

# **2014 IEEE Symposium on Computational Intelligence and Data Mining**

**(CIDM 2014)**

**Orlando, Florida, USA  
9-12 December 2014**



**IEEE Catalog Number: CFP14IDM-POD  
ISBN: 978-1-4799-4517-7**

# TABLE OF CONTENTS

IEEE SSCI 2014 General Committee .....	viii
IEEE CIDM 2014 Committee .....	ix
Keynote Talk: CIDM'14 .....	xii

Wednesday, December 10

9:20AM-10:00AM

**CIDM'14 Keynote Talk**

*What Might be Predicted from Medical Image Minin*  
Lawrence Hall

10:20AM-12:00PM

**CIDM'14 Session 1: Advances in clustering**

Chair: Barbara Hammer

10:20AM	<i>Clustering data over time using kernel spectral clustering with memory</i> .....	1
	Rocco Langone, Raghvendra Mall and Johan A. K. Suykens	
10:40AM	<i>Agglomerative Hierarchical Kernel Spectral Data Clustering</i> .....	9
	Raghvendra Mall, Rocco Langone and Johan Suykens	
11:00AM	<i>Quantum Clustering -- A Novel Method for Text Analysis</i> .....	17
	Ding Liu, Minghu Jiang and Xiaofang Yang	
11:20AM	<i>Generalized Information Theoretic Cluster Validity Indices for Soft Clusterings</i> .....	24
	Yang Lei, James C. Bezdek, Jeffrey Chan, Nguyen Xuan Vinh, Simone Romano and James Bailey	
11:40AM	<i>A Density-Based Clustering of the Self-Organizing Map Using Graph Cut</i> .....	32
	Leonardo Enzo Brito da Silva and Jose Alfredo Ferreira Costa	

1:30PM-3:10PM

**CIDM'14 Session 2: Multitask and Metalearning**

Chair: Rocco Langone

1:30PM	<i>New Bilinear Formulation to Semi-Supervised Classification Based on Kernel Spectral Clustering</i> .....	41
	Vilen Jumutc and Johan Suykens	
1:50PM	<i>Batch Linear Least Squares-based Learning Algorithm for MLMVN with Soft Margins</i> .....	48
	Evgeni Aizenberg and Igor Aizenberg	
2:10PM	<i>Comparing Datasets by Attribute Alignment</i> .....	56
	Jakub Smid and Roman Neruda	
2:30PM	<i>Convex Multi-task Relationship Learning using Hinge Loss</i> .....	63
	Anveshi Charuvaka and Huzefa Rangwala	
2:50PM	<i>Precision-Recall-Optimization in Learning Vector Quantization Classifiers for Improved Medical Classification Systems</i> .....	71
	Thomas Villmann, Marika Kaden, Mandy Lange, Paul Stuermer and Wieland Hermann	

### 3:30PM-5:10PM

#### Special Session: CIDM'14 Session 3: Computational Intelligence for Health and Wellbeing

Chair: Paulo Lisboa

- 3:30PM *BioHCDP: A Hybrid Constituency-Dependency Parser for Biological NLP Information Extraction*..... 78  
Kamal Taha and Mohammed Al Zaabi
- 3:50PM *Classification of iPSC Colony Images Using Hierarchical Strategies with Support Vector Machines*..... 86  
Henry Joutsijoki, Jyrki Rasku, Markus Haponen, Ivan Baldin, Yulia Gizatdinova, Michelangelo Paci, Jyri Saarikoski, Kirsi Varpa, Harri Siirtola, Jorge Avalos-Salguero, Kati Iltanen, Jorma Laurikkala, Kirsi Penttinen, Jari Hyttinen, Katriina Aalto-Setälä and Martti Juhola
- 4:10PM *Semi-supervised source extraction methodology for the nosological imaging of glioblastoma response to therapy* ..... 93  
Sandra Ortega-Martorell, Ivan Olier, Teresa Delgado-Goni, Magdalena Ciezka, Margarida Julia-Sape, Paulo Lisboa and Carles Arus
- 4:30PM *Automatic relevance source determination in human brain tumors using Bayesian NMF*..... 99  
Sandra Ortega-Martorell, Ivan Olier, Margarida Julia-Sape, Carles Arus and Paulo Lisboa
- 4:50PM *Alzheimer's disease patients classification through EEG signals processing*..... 105  
Giulia Ficon, Emanuel Weitschek, Giovanni Felici, Paola Bertolazzi, Simona De Salvo, Placido Bramanti and Maria Cristina De Cola

### Thursday, December 11

#### 10:20AM-12:00PM

#### CIDM'14 Session 4: Mining Relational and Networked data

Chair: John Lee

- 10:20AM *Relational Data Partitioning using Evolutionary Game Theory* ..... 113  
Lawrence O. Hall and Alireza Chakeri
- 10:40AM *Aggregating Predictions vs. Aggregating Features for Relational Classification* ..... 121  
Oliver Schulte and Kurt Routley
- 11:00AM *Ontology Learning with Complex Data Type for Web Service Clustering* ..... 129  
B. T. G. S. Kumara, Incheon Paik, K. R. C. Koswatte and Wuhui Chen
- 11:20AM *Semantic clustering-based cross-domain recommendation* ..... 137  
Anil Kumar, Nitesh Kumar, Muzammil Hussain, Santanu Chaudhury and Sumeet Agarwal
- 11:40AM *Distributed Evolutionary Approach To Data Clustering and Modeling*..... 142  
Mustafa Hajeer, Dasgupta Dipankar, Alexander Semenov and Jari Veijalainen

#### 1:30PM-3:10PM

#### Special Session: CIDM'14 Session 5: High Dimensional Data Analysis

Chair: Thomas Villmann

- 1:30PM *Valid Interpretation of Feature Relevance for Linear Data Mappings* ..... 149  
Benoit Frenay, Daniela Hofmann, Alexander Schulz, Michael Biehl and Barbara Hammer
- 1:50PM *High Dimensional Exploration: A Comparison of PCA, Distance Concentration, and Classification Performance in two fMRI Datasets* ..... 157  
Joset Etzel and Todd Braver
- 2:10PM *Two key properties of dimensionality reduction methods*..... 163  
John A. Lee and Michel Verleysen
- 2:30PM *Generalized kernel framework for unsupervised spectral methods of dimensionality reduction* ..... 171  
Diego Hernan Peluffo-Ordóñez, John Aldo Lee and Michel Verleysen

2:50PM	<i>Evaluating Topic Quality using Model Clustering</i> .....	178
	Vineet Mehta, Rajmonda Caceres, Kevin Carter	

**3:30PM-5:10PM**

**CIDM'14 Session 6: Rule based Modelling, Model Performance, and Interpretability**

Chair: Oliver Schulte

3:30PM	<i>Optimization of the Type-1 and Interval Type-2 Fuzzy Integrators in Ensembles of ANFIS models for Prediction of the Dow Jones Time Series</i> .....	186
	Jesus Soto, Patricia Melin and Oscar Castillo	
3:50PM	<i>Accurate and Interpretable Regression Trees using Oracle Coaching</i> .....	194
	Ulf Johansson, Cecilia Sonstrod and Rikard Konig	
4:10PM	<i>Product Aspect Identification: Analyzing Role of Different Classifiers</i> .....	202
	Xing Yu, Sukanya Manna and Brian N Truong	
4:30PM	<i>Rule Extraction using Genetic Programming for Accurate Sales Forecasting</i> .....	210
	Rikard Konig and Ulf Johansson	
4:50PM	<i>Facial Image Clustering in Stereo Videos Using Local Binary Patterns and Double Spectral Analysis</i> .....	217
	Georgios Orfanidis, Anastasios Tefas, Nikos Nikolaidis and Ioannis Pitas	

**5:10PM-6:45PM**

**Poster Session: SSCI'14 Poster Session**

P132	<i>Weighted Feature-based Classification of Time series Data</i> .....	222
	Ravikumar Penugonda and V. Susheela Devi	
P133	<i>Gender classification of subjects from cerebral blood flow changes using Deep Learning</i> .....	229
	Tomoyuki Hiroyasu, Kenya Hanawa and Utako Yamamoto	
P134	<i>A feature transformation method using genetic programming for two-class classification</i> .....	234
	Tomoyuki Hiroyasu, Toshihide Shiraishi, Tomoya Yoshida and Utako Yamamoto	
P135	<i>Dependency Network Methods for Hierarchical Multi-label Classification of Gene Functions</i> .....	241
	Fabio Fabris and Alex A. Freitas	
P136	<i>A Novel Criterion for Overlapping Communities Detection and Clustering Improvement</i> .....	249
	Alessandro Berti, Alessandro Sperduti and Andrea Burattin	
P137	<i>Incremental Transfer RULES with Incomplete Data</i> .....	257
	Hebah ElGibreen and Mehmet Sabih Aksoy	
P138	<i>Novelty Detection Applied to the Classification Problem Using Probabilistic Neural Network</i> .....	265
	Balvant Yadav and V. Susheela Devi	
P139	<i>A Framework for Initialising a Dynamic Clustering Algorithm: ART2-A</i> .....	273
	Simon Chambers, Ian Jarman and Paulo Lisboa	
P140	<i>Recommendation for Web Services with Domain Specific Context Awareness</i> .....	281
	B. T. G. S. Kumara, Incheon Paik, K. R. C. Koswatte and Wuhui Chen	
P141	<i>Tibetan-Chinese Cross Language Named Entity Extraction Based on Comparable Corpus and Naturally Annotated Resources</i> .....	288
	Yuan Sun, Wenbin Guo and Xiaobing Zhao	

P142	<b><i>Detecting and profiling sedentary young men using machine learning algorithms</i></b> .....	<b>296</b>
	Pekka Siirtola, Riitta Pyky, Riikka Ahola, Heli Koskimaki, Timo Jamsa, Raija Korpelainen and Juha Roning	
P143	<b><i>Patient Level Analytics Using Self-Organising Maps: A Case Study on Type-1 Diabetes Self-care Survey Responses</i></b> .....	<b>304</b>
	Santosh Tirunagari, Norman Poh, Kouros Aliabadi, David Windridge and Deborah Cooke	
P144	<b><i>Interpolation and Extrapolation: Comparison of Definitions and Survey of Algorithms for Convex and Concave Hulls</i></b> .....	<b>310</b>
	Tobias Ebert, Julian Belz and Oliver Nelles	
P145	<b><i>Takagi-Sugeno-Kang Type Collaborative Fuzzy Rule Based System</i></b> .....	<b>315</b>
	Kuang-pen Chou, Mukesh Prasad, Yang-Yin Lin, Sudhanshu Joshi, Chin-Teng Lin and Jyh-Yeong Chang	
P146	<b><i>Recognizing Gym Exercises Using Acceleration Data from Wearable Sensors</i></b> .....	<b>321</b>
	Heli Koskimaki and Pekka Siirtola	
P147	<b><i>What can Spatial Collectives tell us about their environment?</i></b> .....	<b>329</b>
	Zena Wood	
P148	<b><i>Weighted One-Class Classification for Different Types of Minority Class Examples in Imbalanced Data</i></b> .....	<b>337</b>
	Bartosz Krawczyk, Michal Wozniak and Francisco Herrera	
P149	<b><i>A Sparsity-Based Training Algorithm for Least Squares SVM</i></b> .....	<b>345</b>
	Jie Yang and Jun Ma	
P150	<b><i>Wolf Search Algorithm for Attribute Reduction in classification</i></b> .....	<b>351</b>
	Waleed Yamany, Eid Emary and Aboul Ella Hassanien	
P151	<b><i>Alarm prediction in industrial machines using autoregressive LS-SVM models</i></b> .....	<b>359</b>
	Rocco Langone, Carlos Alzate, Abdellatif Bey-Temsamani and Johan A. K. Suykens	
P152	<b><i>Sensor dynamics in high dimensional phase spaces via nonlinear transformations: Application to helicopter loads monitoring</i></b> .....	<b>365</b>
	Julio Valdes, Catherine Cheung and Matthew Li	
P153	<b><i>Automatic Text Categorization Using a System of High-Precision and High-Recall Models</i></b> .....	<b>373</b>
	Dai Li, Yi Murphey and Huang Yinghao	

**Friday, December 12**

**9:30AM-10:30AM**

**Special Session: CIDM'14 Session 7: Business Process Mining, Market Analysis and Process Big Data**

Chair: Andrea Burattin

9:30AM	<b><i>The Use of Process Mining in a Business Process Simulation Context: Overview and Challenges</i></b> .....	<b>381</b>
	Niels Martin, Benoit Depaire and An Caris	
9:50AM	<b><i>Discovering Cross-Organizational Business Rules from the Cloud</i></b> .....	<b>389</b>
	Mario Luca Bernardi, Marta Cimitile and Fabrizio Maggi	
10:10AM	<b><i>GoldMiner: A Genetic Programming based algorithm applied to Brazilian Stock Market</i></b> .....	<b>397</b>
	Alexandre Pimenta, Eduardo Carrano, Ciniro Nametala, Frederico Guimaraes and Ricardo Takahashi	

**11:00AM-12:00PM**

**CIDM'14 Session 8: Educational Data Mining**

Chair: Alexander Schulz

11:00AM	<i>FATHOM: A Neural Network-based Non-verbal Human Comprehension Detection System for Learning Environments</i> .....	403
	Fiona Buckingham, Keeley Crockett, Zuhair Bandar and James O'Shea	
11:20AM	<i>Predicting Student Success Based on Prior Performance</i> .....	410
	Ahmad Slim, Gregory Heileman, Jarred Kozlick and Chaouki Abdallah	
11:40AM	<i>To What Extend Can We Predict Students' Performance? A Case Study in Colleges in South Africa</i> .....	416
	Norman Poh and Ian Smythe	

**1:30PM-3:10PM**

**CIDM'14 Session 9: Modelling and Mining Massive Data Sets**

Chair: Jean-Marc Andreoli

1:30PM	<i>Matching Social Network Biometrics Using Geo-Analytical Behavioral Modeling</i> .....	422
	Mark Rahmes, Kevin Fox, John Delay and Gran Roe	
1:50PM	<i>Massively Parallelized Support Vector Machines based on GPU-Accelerated Multiplicative Updates</i> .....	431
	Connie (Khor Li) Kou and Chao-Hui Huang	
2:10PM	<i>Scaling a Neyman-Pearson Subset Selection Approach Via Heuristics for Mining Massive Data</i> .....	439
	Gregory Ditzler, Matthew Austen, Gail Rosen and Robi Polikar	
2:30PM	<i>MapReduce Guided Approximate Inference Over Graphical Models</i> .....	446
	Ahsanul Haque, Swarup Chandra, Latifur Khan and Michael Baron	
2:50PM	<i>Optimization of Relational Database Usage Involving Big Data (A Model Architecture for Big Data applications)</i> .....	454
	Erin-Elizabeth Durham, Andrew Rosen and Robert Harrison	

**3:30PM-5:10PM**

**CIDM'14 Session 10: Advanced signal processing and data analysis**

Chair: Barbara Hammer

3:30PM	<i>Learning Energy Consumption Profiles from Data</i> .....	463
	Jean-Marc Andreoli	
3:50PM	<i>kNN estimation of the unilateral dependency measure between random variables</i> .....	471
	Angel Cataron, Razvan Andonie and Yvonne Chueh	
4:10PM	<i>Using Data Mining to Investigate Interaction between Channel Characteristics and Hydraulic Geometry Channel Types</i> .....	479
	Leong Lee and Gregory S. Ridenour	
4:30PM	<i>Experimental Studies on Indoor Sign Recognition and Classification</i> .....	489
	Zhen Ni, Siyao Fu, Bo Tang, Haibo He and Xinming Huang	
4:50PM	<i>High-SNR Model Order Selection Using Exponentially Embedded Family and Its Applications to Curve Fitting and Clustering</i> .....	495
	Quan Ding, Steven Kay and Xiaorong Zhang	

<b>AUTHOR INDEX</b> .....	<b>501</b>
---------------------------	------------