

**Proceedings**

**IEEE International Conference  
on Data Mining Workshops  
ICDM Workshops 2008**

**15-19 December 2008  
Pisa, Italy**



Los Alamitos, California  
Washington • Tokyo



**2008 IEEE International  
Conference on Data Mining  
Workshops**

**ICDM Workshops 2008**

**Table of Contents**

Preface

Conference Organizers

RIKD 2008 Message and Committee

DDDM 2008 Message and Committee

MCD 2008 Message and Committee

SSTDm 2008 Message and Committees

HPDM 2008 Message and Committees

FDM 2008 Message and Committee

SADM 2008 Message and Committee

DMDM 2008 Message and Committee

VM 2008 Message and Committee

ADN 2008 Message and Committee

---

**Workshop on Reliability Issues in Knowledge Discovery  
(RIKD 2008)**

Comparing Accuracies of Rule Evaluation Models to Determine Human Criteria on Evaluated Rule Sets .....	1
<i>Hidenao Abe and Shusaku Tsumoto</i>	
Comparing Reliability of Association Rules and OLAP Statistical Tests .....	8
<i>Zhibo Chen, Carlos Ordonez, and Kai Zhao</i>	
Research on Methodology of Classification Mining for Tumor Markers .....	18
<i>Wei Jiang, Min Yao, and Jiekai Yu</i>	
Wavelet-Based Data Perturbation for Simultaneous Privacy-Preserving and Statistics-Preserving .....	27
<i>Lian Liu, Jie Wang, and Jun Zhang</i>	
Online Reliability Estimates for Individual Predictions in Data Streams .....	36
<i>Pedro Pereira Rodrigues, Joao Gama, and Zoran Bosnić</i>	

A Comparative Study of Data Sampling and Cost Sensitive Learning .....	46
<i>Chris Seiffert, Taghi M. Khoshgoftaar, Jason Van Hulse, and Amri Napolitano</i>	
Region Classification with Decision Trees .....	53
<i>J. van Prehn and E. N. Smirnov</i>	
A Study on the Reliability of Case-Based Reasoning Systems .....	60
<i>Ke Wang, James N. K. Liu, and Wei-min Ma</i>	
A Case Study on Classification Reliability .....	69
<i>Honghua Dai</i>	

## **Workshop on Domain Driven Data Mining (DDDM 2008)**

Domain Driven Data Mining (D3M) .....	74
<i>Longbing Cao</i>	
Parameter Tuning for Differential Mining of String Patterns .....	77
<i>Jérémy Besson, Christophe Rigotti, Ieva Mitašiunaite, and Jean-François Boulicaut</i>	
Behavior Informatics and Analytics: Let Behavior Talk .....	87
<i>Longbing Cao</i>	
Scoring Models for Insurance Risk Sharing Pool Optimization .....	97
<i>Nicolas Chapados, Charles Dugas, Pascal Vincent, and Réjean Ducharme</i>	
TransRank: A Novel Algorithm for Transfer of Rank Learning .....	106
<i>Depin Chen, Jun Yan, Gang Wang, Yan Xiong, Weiguo Fan, and Zheng Chen</i>	
One-Class Classification of Text Streams with Concept Drift .....	116
<i>Yang Zhang, Xue Li, and Maria Orłowska</i>	
Post-Processing of Discovered Association Rules Using Ontologies .....	126
<i>Claudia Marinica, Fabrice Guillet, and Henri Briand</i>	
Food Sales Prediction: "If Only It Knew What We Know" .....	134
<i>Patrick Meulstee and Mykola Pechenizkiy</i>	
Discovering Implicit Redundancies in Network Communications for Detecting Inconsistent Values .....	144
<i>Bogdan Tomoyuki Nassu, Takashi Nanya, and Hiroshi Nakamura</i>	
Actionable Knowledge Discovery for Threats Intelligence Support Using a Multi-dimensional Data Mining Methodology .....	154
<i>Olivier Thonnard and Marc Dacier</i>	
Identification of Causal Variables for Building Energy Fault Detection by Semi-supervised LDA and Decision Boundary Analysis .....	164
<i>Keigo Yoshida, Minoru Inui, Takehisa Yairi, Kazuo Machida, Masaki Shioya, and Yoshio Masukawa</i>	

## **Workshop on Mining Complex Data (MCD 2008)**

Unifying Unknown Nodes in the Internet Graph Using Semisupervised Spectral Clustering .....	174
<i>Anat Almog, Jacob Goldberger, and Yuval Shavitt</i>	
Hierarchical Text Categorization in a Transductive Setting .....	184
<i>Michelangelo Ceci</i>	

Towards Combining Structured Pattern Mining and Graph Kernels .....	192
<i>Fabrizio Costa and Björn Bringmann</i>	
k-Nearest Neighbor Classification on First-Order Logic Descriptions .....	202
<i>S. Ferilli, M. Biba, T. M. A. Basile, N. Di Mauro, and F. Esposito</i>	
Semi-supervised Collaborative Clustering with Partial Background Knowledge .....	211
<i>Germain Forestier, Cédric Wemmert, and Pierre Gançarski</i>	
Mining Temporal Patterns with Quantitative Intervals .....	218
<i>Thomas Guyet and René Quiniou</i>	
Plant Protein Localization Using Discriminative and Frequent Partition-Based Subsequences .....	228
<i>S. Vahid Jazayeri and Osmar R. Zaiane</i>	
Clustering Events on Streams Using Complex Context Information .....	238
<i>YongChul Kwon, Wing Yee Lee, Magdalena Balazinska, and Guiping Xu</i>	
Discovering Triggering Events from Longitudinal Data .....	248
<i>Corrado Loglisci and Donato Malerba</i>	
Extension of Partitional Clustering Methods for Handling Mixed Data .....	257
<i>Yosr Naija, Salem Chakhar, Kaouther Blibech, and Riadh Robbana</i>	
Word Sense Discovery for Web Information Retrieval .....	267
<i>Tomasz Nykiel and Henryk Rybinski</i>	
Mining Correlated Pairs of Patterns in Multidimensional Structured Databases .....	275
<i>Tomonobu Ozaki and Takenao Ohkawa</i>	
Association Action Rules .....	283
<i>Zbigniew W. Raś, Agnieszka Dardzińska, Li-Shiang Tsay, and Hanna Wasyluk</i>	
Multiple-Instance Regression with Structured Data .....	291
<i>Kiri L. Wagstaff, Terran Lane, and Alex Roper</i>	
Discovery of Internal and External Hyperclique Patterns in Complex Graph Databases .....	301
<i>Tsubasa Yamamoto, Tomonobu Ozaki, and Takenao Ohkawa</i>	
Harmonic Blind Sound Source Isolation Enhanced by Spectrum Clustering .....	310
<i>Xin Zhang, Wenxin Jiang, and Zbigniew W. Ras</i>	

## **Workshop on Spatial and Spatio-temporal Data Mining (SSTDM 2008)**

A Spatio-temporal Simulation Model for Movement Data Generation .....	320
<i>D. Alberg, M. Last, and S. Elnekave</i>	
Geographic Knowledge Discovery in INGENS: An Inductive Database Perspective .....	326
<i>Annalisa Appice, Anna Ciampi, Antonietta Lanza, Donato Malerba, Antonella Rapolla, and Luisa Vetturi</i>	
Enriching Spatial OLAP with Map Generalization: a Conceptual Multidimensional Model .....	332
<i>Sandro Bimonte, Jérôme Gensel, and Michela Bertolotto</i>	
Risk Assessment of Atmospheric Hazard Releases Using K-Means Clustering .....	342
<i>Guido Cervone, Pasquale Franzese, Yasmin Ezber, and Zafer Boybeyi</i>	

A Robust Graph-Based Algorithm for Detection and Characterization of Anomalies in Noisy Multivariate Time Series .....	349
<i>Haibin Cheng, Pang-Ning Tan, Christopher Potter, and Steven Klooster</i>	
Kernels for the Investigation of Localized Spatiotemporal Transitions of Drought with Support Vector Machines .....	359
<i>Matthew W. Collier and Amy McGovern</i>	
Standards-Based Coastal Sensor Web .....	369
<i>Surya S. Durbha, Roger L. King, Nicolas H. Younan, Santosh A. Rajender, and Shruthi Bheemireddy</i>	
Detection and Exploration of Outlier Regions in Sensor Data Streams .....	375
<i>Conny Franke and Michael Gertz</i>	
Data Mining for Climate Change and Impacts .....	385
<i>Auroop R Ganguly and Karsten Steinhaeuser</i>	
RE-SPaM: Using Regular Expressions for Sequential Pattern Mining in Trajectory Databases .....	395
<i>Leticia I. Gómez and Alejandro A. Vaisman</i>	
Incremental Maintenance of Discovered Spatial Colocation Patterns .....	399
<i>Jiangfeng He, Qinming He, Feng Qian, and Qi Chen</i>	
Speeding up Array Query Processing by Just-In-Time Compilation .....	408
<i>Constantin Jucovschi, Peter Baumann, and Sorin Stancu-Mara</i>	
Extraction of Discriminative Features from Hyperspectral Data .....	414
<i>Habil Kalkan and Yasemin Yardimci</i>	
Scalable Sparse Bayesian Network Learning for Spatial Applications .....	420
<i>Thomas Liebig, Christine Körner, and Michael May</i>	
High Granularity Remote Sensing and Crop Production over Space and Time: NDVI over the Growing Season and Prediction of Cotton Yields at the Farm Field Level in Texas .....	426
<i>Bert Little, Michael Schucking, Brandon Gartrell, Bing Chen, Kenton Ross, and Rodney McKellip</i>	
Web Query Prediction by Unifying Model .....	436
<i>Ning Liu, Jun Yan, Shuicheng Yan, Weiguo Fan, and Zheng Chen</i>	
A Vector-Geometry Based Spatial kNN-Algorithm for Traffic Frequency Predictions .....	442
<i>Michael May, Dirk Hecker, Christine Körner, Simon Scheider, and Daniel Schulz</i>	
Detecting and Tracking Spatio-temporal Clusters with Adaptive History Filtering .....	448
<i>James Rosswog and Kanad Ghose</i>	
A Semi-supervised Learning Algorithm for Recognizing Sub-classes .....	458
<i>Ranga Raju Vatsavai, Shashi Shekhar, and Budhendra Bhaduri</i>	
<b>Workshop on High Performance Data Mining (HPDM 2008)</b>	
Mining Unstructured Text at Gigabyte per Second Speeds .....	468
<i>Alan Ratner</i>	
Chi-Square Test Based Decision Trees Induction in Distributed Environment .....	477
<i>Jie Ouyang, Nilesh Patel, and Ishwar K. Sethi</i>	

Distributed Data Mining Models as Services on the Grid .....	486
<i>Eugenio Cesario and Domenico Talia</i>	
Service Oriented KDD: A Framework for Grid Data Mining Workflows .....	496
<i>Marco Lackovic, Domenico Talia, and Paolo Trunfio</i>	
Exploiting Graphic Card Processor Technology to Accelerate Data Mining Queries in SAP NetWeaver BIA .....	506
<i>Christoph Weyerhaeuser, Tobias Mindnich, Franz Faerber, and Wolfgang Lehner</i>	
Stream-Close: Fast Mining of Closed Frequent Itemsets in High Speed Data Streams .....	516
<i>Ranganath B. N. and M. Narasimha Murty</i>	
Parallel Hierarchical Clustering on Market Basket Data .....	526
<i>Baoying Wang, Qin Ding, and Imad Rahal</i>	
Efficient Distance Computation Using SQL Queries and UDFs .....	533
<i>Sasi K. Pitchaimalai, Carlos Ordonez, and Carlos Garcia-Alvarado</i>	
Distributed Linear Programming and Resource Management for Data Mining in Distributed Environments .....	543
<i>Haimonti Dutta and Hillol Kargupta</i>	
Investigation of Various Matrix Factorization Methods for Large Recommender Systems .....	553
<i>Gábor Takács, István Pilászy, Bottyán Németh, and Domonkos Tikk</i>	

## **Workshop on Foundations of Data Mining (FDM 2008)**

Co-training by Committee: A New Semi-supervised Learning Framework .....	563
<i>Mohamed Farouk Abdel Hady and Friedhelm Schwenker</i>	
Rules Extraction from Multiple Decisions Ordered Information Tables .....	573
<i>Bin Shen, Min Yao, and Zhaohui Wu</i>	
An Efficient Sequential Pattern Mining Algorithm Based on the 2-Sequence Matrix .....	583
<i>Chia-Ying Hsieh, Don-Lin Yang, and Jungpin Wu</i>	
Mining Allocating Patterns in One-Sum Weighted Items .....	592
<i>Yanbo J. Wang, Xinwei Zheng, Frans Coenen, and Cindy Y. Li</i>	
Remarks to Logical Aspects of Measures of Interestingness of Association Rules .....	599
<i>Jan Rauch</i>	
Bounding and Estimating Association Rule Support from Clusters on Binary Data .....	609
<i>Carlos Ordonez, Kai Zhao, and Zhibo Chen</i>	
Reclassification Rules .....	619
<i>Li-Shiang Tsay, Zbigniew W. Ras, and Seunghyun Im</i>	
A Logical Formulation of the Granular Data Model .....	628
<i>Tuan-Fang Fan, Churn-Jung Liau, Tsau-Young (T. Y.) Lin, and Karen Lee</i>	
Multilayer Change-Point Detection on Stock Order Flows by Wavelet Transformation .....	635
<i>Xiaoyan Liu, Xindong Wu, Huaiqing Wang, and Yingfeng Wang</i>	
Statistical Independence and Contingency Matrix .....	643
<i>Shusaku Tsumoto and Shoji Hirano</i>	
An FUSP-Tree Maintenance Algorithm for Record Modification .....	649
<i>Chun-Wei Lin, Tzung-Pei Hong, Wen-Hsiang Lu, and Hsin-Yi Chen</i>	

Hunting for Coherent Co-clusters in High Dimensional and Noisy Datasets .....	654
<i>Meghana Deodhar, Joydeep Ghosh, Gunjan Gupta, Hyuk Cho, and Inderjit Dhillon</i>	
Text Knowledge Mining: An Alternative to Text Data Mining .....	664
<i>D. Sánchez, M.J. Martín-Bautista, I. Blanco, and C. Justicia de la Torre</i>	
Estimating True and False Positive Rates in Higher Dimensional Problems and Its Data Mining Applications .....	673
<i>Andrew Foss and Osmar R. Zaiane</i>	
An Adaptive Pre-filtering Technique for Error-Reduction Sampling in Active Learning .....	682
<i>Michael Davy and Saturnino Luz</i>	
ARUBAS: An Association Rule Based Similarity Framework for Associative Classifiers .....	692
<i>Benoît Depaire, Koen Vanhoof, and Geert Wets</i>	
ZCS Revisited: Zeroth-Level Classifier Systems for Data Mining .....	700
<i>Fani A. Tzima and Pericles A. Mitkas</i>	
A New Graph-Based Algorithm for Clustering Documents .....	710
<i>Airel Pérez Suárez, José Fco. Martínez Trinidad, Jesús Ariel Carrasco Ochoa, and José E. Medina Pagola</i>	
The Set Classification Problem and Solution Methods .....	720
<i>Xia Ning and George Karypis</i>	

## **Workshop on Semantic Aspects in Data Mining (SADM 2008)**

If Constraint-Based Mining is the Answer: What is the Constraint? (Invited Talk) .....	730
<i>Jean-François Boulicaut</i>	
Semantic Features for Multi-view Semi-supervised and Active Learning of Text Classification .....	731
<i>Shiliang Sun</i>	
Ontology-Based Protein-Protein Interactions Extraction from Literature Using the Hidden Vector State Model .....	736
<i>Yulan He, Keiichi Nakata, and Deyu Zhou</i>	
Keyword Extraction Based on Lexical Chains and Word Co-occurrence for Chinese News Web Pages .....	744
<i>Xinghua Li, Xindong Wu, Xuegang Hu, Fei Xie, and Zhaozhong Jiang</i>	
OntoDM: An Ontology of Data Mining .....	752
<i>Panče Panov, Sašo Džeroski, and Larisa Soldatova</i>	
Semantic Annotation and Services for KDD Tools Sharing and Reuse .....	761
<i>Claudia Diamantini and Domenico Potena</i>	
Exploiting Data Semantics to Discover, Extract, and Model Web Sources .....	771
<i>José Luis Ambite, Craig A. Knoblock, Kristina Lerman, Anon Plangprasopchok, Thomas Russ, Cenk Gazen, Steven Minton, and Mark Carman</i>	

## **Workshop on Data Mining for Design and Marketing (DMDM 2008)**

Using Contextual Information to Decrease the Cost of Incorrect Predictions in On-line Customer Behavior Modeling .....	780
<i>M. Gorgoglione, C. Palmisano, and S. Lombardi</i>	
Semantic Analysis Method for Unstructured Data in Telecom Services .....	789
<i>Motoi Iwashita, Ken Nishimatsu, and Shinsuke Shimogawa</i>	
Using Contextual Information in Transactional Segmentation: An Empirical Study in E-Commerce .....	796
<i>M. F. Faraone, M. Gorgoglione, and C. Palmisano</i>	
Simultaneous Co-segmentation and Predictive Modeling for Large, Temporal Marketing Data .....	806
<i>Meghana Deodhar and Joydeep Ghosh</i>	
Title-Composing Support System for Reaching New Audiences .....	816
<i>Yoko Nishihara and Wataru Sunayama</i>	
Innovation Game as Workplace for Sensing Values in Design and Market .....	823
<i>Yukio Ohsawa, Yoshiharu Maeno, Akihiro Takaichi, and Yoko Nishihara</i>	
Character String Analysis and Customer Path in Stream Data .....	829
<i>Katsutoshi Yada</i>	
Combining Behavioral and Social Network Data for Online Advertising .....	837
<i>Abraham Bagherjeiran and Rajesh Parekh</i>	

## **Workshop on Video Mining (VM 2008)**

Semantic Concept Learning through Massive Internet Video Mining .....	847
<i>Peijiang Yuan, Bo Zhang, and Jianmin Li</i>	
Video^M: Multi-video Synopsis .....	854
<i>Teng Li, Tao Mei, In-So Kweon, and Xian-Sheng Hua</i>	
Human Action Recognition by Radon Transform .....	862
<i>Yan Chen, Qiang Wu, and Xiangjian He</i>	
A New Method for Multi-view Face Clustering in Video Sequence .....	869
<i>Panpan Huang, Yunhong Wang, and Ming Shao</i>	
Full-Reference Quality Assessment for Video Summary .....	874
<i>Tongwei Ren, Yan Liu, and Gangshan Wu</i>	
An Efficient Search Algorithm for Content-Based Image Retrieval with User Feedback .....	884
<i>Alex Po Leung and Peter Auer</i>	
Detecting Suspicious Behavior in Surveillance Images .....	891
<i>Daniel Barbará, Carlotta Domeniconi, Zoran Durić, Maurizio Filippone, Richard Mansfield, and Edgard Lawson</i>	



## Workshop on Analysis of Dynamic Networks (ADN 2008)

A Radar for the Internet .....	901
<i>Matthieu Latapy, Clémence Magnien, and Frédéric Ouédraogo</i>	
Temporal Evolution of the UK Web .....	909
<i>Iliaria Bordino, Paolo Boldi, Debora Donato, Massimo Santini, and Sebastiano Vigna</i>	
Characterizing Network Motifs to Identify Spam Comments .....	919
<i>Elaheh Kamaliha, Fatemeh Riahi, Vahed Qazvinian, and Jafar Adibi</i>	
Graph-Based Data Mining in Dynamic Networks: Empirical Comparison of Compression-Based and Frequency-Based Subgraph Mining .....	929
<i>Chang Hun You, Lawrence B. Holder, and Diane J. Cook</i>	
The Impact of Structural Changes on Predictions of Diffusion in Networks .....	939
<i>Mayank Lahiri, Arun S. Maiya, Rajmonda Sulo, Habiba, and Tanya Y. Berger Wolf</i>	
Using Betweenness Centrality to Identify Manifold Shortcuts .....	949
<i>William J. Cukierski and David J. Foran</i>	

## Demos

Semantic Full-Text Search with ESTER: Scalable, Easy, Fast .....	959
<i>Holger Bast, Fabian Suchanek, and Ingmar Weber</i>	
GRAPHITE: A Visual Query System for Large Graphs .....	963
<i>Duen Horng Chau, Christos Faloutsos, Hanghang Tong, Jason I. Hong, Brian Gallagher, and Tina Eliassi-Rad</i>	
Interactive Exploration of Model-Based Automatically Extracted Data .....	967
<i>Anni Coden, Igor Sominsky, and Michael Tanenblatt</i>	
G-REX: A Versatile Framework for Evolutionary Data Mining .....	971
<i>Rikard Konig, Ulf Johansson, and Lars Niklasson</i>	
GeoDMA - A Novel System for Spatial Data Mining .....	975
<i>Thales Sehn Korting, Leila Maria Garcia Fonseca, Maria Isabel Sobral Escada, Felipe Castro da Silva, and Marcelino Pereira dos Santos Silva</i>	
Automatic Multimodal Aggregation of News from Television and the Web .....	979
<i>Alberto Messina</i>	
A KDD Platform Based on the Application Service Provider Paradigm .....	983
<i>Fabio Fumarola, Eliana Salvemini, and Donato Malerba</i>	
A Data Stream Mining System .....	987
<i>Hetal Thakkar, Barzan Mozafari, and Carlo Zaniolo</i>	

## Author Index