

International Conference on Machine Learning and Data Mining MLDM'2009

Program

First Conference Day: July 23, 2009

08.50 **Welcome**
 Petra Perner

Attribute Discretization and Data Preparation

Session Chair: Giorgio Giacinto

09.00 **Improved Comprehensibility and Reliability of Explanations via Restricted Halfspace Discretization**
 Klaus Truemper

09.20 **Selection of Subsets of Ordered Features in Machine Learning**
 O. Seredin, A. Kopylov, V. Mottl

09.40 **Combination of Vector Quantization and Visualization**
 Olga Kurasova, Alma Molyte

10.00 **Discretization of Target Attributes for Subgroup Discovery**
 Katherine Moreland, Klaus Truemper

10.20 **Preserving Privacy in Time Series Data Classification by Discretization**
 Ye Zhu, Yongjian Fu, Huirong Fu

10.40 **Using Resampling Techniques for Better Quality Discretization**
 Taimur Qureshi, Djamel.A.Zighed

11.00 Break

Classification

Session Chair: Rui Camacho

11.15 **A Large Margin Classifier with Additional Features**
 Xinwang Liu, Jianping Yin, En Zhu, Guomin Zhang, Yubin Zhan, Miaomiao Li

11.35 **Sequential EM for Unsupervised Adaptive Gaussian Mixture Model Based Classifier**
 Bashar Awwad Shiekh Hasan, John Q. Gan

11.55 **Efficient AdaBoost Region Classification**
 M. Moed, E.N. Smirnov

12.15 **PMCRI: A Parallel Modular Classification Rule Induction Framework**
 Frederic Stahl, Max Bramer, Mo Adda

12.35 Lunch

14.00 **Dynamic Score Combination: a supervised and unsupervised score combination method**
 Roberto Tronci, Giorgio Giacinto, Fabio Roli

14.20 **ODDboost: Incorporating Posterior Estimates into AdaBoost**

Olga Barinova, Dmitry Vetrov

14.40 **Particular Words Mining and Article Spotting in Old French Gazettes**
rescheduled Loris Eynard, Yann Leydier, and Hubert Emptoz

Ensemble Classifier Learning

Session Chair: Ceci Michelangelo

15.00

Ensemble learning: a study on different variants of the dynamic selection approach
João Mendes-Moreira, Alipio Mario Jorge, Carlos Soares, Jorge Freire de Sousa

15.20

Relevance and Redundancy Analysis for Ensemble Classifiers
Rakkrit Duangsoithong, Terry Windeatt

15.40

Drift-Aware Ensemble Regression
Frank Rosenthal, Peter Benjamin Volk, Martin Hahmann, Dirk Habich, Wolfgang Lehner

16.00

Concept Drifting Detection on Noisy Streaming Data in Random Ensemble Decision Trees
Peipei Li, Xuegang Hu, Qianhui Liang, Yunjun Gao

16.20

Break

Support Vector Machines

Session Chair: Joao Mendes-Moreira

16.35

Fast Local Support Vector Machines for Large Datasets
Nicola Segata, Enrico Blanzieri

16.55

The Effect of Domain Knowledge on Rule Extraction from Support Vector Machines
Nahla Barakat, Andrew P. Bradley

17.15

Towards B-coloring of SOM
Haytham Elghazel, Khalid Benabdeslem

Clustering

Session Chair: Anselmo Cardoso de Paiva

17.35

CSBIterKmeans: A New Clustering Algorithm Based on Quantitative Assessment of the Clustering Quality
Tarek Smaoui, Sascha Müller, Christian Müller-Schloer

17.55

Agent-Based Non-Distributed and Distributed Clustering
Ireneusz Czarnowski and Piotr Jędrzejowicz

18.15

An Evidence Accumulation Approach to Constrained Clustering Combination
Daniel Duarte Abdala, Xiaoyi Jiang

18.35

Poster Session

19.30

End of Session

Second Conference Day: July 24, 2009

Session Chair: Stephan France

- 09.00 **Fast Spectral Clustering with Random Projection and Sampling**
Tomoya Sakai, Atsushi Imiya
- 09.20 **How much true structure has been discovered? Validating Explorative Clustering on a Hold-Out Test Set**
F. Höppner
- 09.40 **Efficient Clustering of Web-Derived Data Sets**
Luís Sarmento, Alexander Kehlenbeck, Eugénio Oliveira, Lyle Ungar
- 10.00 **A probabilistic approach for constrained clustering with topological map**
Khalid Benabdeslem, Jihene Snoussi

Novelty and Outlier Detection

Session Chair: Petra Pernert

- 10.20 **Relational Frequent Patterns Mining for Novelty Detection from Data Streams**
Michelangelo Ceci, Annalisa Appice, Corrado Loglisci, Costantina Caruso, Fabio Fumarola, Carmine Valente, Donato Malerba
- 10.40 Break
- 10.55 **A Comparative Study of Outlier Detection Algorithms**
Charlie Isaksson, Margaret H. Dunham,
- 11.15 **A linear classification method in a very high dimensional space using distributed representation**
rescheduled Takao Kobayashi, Ikuko Shimizu

Learning

Session Chair: Kaspar Riesen

- 11.35 **Concept Learning from (very) Ambiguous Examples**
Dominique Bouthinon, Henry Soldano, Véronique Ventos
- 11.55 **Finding Top-N Pseudo Formal Concepts with Core Intents**
Yoshiaki Okubo, Makoto Haraguchi
- 12.15 **On Fixed Convex Combinations of No-Regret Learners**
Jan-P. Calliess
- 12.35 Lunch
- 14.00 **An Improved Tabu Search (ITS) Algorithm based on Open Cover Theory for Global Extremums**
Kemal Yükses, Serhat Cakaloglu
- 14.20 **The Needles-In-Haystack Problem**
Katherine Moreland, Klaus Truemper

Association Rules and Pattern Mining

Session Chair: Georg Ruß

- 14.40 **Mining Multiple Level Non-Redundant Association Rules through Two-Fold Pruning of Redundancies**

Corrado Loglisci , Donato Malerba

- 15.00 **Pattern Mining with Natural Language Processing: An exploratory approach**
Ana Cristina Mendes, Cláudia Antunes
- 15.20 **Is the distance compression effect overstated? Some theory and experimentation**
Stephen France, Douglas Carroll
- 15.40 Break

Data Mining on Multimedia Data

Session Chair: Yanbo J. Wang

- 15.55 **An evidence-driven probabilistic inference framework for semantic image understanding**
Spiros Nikolopoulos, Georgios. Th. Papadopoulos, Ioannis Kompatsiaris, Ioannis Patras
- 16.15 **Detection of Masses in Mammographic Images Using Simpson's Diversity Index in Circular Regions and SVM**
André Pereira Nunes, Aristófanés Corrêa Silva, Anselmo Cardoso de Paiva
- 16.45 **Mining Lung Shape from X-ray Images**
Vassili Kovalev, Aliaksandr Prus, Pavel Vankevich
- 17.05 **A Wavelet-based Method for Detecting Seismic Anomalies in Remote Sensing Satellite Data**
Pan Xiong, Yaxin Bi, Xuhui Shen
- 18.00 **Banquet**

Third Conference Day: July 25, 2009

Session Chair: Piotr Jedrzejowicz

- 09.00 **Spectrum Steganalysis of WAV Audio Streams**
Qingzhong Liu , Andrew H. Sung, Mengyu Qiao
- 09.20 **Audio-based Emotion Recognition in Judicial Domain: A Multilayer Support Vector Machines Approach**
E. Fersini, E. Messina, G. Arosio, F. Archetti
- 09.40 **Learning with a Quadruped Chopstick Robot**
Wei-Chung Lee, Jong-Chen Chen, Shou-zhe Wu, Kuo-Ming Lin
- 10.00 **Dissimilarity Based Vector Space Embedding of Graphs Using Prototype Reduction Schemes**
Kaspar Riesen, Horst Bunke

Text Mining

Session Chair: Sara Colantonio

- 10.20 **Using graph-kernels to represent semantic information in text classification**
Teresa Goncalves, Paulo Quaresma
- 10.40 **A General Framework of Feature Selection for Text Categorization**
Hongfang Jing, Bin Wang, Yahui Yang, Yan Xu
- 11.00 Break

- 11.15 **New Semantic Similarity Based Model for Text Clustering Using Extended Gloss Overlaps**
Walaa K. Gad, Mohamed S. Kamel

Aspects of Data Mining

Session Chair: Enrico Blanzieri

- 11.35 **Learning Betting Tips from Users' Bet Selections**
Erik Štrumbelj, Marko Robnik Šikonja, Igor Kononenko
- 11.55 **An Approach to Web-scale Named-Entity Disambiguation**
Luís Sarmento, Alexander Kehlenbeck, Eugénio Oliveira, Lyle Ungar
- 12.15 **A general learning method for automatic title extraction from HTML pages**
Sahar Changuel, Nicolas Labroche, Bernadette Bouchon-Meunier
- 12.35 **Regional Pattern Discovery in Geo-Referenced Datasets Using PCA**
Oner Ulvi Celepcikay, Christoph F. Eick, Carlos Ordonez
- 12.55 Lunch
- 14.30 **Memory-Based Modeling of Seasonality for Prediction of Climatic Time Series**
Daniel Nikovski, Ganesan Ramachandran
- 14.50 **A Neural Approach for SME's Credit Risk Analysis in Turkey**
Gülner Derelioglu, Fikret Gürgen , Nesrin Okay
- 15.10 **Assisting Data Mining Through Automated Planning**
Fernando Fernández, Daniel Borrajo, Susana Fernández, David Manzano
- 15.30 **Predictions with Confidence in Applications**
Mikhail Dashevskiy, Zhiyuan Luo
- 15.50 Break

Data Mining in Medicine

Session Chair: Petra Perner

- 16.05 **Aligning Bayesian Network Classifiers with Medical Contexts**
Linda C. van der Gaag, Silja Renooij, Ad Feelders, Arend de Groote, Marinus J.C. Eijkemans, Frank J. Broekmans, Bart C.J.M. Fauser
- 16.35 **Assessing the Eligibility of Kidney Transplant Donors**
Francisco Reinaldo, Carlos Fernandes, Md. Anishur Rahman, Andreia Malucelli and Rui Camacho
- 16.55 **Lung Nodules Classification in CT Images Using Simpson's Index, Geometrical Measures and One-Class SVM**
Cleriston Araujo da Silva, Aristófanés Corrêa Silva, Stelmo Magalhães, Barros Nett, Anselmo Cardoso de Paiva, Geraldo Braz Junior, Rodolfo Acatauassú Nunes
- 17.15 **The SSK29 algorithm for the early detection of ovarian cancer with increased sensitivity**
Brian Burford
- 17.35 **Discussion and Best Paper Award**
- 18.00 **End of Conference**
Farewell

Poster Session

Aspects of Data Mining

Building Classifiers to Identify Split Files

Pam Green, Peter C.R. Lane, Austen Rainer, Sven-Bodo Scholz

Data Mining Application in Goldbach Conjecture

Yanbo J. Wang

A Visualized Data Analysis Framework for Online-Shopping Products

Chung-Chian Hsu, Jing-Wei Lin, Chien-Hsin Chen

Visual Data Mining of Agriculture Data

Georg Ruß, Rudolf Kruse, Martin Schneider, Peter Wagner

Data Mining on Multimedia Data

SVM Based Action Scene Detection

Liang-Hua Chen, Chih-Wen Su, Chi-Feng Weng, Hong-Yuan Liao

A Two-fold PCA-Approach for Inter-Individual Recognition of Emotions in Natural Walking

Michelle Karg, Robert Jenke, Kolja Kühnlenz, Martin Buss

A Semantic Framework for Mining Novel Knowledge from Multimedia Data

Sara Colantonio, Igor Gurevich, Ovidio Salvetti, Yulia Trusova

Some Improvements of the ILP System HYPER Used in the Context of Robotic Learning

Joona Laukkanen, Georgi Stojanov, Vesna Poprcova

Theory

Robust Supervised Classifiers for Nondegenerate Generalized Gaussian Pattern Distributions with a Bounded Variance

Georgy Shevlyakov, Maya Shevlyakova, Vladimir Shin

How much Testing is Enough? A Learning Approach

Asem Eltaher, Markus Maurer, Thomas Form, Mohamed Ayeb

Performance Gain of Different Parallelization Approaches for Growing Neural Gas

Alexander Adam, Sebastian Leuoth, Wolfgang Benn

The Best Combining of Classifiers with Prior Probabilities

Zhong-Liang Fu, Yu Yao, Xiang-Hui Zhao

Approximation of the Parameters of a Readability Formula by Robust Regression

Tim von der Brück

Tree kernel and Feature Vector Methods for Formal Semantic Requests Classification

François Bouchet, Jean-Paul Sansonnet

Towards SISI – a Self Adaptive Insert Strategy for the Intelligent Cluster Index (ICIX)

Sebastian Leuoth, Wolfgang Benn

Optimization of the Weighting Exponent in the Fuzzy c-means Clustering

Liyun Chu, W.K. Wong, S.Y.S. Leung

Distance and Similarity

Clustering and Separating of a Set of Members in Terms of Mutual Distances and Proximities
Sergey D. Dvoenko