

2007 6th International Conference on Machine Learning and Applications

(ICMLA 2007)

**Cincinnati, Ohio, USA
13 – 15 December 2007**



**IEEE Catalog Number: CFP07592-PRT
ISBN: 978-1-4244-3035-2**

Table of Contents

Sixth International Conference on Machine Learning and Applications

ICMLA 2007

Preface

Conference Committees

International Program Committee

Invited Speakers

Machine Learning for Information Management: Some Promising Directions

Dr. William Cohen, Carnegie Mellon University, USA

Probabilistic Graphical Models—Theory, Algorithm, and Application

Dr. Eric Xing, Carnegie Mellon University, USA

Machine Learning Challenges in Chemoinformatics and Drug Screening and Design

Professor Pierre Baldi, University of California Irvine, USA

Text Mining and Ontology Applications in Bioinformatics and GIS

Professor Shankant B. Navathe, Georgia Institute of Technology, USA

Neural Networks with Complex-Valued Neurons for Recurrent and Feedforward Architectures

Professor Jacek M. Zurada, University of Louisville, USA

Learning in Biomedicine and Bioinformatics Using Affinity Propagation

Dr. Brendan J. Frey, University of Toronto, Canada

Technologies and Solutions for Trend Detection in Public Literature for Biomarker Discovery

Dr. Bernd Wachmann, Siemens Corporate Research, NJ, USA

Support Vector Machines

An Optimization Method for Selecting Parameters in Support Vector Machines _____ 1

Yulin Dong, Zhonghang Xia, Manghui Tu, and Guangming Xing

Large Scale Classification with Support Vector Machine Algorithms _____ 7

Thanh-Nghi Do and Jean-Daniel Fekete

Modifying Kernels Using Label Information Improves SVM Classification Performance _____ 13

Renqiang Min, Anthony Bonner, and Zhaolei Zhang

Evolving Kernel Functions for SVMs by Genetic Programming _____ 19

Laura Diosan, Alexandrina Rogozan, and Jean Pierre Pecuchet

Sparsity Regularization Path for Semi-Supervised SVM _____ 25

Gilles Gasso, Karina Zapien, and Stephane Canu

Evolutionary Methods

Improving Gene Expression Programming Performance by Using Differential Evolution _____	31
<i>Qiongyun Zhang, Chi Zhou, Weimin Xiao, and Peter C. Nelson</i>	
Using Genetic Programming for the Induction of Oblique Decision Trees _____	38
<i>Amin Shali, Mohammad Reza Kangavari, and Bahareh Bina</i>	
Extraction of Minimum Decision Algorithm Using Rough Sets and Genetic Algorithms _____	44
<i>Michiyuki Hirokane, Shusaku Kouno, and Yasutoshi Nomura</i>	
Design of Simple Limit Cycles with Recurrent Neural Networks for Oscillatory Control _____	50
<i>Guillaume Jouffroy</i>	
Multi-Stages Genetic Algorithms: Introducing Temporal Structures to Facilitate Selection of Optimal Evolutionary Paths _____	56
<i>Ting Qian</i>	
A New Ant Evolution Algorithm to Resolve TSP Problem _____	62
<i>QingBao Zhu and ShuYan Chen</i>	

Data Mining

An Efficient Constraint-Based Closed Set Mining Algorithm _____	67
<i>Haiyun Bian, Raj Bhatnagar, and Barrington Young</i>	
Rare Itemset Mining _____	73
<i>Mehdi Adda, Lei Wu, and Yi Feng</i>	
Semantic Partition Based Association Rule Mining across Multiple Databases Using Abstraction _____	81
<i>P. Santhi Thilagam and Ananthanarayana V.S</i>	
An Itemset-Driven Cluster-Oriented Approach to Extract Compact and Meaningful Sets of Association Rules _____	87
<i>Claudio Haruo Yamamoto, Maria Cristina Ferreira de Oliveira, Magaly Lika Fujimoto, and Solange Oliveira Rezende</i>	
Clustering Categorical Data Based on Maximal Frequent Itemsets _____	93
<i>Dadong Yu, Dongbo Liu, Rui Luo, and Jianxin Wang</i>	

Learning

Learning Algorithms for Grammars of Variable Arity Trees _____	98
<i>Neetha Sebastian and Kamala Krithivasan</i>	
An Incremental Viterbi Algorithm _____	104
<i>Jason Bobbin</i>	
Phase Transition and Heuristic Search in Relational Learning _____	112
<i>Erick Alphonse and Aomar Osmani</i>	
Learning Complex Problem Solving Expertise from Failures _____	118
<i>Cristina Boicu, Gheorghe Tecuci, and Mihai Boicu</i>	
Bias-Variance Tradeoff in Hybrid Generative-Discriminative Models _____	124
<i>Guillaume Bouchard</i>	

Text and Multimedia Learning

Combining Active Learning and Relevance Vector Machines for Text Classification _____	130
<i>Catarina Silva and Bernardete Ribeiro</i>	
Efficient Label Propagation for Interactive Image Segmentation _____	136
<i>Fei Wang, Xin Wang, and Tao Li</i>	
An OCR-Independent Character Segmentation Using Shortest-Path in Grayscale Document Images _____	142
<i>Jia Tse, Christopher Jones, Dean Curtis, and Evangelos Yfantis</i>	
Boosting Inductive Transfer for Text Classification Using Wikipedia _____	148
<i>Somnath Banerjee</i>	
Discover the Power of Social and Hidden Curriculum to Decision Making: Experiments with Enron Email and Movie Newsgroups _____	154
<i>Hung-Ching Chen, Mark Goldberg, Malik Magdon-Ismail, and William A. Wallace</i>	

Application I

Soft Failure Detection Using Factorial Hidden Markov Models _____	160
<i>Guillaume Bouchard and Jean-Marc Andreoli</i>	
Memory-Based Context-Sensitive Spelling Correction at Web Scale _____	166
<i>Andrew Carlson and Ian Fette</i>	
Covariance Matrix Computations with Federated Databases _____	172
<i>Barrington Young, Raj Bhatnagar, Giridhar Tatavarty, and Haiyun Bian</i>	
Recognition of Ultrasonic Multi-Echo Sequences for Autonomous Symbolic Indoor Tracking _____	178
<i>André Stuhlsatz</i>	

Learning

An Agent Based System for California Electricity Market: A Perspective of Myopic Machine Learning _____	186
<i>Toshiyuki Sueyoshi and Gopalakrishna Reddy Tadiparthi</i>	
Predicting Building Contamination Using Machine Learning _____	192
<i>Shawn Martin and Sean McKenna</i>	
Bootstrapping Algorithms for an Application in the Automotive Domain _____	198
<i>Martin Schierle and Sascha Schulz</i>	

Classification

An Application of a Rule-Based Model in Software Quality Classification _____	204
<i>Taghi M. Khoshgoftaar, Lofton A. Bullard, and Kehan Gao</i>	
Estimating Class Probabilities in Random Forests _____	211
<i>Henrik Boström</i>	
Effect of Synaptic Weight Assignment on Spiking Neuron Response _____	217
<i>Robert H. Fujii and Taiki Ichishita</i>	

Application II

A Comparison of Two Algorithms for Predicting the Condition Number _____	223
<i>Dianwei Han and Jun Zhang</i>	
Evolutionary Sound Matching: A Test Methodology and Comparative Study _____	229
<i>Thomas J. Mitchell and David P. Creasey</i>	
Learning to Evaluate Conditional Partial Plans _____	235
<i>Slawomir Nowaczyk and Jacek Malec</i>	

Regression/Prediction

A Spline Based Regression Technique on Interval Valued Noisy Data _____	241
<i>Balaji Kommineni, Shubhankar Basu, and Ranga Vemuri</i>	
A New Time Series Prediction Algorithm Based on Moving Average of n^{th} -Order Difference _____	248
<i>Yang Lan and Daniel Neagu</i>	
Machine Learned Regression for Abductive DNA Sequencing _____	254
<i>David J. Thornley, Maxim Zverev, and Stavros Petridis</i>	
Automatic Medical Coding of Patient Records via Weighted Ridge Regression _____	260
<i>Jian-Wu Xu, Shipeng Yu, Jinbo Bi, Lucian Vlad Lita, Radu Stefan Niculescu, and R. Bharat Rao</i>	
Constructive Neural Network Ensemble for Regression Tasks in High Dimensional Spaces _____	266
<i>Adeline Schmitz and Hamid Hefazi</i>	

Machine Learning in Web Based Real-Time Applications

Web-Based Maze Robot Learning Using Fuzzy Motion Control System _____	274
<i>Nihat Yilmaz and Seref Sagiroglu</i>	
Web Based Machine Learning for Language Identification and Translation _____	280
<i>Seref Sagiroglu, Uraz Yavanoglu, and Esra Nergis Guven</i>	
A Web Based Adaptive Educational System _____	286
<i>Hamdi Tolga Kahraman, Ilhami Colak, and Seref Sagiroglu</i>	

Dimensionality Reduction

Dimensionality Reduction for Active Learning with Nearest Neighbour Classifier in Text Categorisation Problems _____	292
<i>Michael Davy and Saturnino Luz</i>	
Scalable Optimal Linear Representation for Face and Object Recognition _____	298
<i>Yiming Wu, Xiuwen Liu, and Washington Mio</i>	

Induction, Model Selection and Evaluation

Model Evaluation for Prognostics: Estimating Cost Saving for the End Users _____	304
<i>Chunsheng Yang and Sylvain Létourneau</i>	
Rule Refinement with Extended Data Expression _____	310
<i>Jung Min Kong, Dong-Hun Seo, and Won Don Lee</i>	
A Simultaneous Two-Level Clustering Algorithm for Automatic Model Selection _____	316
<i>Guénaél Cabanes and Younès Bennani</i>	

LEONARDO – The Computational Intelligence (CI) Model Selection Wizard _____	322
<i>Peter Owotoki and Friedrich Mayer-Lindenberg</i>	
Control of a Re-Entrant Line Manufacturing Model with a Reinforcement Learning Approach _____	330
<i>José A. Ramírez-Hernández and Emmanuel Fernandez</i>	

Reinforcement Learning/ Learning from Imbalanced Data/ Bayesian Networks

Supervised Reinforcement Learning Using Behavior Models _____	336
<i>Victor Uc-Cetina</i>	
Performance Evaluation of EANT in the RoboCup Keepaway Benchmark _____	342
<i>Jan Hendrik Metzen, Mark Edgington, Yohannes Kassahun, and Frank Kirchner</i>	
Learning with Limited Minority Class Data _____	348
<i>Taghi M. Khoshgoftaar, Chris Seiffert, Jason Van Hulse, Amri Napolitano, and Andres Folleco</i>	
Combining Multi-Distributed Mixture Models and Bayesian Networks for Semi-Supervised Learning _____	354
<i>Manuel Stritt, Lars Schmidt-Thieme, and Gerhard Poeppel</i>	
Using Evolutionary Sampling to Mine Imbalanced Data _____	363
<i>Dennis J. Drown, Taghi M. Khoshgoftaar, and Ramaswamy Narayanan</i>	
Learning Bayesian Networks Consistent with the Optimal Branching _____	369
<i>Alexandra M. Carvalho and Arlindo L. Oliveira</i>	

Unsupervised Learning

Manifold Clustering via Energy Minimization _____	375
<i>Qiyong Guo, Hongyu Li, Wenbin Chen, I-Fan Shen, and Jussi Parkkinen</i>	
Co-Clustering by Similarity Refinement _____	381
<i>Jian Zhang</i>	
2IBGSOM: Interior and Irregular Boundaries Growing Self-Organizing Maps _____	387
<i>Thouraya Ayadi, Tarek M. Hamdani, Adel M. Alimi, and Mohamed A. Khabou</i>	
Comparison of Semantic and Single Term Similarity Measures for Clustering Turkish Documents _____	393
<i>Bülent Yücesoy and Şule Gündüz Oğüdücü</i>	
A Music Recommendation System with a Dynamic K-means Clustering Algorithm _____	399
<i>Dongmoon Kim, Kun-su Kim, Kyo-Hyun Park, Jee-Hyong Lee, and Keon Myung Lee</i>	

Feature Extraction and Selection / Ensemble Learning

Tracking Recurrent Concept Drift in Streaming Data Using Ensemble Classifiers _____	404
<i>Sasthakumar Ramamurthy and Raj Bhatnagar</i>	
Feature Extraction Using Random Matrix Theory Approach _____	410
<i>Viktoria Rojkova and Mehmed Kantardzic</i>	
Toward Optimal Selection of Feature Clusters _____	417
<i>Lei Yu and Hao Li</i>	
A Hierarchical Feature Decomposition Clustering Algorithm for Unsupervised Classification of Document Image Types _____	423
<i>Dean Curtis, Vitaliy Kubushyn, E.A. Yfantis, and Michael Rogers</i>	
Enhanced Recursive Feature Elimination _____	429
<i>Xue-wen Chen and Jong Cheol Jeong</i>	

Application III

Learning a Knowledge Base of Ontological Concepts for High-Level Scene Interpretation _____	436
<i>Johannes Hartz and Bernd Neumann</i>	
Web-Based Parallel Corpora for Statistical Machine Translation _____	444
<i>Bo Li, Juan Liu, and Wenjuan Shi</i>	
Uncertainty Optimization for Robust Dynamic Optical Flow Estimation _____	450
<i>Volker Willert, Marc Toussaint, Julian Eggert, and Edgar Körner</i>	
Direct Calculation of Predictions for K29/K29* _____	458
<i>Brian Burford</i>	
Learning: An Effective Approach in Endgame Chess Board Evaluation _____	464
<i>Mehdi Samadi, Zohreh Azimifar, and Mansour Zolghadri Jahromi</i>	
Hybrid Neural Network Based Model for Predicting the Performance of a Two Stroke Spark Ignition Engine _____	470
<i>Mohmad Marouf Wani and M. Arif Wani</i>	

Machine Learning Applications in Bioinformatics

Predicting Binding Sites in the Mouse Genome _____	476
<i>Yi Sun, Mark Robinson, Rod Adams, Neil Davey, and Alistair Rust</i>	
Reducing a Biomarkers List via Mathematical Programming: Application to Gene Signatures to Detect Time-Dependent Hypoxia in Cancer _____	482
<i>Glenn Fung, Renaud. Seigneuric, Sriram Krishnan, R. Bharat Rao, Brad G. Wouters, and Philippe Lambin</i>	
Polynomial and RBF Kernels as Marker Selection Tools—A Breast Cancer Case Study _____	488
<i>Michalis E. Blazadonakis and Michalis Zervakis</i>	
Statistical and Biological Validation Methods in Cluster Analysis of Gene Expression _____	494
<i>Daniele Yumi Sunaga, Julio Cesar Nievola, and Milton Pires Ramos</i>	
Support Vector Machine Classification of Probability Models and Peptide Features for Improved Peptide Identification from Shotgun Proteomics _____	500
<i>Bobbie-Jo M. Webb-Robertson, Christopher S. Oehmen, and William R. Cannon</i>	
Identifying Functional Binding Motifs of Tumor Protein p53 Using Support Vector Machines _____	506
<i>Amit U. Sinha, Mukta Phatak, Raj Bhatnagar, and Anil G. Jegga</i>	

Workshop Session: Machine Learning Applications in Biomedicine

Use of Neural Networks to Predict Adverse Outcomes from Acute Coronary Syndrome for Male and Female Patients _____	512
<i>Claire L. McCullough, Andrew J. Novobilski, and Francis M. Fesmire</i>	
The Natural Induction System AQ21 and its Application to Data Describing Patients with Metabolic Syndrome: Initial Results _____	518
<i>Janusz Wojtusiak, Ryszard S. Michalski, Thipkesone Simanivanh, and Anna V. Baranova</i>	
Understanding Challenges in Preserving and Reconstructing Computer-Assisted Medical Decision Processes _____	524
<i>Sang-Chul Lee and Peter Bajcsy</i>	
Semi-Supervised Active Learning for Modeling Medical Concepts from Free Text _____	530
<i>Rómer Rosales, Praveen Krishnamurthy, and R. Bharat Rao</i>	

A Statistical Algorithm to Discover Knowledge in Medical Data Sources _____	537
<i>Alexander J. Senf, Carl Leonard, and James DeLeo</i>	
A Clustering-Based Approach to Predict Outcome in Cancer Patients _____	541
<i>Kai Xing, Dechang Chen, Donald Henson, and Li Sheng</i>	

Workshop Session: Machine Learning Applications in Genomics

Maximum Likelihood Quantization of Genomic Features Using Dynamic Programming _____	547
<i>Mingzhou (Joe) Song, Robert M. Haralick, and Stéphane Boissinot</i>	
CLASSEQ: Classification of Sequences via Comparative Analysis of Multiple Genomes _____	554
<i>Kwangmin Choi, Youngik Yang, and Sun Kim</i>	
Biomarker Identification by Knowledge-Driven Multi-Level ICA and Motif Analysis _____	560
<i>Li Chen, Chen Wang, Ie-Ming Shih, Tian-Li Wang, Zhen Zhang, Yue Wang, Robert Clarke, Eric Hoffman, and Jianhua Xuan</i>	
Generalized Sequence Signatures through Symbolic Clustering _____	567
<i>Dietmar Dorr and Anne Denton</i>	
SVMotif: A Machine Learning Motif Algorithm _____	573
<i>Mark A. Kon, Yue Fan, Dustin Holloway, and Charles DeLisi</i>	

Workshop Session: Machine Learning Applications in Proteomics

Association Learning in SOMs for Fuzzy-Classification _____	581
<i>Thomas Villmann, Frank-Michael Schleif, Martijn van der Werff, Andre Deelder, and Rob Tollenaar</i>	
SEBINI-CABIN: An Analysis Pipeline for Biological Network Inference, with a Case Study in Protein-Protein Interaction Network Reconstruction _____	587
<i>Ronald C. Taylor, Mudita Singhal, Don S. Daly, Kelly Domico, Amanda M. White, Deanna L. Auberry, Kenneth J. Auberry, Brian Hooker, Greg Hurst, Jason McDermott, W. Hayes McDonald, Dale Pelletier, Denise Schmoyer, and William R. Cannon</i>	
Alignment of Multiple Proteins with an Ensemble of Hidden Markov Models _____	594
<i>Jia Song, Chunmei Liu, Yinglei Song, Junfeng Qu, and Gurdeep S. Hura</i>	
Improvement of Bayesian Network Inference Using a Relaxed Gene Ordering _____	600
<i>Dongxiao Zhu and Hua Li</i>	

Workshop Session: Machine Learning Applications in Transcriptomics

Feature Extraction from Microarray Expression Data by Integration of Semantic Knowledge _____	606
<i>Young-Rae Cho, Xian Xu, Woochang Hwang, and Aidong Zhang</i>	
Normalized Linear Transform for Cross-Platform Microarray Data Integration _____	612
<i>Huilin Xiong, Ya Zhang, and Xue-wen Chen</i>	
Logistic Ensembles for Random Spherical Linear Oracles _____	618
<i>Leif E. Peterson and Matthew A. Coleman</i>	
Cox's Proportional Hazards Model with L_p Penalty for Biomarker Identification and Survival Prediction _____	624
<i>Zhenqiu Liu</i>	
Modeling Spatial-Temporal Epidemics Using STBL Model _____	629
<i>Lynne Billard, Duck-Ki Kim, Chan-Hee Lee, Sung Duck Lee, Keon-Myung Lee, and Sung -Soo Kim</i>	

Author Index