabir University of Technology, Iran

**A New Hybrid Method for Bayesian Network Learning With Dependency Constraints  
[#9036] .....53**

Oliver Schulte, Gustavo Frigo, Russell Greiner and Hassan Khosravi  
School of Computing Science, Simon Fraser University, Canada; Department of Computing Science, University of Alberta, Canada

**A Pillar Algorithm for K-Means Optimization by Distance Maximization for Initial Centroid Designation  
[#9038] .....61**

Ali Ridho Barakbah and Yasushi Kiyoki  
Graduate School of Media and Governance, Keio University, Japan, Japan; Faculty of Environmental Information, Keio University, Japan, Japan

**Session CIDM-3: CI/Probabilistic/Statistical and other methods III**

*Monday, March 30, 2:00PM-4:00PM, Room: Hermitage C, Chair: Giuseppe Di Fatta, University of Reading, UK*

**An Adaptive Ensemble Classifier for Concept Drifting Stream [#9043] .....69**

Dengyuan Wu, Ying Liu, Ge Gao, Zhendong Mao, Weishan Ma and Tao He  
Institute of Computing Technology, Chinese Academy of Sciences, China; Fictitious Economy and Data Science Research Center, Graduate University of Chinese Academy of Sciences, China; University of Virginia, China

**Missing Traffic Flow Data Prediction using Least Squares Support Vector Machines in Urban Arterial Streets [#9045] .....76**

Yang Zhang and Yuncai Liu  
Shanghai Jiao Tong University, China

**Dynamic classifier selection using clustering for spam detection [#9053] .....84**

Mehrnoush Famil saeedian and Hamid Beigy  
Sharif University of Technology, Iran

**Skill Rating by Bayesian Inference [#9060] .....89**

Giuseppe Di Fatta, Guy Haworth and Kenneth Regan  
The University of Reading, United Kingdom; University at Buffalo, The State University of New York, United States

**Clustering-based Activity Classification with a Wrist-worn Accelerometer Using Basic Features  
[#9066] .....95**

Pekka Siirtola, Perttu Laurinen, Eija Haapalainen, Juha Roning and Hannu Kinnunen  
University of Oulu, Finland; Polar Electro Oy, Finland

**Session CIDM-4: Data Understanding, rule extraction, logical models I**

*Monday, March 30, 4:30PM-6:30PM, Room: Hermitage C, Chair: Kyoko Fukuda, University of Canterbury, NZ*

**Decision Trees as Information Source for Attribute Selection [#9013] ..... 101**

Kyoko Fukuda and Brent Martin  
University of Canterbury, New Zealand

**An Enhanced Data Mining Life Cycle [#9018] ..... 109**

Markus Hofmann and Brendan Tierney  
Institute of Technology Blanchardstown, Ireland; Dublin Institute of Technology, Ireland

**Local Mining of Association Rules with Rule Schemas [#9047] ..... 118**

Andrei Olaru, Claudia Marinica and Fabrice Guillet  
Departement of Computer Science, University Politehnica of Bucharest, Romania; LINA - Ecole polytechnique de l'Universite de Nantes, France

**Extracting Hot Spots of Basic and Complex Topics From Time Stamped Documents  
[#9089] ..... 125**

Wei Chen and Parvathi Chundi  
University of Nebraska at Omaha, United States

**Session CIDM-5: Applications to biomedicine, e-commerce, engineering, etc I**

*Tuesday, March 31, 8:30AM-10:30AM, Room: Hermitage C, Chair: Anna L Buczak, JHU APL, USA*

**Hierarchically Classifying Documents with Multiple Labels  
[#9007] ..... 133**

Andrew Mayne and Russell Perry  
Hewlett Packard Labs, United Kingdom

**Determining the Strength of the Propensities of a Blog Network [#9024] ..... 140**

Seok-Ho Yoon, Sang-Wook Kim and Sunju Park  
Hanyang University, Korea (South); Yonsei University, Korea (South)

**Mining Electronic Medical Records for Patient Care Patterns [#9076] ..... 146**

Anna Buczak, Linda Moniz, Brian Feighner and Joseph Lombardo  
JHU APL, United States

**Gender Identification from E-mails [#9081] ..... 154**

Na Cheng, Xiaoling Chen, Rajarathnam Chandramouli and K.P. Subbalakshmi  
Stevens Institute of Technology, United States

**Session CIDM-6: Applications to biomedicine, e-commerce, engineering, etc II**  
*Tuesday, March 31, 11:00AM-1:00PM, Room: Hermitage C, Chair: Sangoh Jeong, Samsung Information Systems America, USA*

**Generalized Random Rotation Perturbation for Vertically Partitioned Data Sets**  
**[#9015] ..... 159**

Zhenmin Lin, Jie Wang, Lian Liu and Jun Zhang  
University of Kentucky, United States; Univesity of Kentucky, United States

**Mining for Insider Threats in Business Transactions and Processes** **[#9046] ..... 163**

William Eberle and Lawrence Holder  
Tennessee Technological University, United States; Washington State University, United States

**Density-Based Clustering of Polygons** **[#9086] ..... 171**

Deepti Joshi, Ashok Samal and Leen-Kiat Soh  
University of Nebraska-Lincoln, United States

**Non-collaborative Interest Mining for Personal Devices** **[#9091] ..... 179**

Sangoh Jeong, Doreen Cheng, Henry Song and Swaroop Kalasapur  
Samsung Information Systems America, United States

**Session CIDM-7: Applications to biomedicine, e-commerce, engineering, etc III**  
*Tuesday, March 31, 2:00PM-4:00PM, Room: Hermitage C, Chair: Alessando Sperduti, University of Padova, Italy*

**An Empirical Study of Bagging and Boosting Ensembles for Identifying Faulty Classes in Object-Oriented Software** **[#9059] ..... 187**

Hamoud Aljamaan and Mahmoud Elish  
King Fahd University of Petroleum and Minerals, Saudi Arabia

**Automatic Analysis of Eye Tracking Data for Medical Diagnosis** **[#9061] ..... 195**

Filippo Galgani, Yiwen Sun, Pier Luca Lanzi and Jason Leigh  
Politecnico di Milano, Italy; University of Illinois at Chicago, United States

**Application of the Preference Learning Model to a Human Resources Selection Task**  
**[#9062] ..... 203**

Fabio Aiolli, Michele De Filippo and Alessandro Sperduti  
University of Padova, Italy

**Practical Fuzzy Decision Trees** **[#9069] ..... 211**

Na'el Abu-halaweh and Robert Harrison  
Georgia State University, United States

**Session CIDM-8: Data Understanding, rule extraction, logical models II**

*Tuesday, March 31, 4:30PM-6:30PM, Room: Hermitage C, Chair: Janusz Kacprzyk, Polish Academy of Sciences, Poland*

**Data Mining via Protoform Based Linguistic Summaries: Some Possible Relations to Natural Language Generation [#9064] .....217**

Janusz Kacprzyk and Slawomir Zadrozny  
Systems Research Institute, Polish Academy of Sciences, Poland

**Handling Continuous Attributes in Ant Colony Classification Algorithms [#9072] .....225**

Fernando Otero, Alex Freitas and Colin Johnson  
University of Kent, United Kingdom

**A Novel Data Clustering Algorithm Based on Electrostatic Field Concepts [#9077] .....232**

Masoumeh Kalantari Khandani, Parvaneh Saeedi, Yaser P. Fallah and Mehdi K. Khandani  
Simon Fraser University, Canada; University of California Berkeley, United States; University of Maryland, College Park, United States

**Evolving Decision Trees Using Oracle Guides [#9079] .....238**

Ulf Johansson and Lars Niklasson  
University of Boras, Sweden; University of Skovde, Sweden

**Ensemble Member Selection Using Multi-Objective Optimization [#9082] .....245**

Tuve Lofstrom, Ulf Johansson and Henrik Bostrom  
University of Boras, Sweden; University of Skovde, Sweden

**Session CIDM-9: Text, graph and web mining**

*Wednesday, April 1, 8:30AM-10:30AM, Room: Hermitage C, Chair: Vladimir Bartik, Brno University of Technology, Czech Republic*

**Association Based Classification for Relational Data and Its Use in Web Mining [#9063] .....252**

Vladimir Bartik  
Brno University of Technology, Czech Republic

**Empirical Comparison of Graph Classification Algorithms [#9071] .....259**

Nikhil Ketkar, Lawrence Holder and Diane Cook  
Washington State University, United States

**Faster Computation of the Direct Product Kernel for Graph Classification [#9084] .....267**

Nikhil Ketkar, Lawrence Holder and Diane Cook  
Washington State University, United States

**Dependable Performance Analysis for Fuzzy Clustering of Web Usage Data [#9090] .....275**

Amir Ketata, Sudhir Mudur and Nematollaah Shiri

Department of Computer Science and Software Engineering, Concordia University, Montreal, Quebec, Canada

**Discovering Relational Knowledge from Two Disjoint Sets of Literatures Using Inductive Logic Programming [#9095] .....283**

Supphachai Thaicharoen, Tom Altman, Katheleen Gardiner and Krzysztof Cios

University of Colorado Denver, United States; Virginia Commonwealth University, United States

**Session CIDM-10: Mining spatial and spatial-temporal data**

*Wednesday, April 1, 11:00AM-1:00PM, Room: Hermitage C, Chair: Rachsuda Jiamthaphaksin, University of Houston*

**LD-BSCA: A Local-Density Based Spatial Clustering Algorithm [#9031] .....291**

Guiyi Wei and Haiping Liu

College of Computer Science and Information Engineering, Zhejiang Gongshang University, China

**Trajectory Clustering in Road Network Environment [#9050] .....299**

Jung-Im Won, Sang-Wook Kim, Ji-Haeng Baek and Junghoon Lee

Hanyang University, Korea (South); Jeju National University, Korea (South)

**An Architecture and Algorithms for Multi-Run Clustering [#9051] .....306**

Rachsuda Jiamthaphaksin, Christoph Eick and Vadeerat Rinsurongkawong

Computer Science Department, University of Houston, United States

**MWASP: Multiple-Width Approximate Sequential Patterns [#9068] .....314**

Kelly Kingchi Yip and David Nembhard

Department of Industrial and Manufacturing Engineering, Pennsylvania State University, United States

**Fuzzy P-Mode Prototypes: A Generalization of Frequency-Based Cluster Prototypes for Clustering Categorical Objects [#9094] .....320**

Mahnhoon Lee

Thompson Rivers University, Canada

**Session CIDM-11: Mining of very large datasets, scalability**

*Wednesday, April 1, 2:00PM-4:00PM, Room: Hermitage C, Chair: P. Krishna Reddy, International Institute of Information Technology - Hyderabad, India*

**Diversity Analysis on Imbalanced Data Sets by Using Ensemble Models [#9014] .....324**

Shuo Wang and Xin Yao

School of Computer Science, University of Birmingham, United Kingdom

<b>Large-scale Attribute Selection using Wrappers [#9025]</b> .....	<b>332</b>
Martin Guetlein, Eibe Frank, Mark Hall and Andreas Karwath Albert-Ludwigs-Universitaet Freiburg, Germany; University of Waikato, New Zealand; Pentaho Corporation, United States	
<b>An Improved Multiple Minimum Support Based Approach to Mine Rare Association Rules [#9074]</b> .....	<b>340</b>
R. Uday Kiran and P. Krishna Reddy International Institute of Information Technology - Hyderabad, India	
<b>Data Mining with Ensembles of Fuzzy Decision Trees [#9075]</b> .....	<b>348</b>
Christophe Marsala University Pierre et Marie Curie - Paris 6, France	
<b>The Locality of RBF-SVM for Incremental Learning [#9087]</b> .....	<b>355</b>
Wael Emara and Mehmed Kantardzic University of Louisville, United States	
<b>Session CIDM-12: Feature extraction, selection, aggregation, construction</b> <i>Wednesday, April 1, 4:30PM-6:30PM, Room: Hermitage C, Chair: Christophe Marsala, Universite Pierre et Marie Curie, France</i>	
<b>Parametric Subspace Analysis for dimensionality reduction and classification [#9033]</b> .....	<b>363</b>
Nhat Vo, Duc Vo, Bill Moran, and Subhash Challa University of Melbourne, Australia; University of Technology, Sydney, Australia	
<b>Incremental Wrapper-based Subset Selection with Replacement: an advantageous alternative to sequential forward selection [#9052]</b> .....	<b>367</b>
Pablo Bermejo, Jose Antonio Gamez and Jose Miguel Puerta Universidad de Castilla-La Mancha, Spain	
<b>Analysis and Visualization of the Geographical Distribution of Atlantic Forest Bromeliads Species [#9073]</b> .....	<b>375</b>
Stainam Brandao, Wagner Silva, Luis Silva, Vladimir Fagundes and Carlos de Mello COPPE/UFRJ - Computer Science Department, Graduate School of Engineering, Federal University of Rio de Janeiro, Brazil	
<b>Maintaining Only Frequent Itemsets to Mine Approximate Frequent Itemsets over Online Data Streams [#9085]</b> .....	<b>381</b>
Yongyan Wang, Kun Li and Hongan Wang Institute of Software, Chinese Academy of Sciences, China	



**Session CIDM-13: Mining of signals and data stream**

*Thursday, April 2, 8:30AM-10:30AM, Room: Hermitage C, Chair: Xin Feng, Marquette University, USA*

**Regularized Extreme Learning Machine [#9020] .....389**

Wanyu Deng, Qinghua Zheng, Lin Chen

Xi'an Jiaotong University, China; Institute of Post & Telecommunication of Xi'an, China

**Relevance Weighting of Multi-Term Queries for Vector Space Model [#9048] .....396**

Louis Wang

International School of Minnesota, United States

**An approximation algorithm for finding skeletal points for density based clustering approaches [#9093] .....403**

Soheil Hassas Yeganeh, Jafar Habibi, Hassan Abolhassani, Mahdi Abbaspour Tehrani and Jamshid Esmaelnezhad

Sharif University of Technology, Iran

**Mining Multiple Temporal Patterns of Complex Dynamic Data Systems [#9099] .....411**

Xin Feng and Odilon Senyana

Marquette University, United States

**AUTHOR INDEX**