The IADIS European Conference on Data Mining 2009 took place in Algarve, Portugal, 18-20 June, 2009. This conference was part of the Multi Conference on Computer Science and Information Systems 2009 (MCCSIS), 17 - 23 June 2009, which had a total of 1131 submissions.

The IADIS European Conference on Data Mining (ECDM'09) aimed to gather researchers and application developers from a wide range of data mining related areas such as statistics, computational intelligence, pattern recognition, databases and visualization. ECDM’09 had the goal to advance the state of the art in data mining field and its various real world applications. It provided opportunities for technical collaboration among data mining and machine learning researchers around the globe.

The IADIS European Conference on Data Mining 2009 received 63 submissions from more than 19 countries. Each submission had been anonymously reviewed by an average of four independent reviewers, to ensure that the final accepted submissions were of a high standard. Consequently only 14 full papers were published which meant an acceptance rate of about 22%. A few more papers were accepted as short papers, reflection papers and posters. The best paper authors were invited to publish an extended version of their paper in the IADIS International Journal on Computer Science and Information Systems (ISSN: 1646-3692) and also in other selected journals, including journals from Inderscience.

The submissions were accepted under the following areas of interest:

**Core Data Mining Topics**
- Parallel and distributed data mining algorithms
- Data streams mining
- Graph mining
- Spatial data mining
- Text video, multimedia data mining
- Web mining
- Pre-processing techniques
- Visualization
- Security and information hiding in data mining

**Data Mining Applications**
- Databases
- Bioinformatics
- Biometrics
- Image analysis
- Financial modeling
- Forecasting
- Classification
- Clustering

Besides the presentation of full papers, short papers, reflection papers and posters, the conference also included two keynote presentations from internationally distinguished researchers, Professor Kurosh Madani, Images, Signals and Intelligence Systems Laboratory (LISSI / EA 3956) PARIS XII University, Senart-Fontainebleau Institute of Technology, France and Dr. Claude C. Chibelushi, Faculty of Computing, Engineering & Technology, Staffordshire University, United Kingdom. In addition, the Conference also offered a tutorial by Professor Kurosh Madani.


Overall the Conference offered an opportunity to all their participants to discuss with success the most significant aspects regarding the theme Data Mining. It served as a forum that gathered researchers, practitioners, students and anyone that was working or studying in the field of the Data Mining.
Keynote Presentations:

**K.1 – TOWARD HIGHER LEVEL OF INTELLIGENT SYSTEMS FOR COMPLEX DATA PROCESSING AND MINING** by Professor Kurosh Madani, Images, Signals and Intelligence Systems Laboratory (LISSI / EA 3956), PARIS XII University, Senart-Fontainebleau Institute of Technology, France

**Abstract:**

Real world applications and especially those dealing with complex data mining ones make quickly appear the insufficiency of academic (called also sometime theoretical) approach in solving such categories of problems. The difficulties appear since definition of the “problem’s solution” notion. In fact, academic approaches often begin by problem’s constraints simplification in order to obtain a “solvable” model (here, solvable model means a set of mathematically solvable relations or equations describing a processing flow, a behavior, a set of phenomena, etc...). If the theoretical consideration is a mandatory step to study a given problem's solvability, for a very large number of real world dilemmas, it doesn't lead to a solvable or realistic solution. Difficulty could be related to several issues among which:

- large number of parameters to be taken into account making conventional mathematical tools inefficient,
- strong nonlinearity of the data (describing a complex behavior or ruling relationship between involved data), leading to unsolvable equations,
- partial or total inaccessibility to relevant features (relevant data), making the model insignificant,
- subjective nature of relevant features, parameters or data, making the processing of such data or parameters difficult in the frame of conventional quantification,
- necessity of expert’s knowledge, or heuristic information consideration,
- imprecise information or data leakage.

Examples illustrating the above-mentioned difficulties are numerous and may concern various areas of real world or industrial applications. As first example, one can emphasize difficulties related to economical and financial modeling (data mining, features' extraction and prediction), where the large number of parameters, on the one hand, and human related factors, on the other hand, make related real world problems among the most difficult to solve. Another illustrative example concerns the delicate class of dilemmas dealing with complex data's and multifaceted information's processing, especially when processed information (representing patterns, signals, images, etc.) are strongly noisy or involve deficient data. In fact, real world and industrial applications, comprising image analysis, systems and plants safety, complex manufacturing and processes optimization, priority selection and decision,, classification and clustering are often those belonging to such class of dilemmas.

If much is still to discover about how the animal’s brain trains and self-organizes itself in order to process and mining so various and so complex information, a number of recent advances in “neurobiology” allow already highlighting some of key mechanisms of this marvels machine. Among them one can emphasizes brain’s “modular” structure and its “self-organizing” capabilities. In fact, if our simple and inappropriate binary technology remains too primitive to achieve the processing ability of these marvels mechanisms, a number of those highlighted points could already be sources of inspiration for designing new machine learning approaches leading to higher levels of artificial systems’ intelligence.

This plenary talk deals with machine learning based modular approaches which could offer powerful solutions to overcome processing difficulties in the aforementioned frame. It focuses machine learning based modular approaches which take advantage from self-organizing multi-modeling (“divide and conquer” paradigm). If the machine learning capability provides processing system's adaptability and offers an appealing alternative for fashioning the processing technique adequacy, the modularity may result on a substantial reduction of treatment’s complexity. In fact, the modularity issued complexity reduction may be obtained from several instances: it may result from distribution of computational effort on several modules (multi-modeling and macro parallelism); it can emerge from cooperative or concurrent contribution of several processing modules in handling a same task (mixture of experts); it may drop from the modules’ complementary contribution (e.g. specialization of a module on treating a given task to be performed).

One of the most challenging classes of data processing and mining dilemmas concerns the situation when no a priori information (or hypothesis) is available. Within this frame, a self-organizing modular machine learning approach, combining “divide and conquer” paradigm and “complexity estimation” techniques called self-organizing “Tree-like Divide To Simplify” (T-DTS) approach will be described and evaluated.
K.2 – HCI THROUGH THE ‘HC EYE’ (HUMAN-CENTRED EYE):
CAN COMPUTER VISION INTERFACES EXTRACT THE
MEANING OF HUMAN INTERACTIVE BEHAVIOUR?

by Dr. Claude C. Chibelushi, Faculty of Computing, Engineering & Technology,
Staffordshire University, United Kingdom

Abstract:

Some researchers advocating a human-centred computing perspective have been investigating new methods for interacting with computer systems. A goal of these methods is to achieve natural, intuitive and effortless interaction between humans and computers, by going beyond traditional interaction devices such as the keyboard and the mouse. In particular, significant technical advances have been made in the development of the next generation of human computer interfaces which are based on processing visual information captured by a computer. For example, existing image analysis techniques can detect, track and recognise humans or specific parts of their body such as faces and hands, and they can also recognise facial expressions and body gestures.

This talk will explore technical developments and highlight directions for future research in digital image and video analysis which can enhance the intelligence of computers by giving them, for example, the ability to understand the meaning of communicative gestures made by humans and recognise context-relevant human emotion. The talk will review research efforts towards enabling a computer vision interface to answer the what, when, where, who, why, and how aspects of human interactive behaviour. The talk will also discuss the potential impacts and implications of technical solutions to problems arising in the context of human computer interaction. Moreover, it will suggest how the power of the tools built onto these solutions can be harnessed in many realms of human endeavour.

Best Papers:

- AN EXPERIMENTAL STUDY OF THE DISTRIBUTED CLUSTERING FOR AIR POLLUTION PATTERN RECOGNITION IN SENSOR NETWORKS by Yajie Ma, Wuhan University of Science and Technology, China

Abstract:

In this paper, we make an experimental study of the urban air pollution pattern analysis within MESSAGE system. A hierarchical network framework consisted of mobile sensors and stationary sensors is designed. A sensor gateway core architecture is developed which is suited to grid-based computation. Then we make experimental analysis including the identification of pollution hotspots and the dispersion of pollution clouds based on a real-time peer-to-peer clustering algorithm. Our results provide a typical air pollution pattern in urban environment which gives a real-time track of the air pollution variation.

- CONTINUOUS-TIME HIDDEN MARKOV MODELS FOR THE COPY NUMBER ANALYSIS OF GENOTYPING ARRAYS by Matthew Kowgier and Rafal Kustra, Dalla Lana School of Public Health, University of Toronto, Canada

Abstract:

We present a novel Hidden Markov Model for detecting copy number variations (CNV) from genotyping arrays. Our model is a novel application of HMM to inferring CNVs from genotyping arrays: it assumes a continuous time framework and is informed by prior findings from previously analysed real data. This framework is also more realistic than discrete-time models which are currently used since the underlying genomic sequence is few hundred times denser than the array data. We show how to estimate the model parameters using a training data of normal samples whose CNV regions have been confirmed, and present results from applying the model to a set of HapMap samples containing aberrant SNPs.
Committees:

Program Chair: Ajith P. Abraham, School of Computer Science, Chung-Ang University, South Korea

Conference Co-Chairs:

- Piet Kommers, University of Twente, The Netherlands
- Pedro Isaias, Universidade Aberta (Portuguese Open University), Portugal
- Nian-Shing Chen, National Sun Yat-sen University, Taiwan

Committee Members:

- Abdel-Badeeh M. Salem, Ain Shams University, Egypt
- Akihiro Inokuchi, Osaka University, Japan
- Alessandra Raffaetà, Università Ca’ Foscari di Venezia, Italy
- Alexandros Nanopoulos, University of Hildesheim, Germany
- Alfredo Cuzzocrea, University of Calabria, Italy
- Anastasios Dimou, Informatics and Telematics Institute, Greece
- Andreas König, TU Kaiserslautern, Germany
- Annalisa Appice, Università degli Studi di Bari, Italy
- Arnab Bhattacharya, I.I.T. Kanpur, India
- Artchil Maysuradze, Moscow University, Russia
- Ben Kao, The University of Hong Kong, Hong Kong
- Carson Leung, University of Manitoba, Canada
- Chao Luo, University of Technology, Sydney, Australia
- Christos Makris, University of Patras, Greece
- Claudio Lucchese, Università Ca’ Foscari di Venezia, Italy
- Claudio Silvestri, Università di Ca’ Foscari di Venezia, Italy
- Dan Wu, University of Windsor, Canada
- Daniel Kudenko, University of York, UK
- Daniel Pop, University of the West Timisoara, Romania
- Daniela Zaharie, West University of Timisoara, Romania
- Daoqiang Zhang, Nanjing University of Aeronautics and Astronautics, China
- David Cheung, University of Hong Kong, Hong Kong
- Dimitrios Katsaros, University of Thessaly, Greece
- Dino Ienco, Università di Torino, Italy
- Edward Hung, Hong Kong Polytechnic University, Hong Kong
- Eugenio Cesario, Università della Calabria, Italy
- Fotis Lazarinis, Technological Educational Institute, Greece
- Francesco Folino, University of Calabria, Italy
- Gabriela Kokai, Friedrich Alexander University, Germany
- George Pallis, University of Cyprus, Cyprus
- Georgios Yannakakis, IT-University of Copenhagen, Denmark
- Hamed Nassar, Suez Canal University, Egypt
- Harish Karnick, IIT Kanpur, India
- Hui Xiong, Rutgers University, USA
- Ingrid Fischer, University of Konstanz, Germany
- Ioannis Kopanakis, Technological Educational Institute of Crete, Greece
- Jason Wang, New Jersey Institute of Technology, USA
- Jia-Yu Pan, Google Inc., USA
- Jialie Shen, Singapore Management University, Singapore
- John Kouris, University of Patras, Greece
- José M. Peña, Technical University of Madrid, Spain
- Jun Huan, University of Kansas, USA
- Junjie Wu, Beijing University of Aeronautics and Astronautics, China
- Justin Dauwels, MIT, USA