2007 6th International Conference on Machine Learning and Applications

(ICMLA 2007)

Cincinnati, Ohio, USA
13 – 15 December 2007
# 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 Shamkant 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**  
  *Yulin Dong, Zhonghang Xia, Manghui Tu, and Guangming Xing*  
  page 1

- **Large Scale Classification with Support Vector Machine Algorithms**  
  *Thanh-Nghi Do and Jean-Daniel Fekete*  
  page 7

- **Modifying Kernels Using Label Information Improves SVM Classification Performance**  
  *Renqiang Min, Anthony Bonner, and Zhaolei Zhang*  
  page 13

- **Evolving Kernel Functions for SVMs by Genetic Programming**  
  *Laura Diosan, Alexandrina Rogozan, and Jean Pierre Pecuchet*  
  page 19

- **Sparsity Regularization Path for Semi-Supervised SVM**  
  *Gilles Gasso, Karina Zapien, and Stephane Canu*  
  page 25
Evolutionary Methods

Improving Gene Expression Programming Performance by Using Differential Evolution
Qiongyun Zhang, Chi Zhou, Weimin Xiao, and Peter C. Nelson

Using Genetic Programming for the Induction of Oblique Decision Trees
Amin Shali, Mohammad Reza Kangavari, and Bahareh Bina

Extraction of Minimum Decision Algorithm Using Rough Sets and Genetic Algorithms
Michiyuki Hirokane, Shusaku Kouno, and Yasutoshi Nomura

Design of Simple Limit Cycles with Recurrent Neural Networks for Oscillatory Control
Guillaume Jouffroy

Multi-Stages Genetic Algorithms: Introducing Temporal Structures to Facilitate Selection of Optimal Evolutionary Paths
Ting Qian

A New Ant Evolution Algorithm to Resolve TSP Problem
QingBao Zhu and ShuYan Chen

Data Mining

An Efficient Constraint-Based Closed Set Mining Algorithm
Haiyun Bian, Raj Bhatnagar, and Barrington Young

Rare Itemset Mining
Mehdi Adda, Lei Wu, and Yi Feng

Semantic Partition Based Association Rule Mining across Multiple Databases Using Abstraction
P. Santhi Thilagam and Ananthanarayana V.S

An Itemset-Driven Cluster-Oriented Approach to Extract Compact and Meaningful Sets of Association Rules
Claudio Haruo Yamamoto, Maria Cristina Ferreira de Oliveira, Magaly Lika Fujimoto, and Solange Oliveira Rezende

Clustering Categorical Data Based on Maximal Frequent Itemsets
Dadong Yu, Dongbo Liu, Rui Luo, and Jianxin Wang

Learning

Learning Algorithms for Grammars of Variable Arity Trees
Neetha Sebastian and Kamala Krithivasan

An Incremental Viterbi Algorithm
Jason Bobbin

Phase Transition and Heuristic Search in Relational Learning
Erick Alphonse and Aomar Osmani

Learning Complex Problem Solving Expertise from Failures
Cristina Boicu, Gheorghe Tecuci, and Mihai Boicu

Bias-Variance Tradeoff in Hybrid Generative-Discriminative Models
Guillaume Bouchard
Text and Multimedia Learning

Combining Active Learning and Relevance Vector Machines for Text Classification
Catarina Silva and Bernardete Ribeiro
130

Efficient Label Propagation for Interactive Image Segmentation
Fei Wang, Xin Wang, and Tao Li
136

An OCR-Independent Character Segmentation Using Shortest-Path in Grayscale Document Images
Jia Tse, Christopher Jones, Dean Curtis, and Evangelos Yfantis
142

Boosting Inductive Transfer for Text Classification Using Wikipedia
Somnath Banerjee
148

Discover the Power of Social and Hidden Curriculum to Decision Making: Experiments with Enron Email and Movie Newsgroups
Hung-Ching Chen, Mark Goldberg, Malik Magdon-Ismail, and William A. Wallace
154

Application I

Soft Failure Detection Using Factorial Hidden Markov Models
Guillaume Bouchard and Jean-Marc Andreoli
160

Memory-Based Context-Sensitive Spelling Correction at Web Scale
Andrew Carlson and Ian Fette
166

Covariance Matrix Computations with Federated Databases
Barrington Young, Raj Bhatnagar, Giridhar Tatavarty, and Haiyun Bian
172

Recognition of Ultrasonic Multi-Echo Sequences for Autonomous Symbolic Indoor Tracking
André Stuhlsatz
178

Learning

An Agent Based System for California Electricity Market: A Perspective of Myopic Machine Learning
Toshiyuki Sueyoshi and Gopalakrishna Reddy Tadiparthi
186

Predicting Building Contamination Using Machine Learning
Shawn Martin and Sean McKenna
192

Bootstrapping Algorithms for an Application in the Automotive Domain
Martin Schierle and Sascha Schulz
198

Classification

An Application of a Rule-Based Model in Software Quality Classification
Taghi M. Khoshgoftaar, Lofton A. Bullard, and Kehan Gao
204

Estimating Class Probabilities in Random Forests
Henrik Boström
211

Effect of Synaptic Weight Assignment on Spiking Neuron Response
Robert H. Fujii and Taiki Ichishita
217
Application II

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

Regression/Prediction

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

Machine Learning in Web Based Real-Time Applications

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

Dimensionality Reduction

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

Induction, Model Selection and Evaluation

Model Evaluation for Prognostics: Estimating Cost Saving for the End Users
Chunsheng Yang and Sylvain Létourneau 304
Rule Refinement with Extended Data Expression
Jung Min Kong, Dong-Hun Seo, and Won Don Lee 310
A Simultaneous Two-Level Clustering Algorithm for Automatic Model Selection
Guénaël Cabanes and Younès Bemmani 316
Reinforcement Learning/ Learning from Imbalanced Data/ Bayesian Networks

Supervised Reinforcement Learning Using Behavior Models 336
\textit{Victor Uc-Cetina}

Performance Evaluation of EANT in the RoboCup Keepaway Benchmark 342
\textit{Jan Hendrik Metzen, Mark Edgington, Yohannes Kassahun, and Frank Kirchner}

Learning with Limited Minority Class Data 348
\textit{Taghi M. Khoshgoftaar, Chris Seiffert, Jason Van Hulse, Amri Napolitano, and Andres Folleco}

Combining Multi-Distributed Mixture Models and Bayesian Networks for Semi-Supervised Learning 354
\textit{Manuel Stritt, Lars Schmidt-Thieme, and Gerhard Poeppel}

Using Evolutionary Sampling to Mine Imbalanced Data 363
\textit{Dennis J. Drown, Taghi M. Khoshgoftaar, and Ramaswamy Narayanan}

Learning Bayesian Networks Consistent with the Optimal Branching 369
\textit{Alexandra M. Carvalho and Arlindo L. Oliveira}

Unsupervised Learning

Manifold Clustering via Energy Minimization 375
\textit{Qiyong Guo, Hongyu Li, Wenbin Chen, I-Fan Shen, and Jussi Parkkinen}

Co-Clustering by Similarity Refinement 381
\textit{Jian Zhang}

2IBGSOM: Interior and Irregular Boundaries Growing Self-Organizing Maps 387
\textit{Thouraya Ayadi, Tarek M. Hamdani, Adel M. Alimi, and Mohamed A. Khabou}

Comparison of Semantic and Single Term Similarity Measures for Clustering Turkish Documents 393
\textit{Bülent Yücesoy and Şule Gündüz Oğudücü}

A Music Recommendation System with a Dynamic K-means Clustering Algorithm 399
\textit{Dongmoon Kim, Kun-su Kim, Kyo-Hyun Park, Jee-Hyong Lee, and Keon Myung Lee}

Feature Extraction and Selection / Ensemble Learning

Tracking Recurrent Concept Drift in Streaming Data Using Ensemble Classifiers 404
\textit{Sasthakumar Ramamurthy and Raj Bhatnagar}

Feature Extraction Using Random Matrix Theory Approach 410
\textit{Viktoria Rajkova and Mehmed Kantardzic}

Toward Optimal Selection of Feature Clusters 417
\textit{Lei Yu and Hao Li}

A Hierarchical Feature Decomposition Clustering Algorithm for Unsupervised Classification of Document Image Types 423
\textit{Dean Curtis, Vitaliy Kubushyn, E.A. Yfantis, and Michael Rogers}

Enhanced Recursive Feature Elimination 429
\textit{Xue-wen Chen and Jong Cheol Jeong}
Application III

Learning a Knowledge Base of Ontological Concepts for High-Level Scene Interpretation
Johannes Hartz and Bernd Neumann

Web-Based Parallel Corpora for Statistical Machine Translation
Bo Li, Juan Liu, and Wenjuan Shi

Uncertainty Optimization for Robust Dynamic Optical Flow Estimation
Volker Willert, Marc Toussaint, Julian Eggert, and Edgar Körner

Direct Calculation of Predictions for K29/K29*
Brian Burford

Learning: An Effective Approach in Endgame Chess Board Evaluation
Mehdi Samadi, Zohreh Azimifar, and Mansour Zolghadri Jahromi

Hybrid Neural Network Based Model for Predicting the Performance of a Two Stroke Spark Ignition Engine
Mohmad Marouf Wani and M. Arif Wani

Machine Learning Applications in Bioinformatics

Predicting Binding Sites in the Mouse Genome
Yi Sun, Mark Robinson, Rod Adams, Neil Davey, and Alistair Rust

Reducing a Biomarkers List via Mathematical Programming: Application to Gene Signatures to Detect Time-Dependent Hypoxia in Cancer
Glenn Fung, Renaud. Seigneuric, Sriram Krishnan, R. Bharat Rao, Brad G. Wouters, and Philippe Lambin

Polynomial and RBF Kernels as Marker Selection Tools—A Breast Cancer Case Study
Michalis E. Blazadonakis and Michalis Zervakis

Statistical and Biological Validation Methods in Cluster Analysis of Gene Expression
Daniele Yumi Sunaga, Julio Cesar Nievola, and Milton Pires Ramos

Support Vector Machine Classification of Probability Models and Peptide Features for Improved Peptide Identification from Shotgun Proteomics
Bobbie-Jo M. Webb-Robertson, Christopher S. Oehmen, and William R. Cannon

Identifying Functional Binding Motifs of Tumor Protein p53 Using Support Vector Machines
Amit U. Sinha, Mukta Phatak, Raj Bhatnagar, and Anil G. Jegga

Workshop Session: Machine Learning Applications in Biomedicine

Use of Neural Networks to Predict Adverse Outcomes from Acute Coronary Syndrome for Male and Female Patients
Claire L. McCullough, Andrew J. Novobilski, and Francis M. Fesmire

The Natural Induction System AQ21 and its Application to Data Describing Patients with Metabolic Syndrome: Initial Results
Janusz Wojtusiak, Ryszard S. Michalski, Thipkesone Simanivanh, and Anna V. Baranova

Understanding Challenges in Preserving and Reconstructing Computer-Assisted Medical Decision Processes
Sang-Chul Lee and Peter Bajcsy

Semi-Supervised Active Learning for Modeling Medical Concepts from Free Text
Róméros Rosales, Praveen Krishnamurthy, and R. Bharat Rao
Workshop Session: Machine Learning Applications in Genomics

Maximum Likelihood Quantization of Genomic Features Using Dynamic Programming
Mingzhou (Joe) Song, Robert M. Haralick, and Stéphane Boissinot

CLASSEQ: Classification of Sequences via Comparative Analysis of Multiple Genomes
Kwangmin Choi, Youngik Yang, and Sun Kim

Biomarker Identification by Knowledge-Driven Multi-Level ICA and Motif Analysis
Li Chen, Chen Wang, Ie-Ming Shih, Tian-Li Wang, Zhen Zhang, Yue Wang, Robert Clarke, Eric Hoffman, and Jianhua Xuan

Generalized Sequence Signatures through Symbolic Clustering
Dietmar Dorr and Anne Denton

SVMotif: A Machine Learning Motif Algorithm
Mark A. Kon, Yue Fan, Dustin Holloway, and Charles DeLisi

Workshop Session: Machine Learning Applications in Proteomics

Association Learning in SOMs for Fuzzy-Classification
Thomas Villmann, Frank-Michael Schleif, Martijn van der Werff, Andre Deelder, and Rob Tollemaal

SEBINI-CABIN: An Analysis Pipeline for Biological Network Inference, with a Case Study in Protein-Protein Interaction Network Reconstruction
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

Alignment of Multiple Proteins with an Ensemble of Hidden Markov Models
Jia Song, Chunmei Liu, Yinglei Song, Junfeng Qu, and Gurdeep S. Hura

Improvement of Bayesian Network Inference Using a Relaxed Gene Ordering
Dongxiao Zhu and Hua Li

Workshop Session: Machine Learning Applications in Transcriptomics

Feature Extraction from Microarray Expression Data by Integration of Semantic Knowledge
Young-Rae Cho, Xian Xu, Woochang Hwang, and Aidong Zhang

Normalized Linear Transform for Cross-Platform Microarray Data Integration
Huilin Xiong, Ya Zhang, and Xue-wen Chen

Logistic Ensembles for Random Spherical Linear Oracles
Leif E. Peterson and Matthew A. Coleman

Cox’s Proportional Hazards Model with \( L_p \) Penalty for Biomarker Identification and Survival Prediction
Zhenqiu Liu

Modeling Spatial-Temporal Epidemics Using STBL Model
Lynne Billard, Duck-Ki Kim, Chan-Hee Lee, Sung Duck Lee, Keon-Myung Lee, and Sung-Soo Kim

Author Index