The 2nd International Workshop on Inductive Modelling (IWIM07) will be held in Prague on September 23-26, 2007. The initial Workshop took place in Kyiv, Ukraine, in July 2005 following the International Conference on Inductive Modelling (ICIM'2002) in Lviv, Ukraine, in May 2002. The series of conferences and workshops is the only international forum that focuses on theory, algorithms, applications, solutions, and new developments of data mining and knowledge extraction technologies which originate from Ivakhnenko’s Group Method of Data Handling (GMDH) as a typical inductive modelling method. Built on principles of self-organization inductive modelling has been developing and using in several key areas for over 30 years now and can be found in data mining technologies like Polynomial Neural Networks, Adaptive Learning Networks, or Statistical Learning Networks. More recent developments also utilize Genetic Algorithms or the idea of Active Neurons and multi-leveled self-organization to build models from data.

The motivation of this 2nd workshop is to analyze the state-of-the-art of modelling methods that inductively generate models from data, to discuss concepts of an automated knowledge discovery workflow, to share new ideas on model validation and visualization, to present novel applications in different areas, and to give inspiration and background on how inductive modelling can evolve and contribute given the current global challenges.
# Table of Contents

- Welcome and Organizing Committee ................................................. 1
- Agenda at a Glance ........................................................................ 4
- Opening Ceremony ........................................................................ 5
- Tutorials ....................................................................................... 6
- Welcome Reception ......................................................................... 7
- Session: Principles & Theoretical Foundations I ............................ 8
- Session: Principles & Theoretical Foundations II & High Performance Computing ................................................................. 9
- Session: Inductive Modeling vs. Computational Intelligence ........ 10
- Session: Real-world Applications I ................................................ 11
- Session: Optimal Complexity of inductive Models I ...................... 13
- Session: Optimal Complexity of inductive Models II ...................... 14
- Session: Social Event .................................................................... 14
- Session: Real-world Applications II .............................................. 15
- Session: Data Mining ..................................................................... 16
- Session: Optimization & Evolution of Inductive Models and Neural Networks I ................................................................. 17
- Session: Time Series ..................................................................... 18
- Session: Optimization & Evolution of Inductive Models and Neural Networks II ................................................................. 19
- Session: Optimization & Evolution of Inductive Models and Neural Networks III ................................................................. 19
- Session: Posters ............................................................................ 20
- Session: Plenary Session ................................................................. 22
- Session: Closing Ceremony ............................................................. 23
**Organizing Committee**

**Honorary Chair**
Alexey Ivakhnenko  
Professor, Academician of NAS of Ukraine

**Workshop Chair**
**Volodymyr Stepanenko**  
Professor, Dr. Sci  
stepanenko@irtc.org.ua  
International Research and Training Center  
of Information Technologies and Systems  
of the NAS and MES of Ukraine

**Workshop Co-Chair**
**Godfrey Onwubolu**  
Professor of Engineering  
onwubolu_g@usp.ac.fj  
School of Engineering & Physics  
Faculty of Science & Technology  
University of the South Pacific

**Local Chair**
**Miroslav Šnorek**  
Professor  
snorek@fel.cvut.cz  
Department of Computer Science and Engineering  
Faculty of Electrical Engineering  
Czech Technical University in Prague

**Secretary**
**Pavel Kordík**  
Assistant Professor  
kordikp@fel.cvut.cz  
Department of Computer Science and Engineering,  
Faculty of Electrical Engineering,  
Czech Technical University in Prague

**Program Committee**
Volodymyr Stepashko  
Cybernetics Centre, Kyiv, Ukraine  
Godfrey Onwubolu  
University of the South Pacific, Suva, Fiji  
Miroslav Šnorek  
Czech Technical University, Czech Republic  
Hitoshi Iba  
The University of Tokyo, Japan  
Witold Pedrycz  
University of Alberta, Canada  
Sung-Kwon Oh  
Wookwang University, South Korea  
Tadashi Kondo  
The University of Tokushima, Japan  
Nikolay I. Nikolaev  
University of London, UK  
Tatyana Aksyonova  
Grenoble University, France  
Frank Lemke  
KnowledgeMiner Software, Berlin, Germany  
Pavel Kordík  
Czech Technical University, Czech Republic  
He Changzheng  
Sichuan University, Chengdu, China  
N. Nariman-Zadeh  
The University of Gilan, Iran  
Hambarsum Bozdogan  
The University of Tennessee, USA  
Panos Liatsis  
City University London, UK  
John Elder  
Elder Research, USA

**Local Organizing Committee**
Miroslav Šnorek  
Pavel Kordík  
Jan Koutník  
Miroslav Skrbek  
Miroslav Čepek  
Pavel Náplava  
Petr Burian  
Zdeněk Buk  
Rudolf Marek  
Jan Drchal
Agenda at a Glance

Saturday September 22nd 2007:
Arrival of participants, accommodation in Masaryk college and other hotels, sight seeing Prague at night.

Sunday September 23rd 2007: Tutorials Day
8:30 – 9:30: Registration of participants
9:30 Opening ceremony
10:00 – 10:45 TUTORIAL 1, Volodomyr Stepashko:
GMDH-based Inductive Modelling: Theoretical, Structural, and Applied Aspects

10:45 – 11:30 TUTORIAL 2, Frank Lemke:
Multi-leveled Self-organization

Lunch
12:30 – 13:15 TUTORIAL 3, Pavel Kordík:
GMDH and FAKE GAME: Evolving ensembles of inductive models

13:15 – 14:00 TUTORIAL 4, John Elder:
Top 10 Inductive Modeling Mistakes

14:00 – 14:15 Coffee break
14:15 – 16:30 Plenary Session
18:00 – 21:00 Welcome reception

Monday September 24th 2007:
9:30 – 11:30 2 parallel sections of contribution presentations
Lunch
12:30 – 14:30 2 parallel sections of contribution presentations
Coffee break
15:00 – 16:30 2 parallel sections of contribution presentations
19:00 – 22:00 Social event – boat cruise on the Vltava river

Tuesday September 25th 2007:
9:30 – 11:30 2 parallel sections of contribution presentations
Lunch
12:30 – 14:30 2 parallel sections of contribution presentations
Coffee break
15:00 – 15:30 2 parallel sections of contribution presentations
15:30 – 16:30 Poster section and backstage discussions
18:00 Individual program

Wednesday September 26th 2007:
9:30 – 11:00 Plenary Session
11:00 – 11:30 Closing ceremony
Lunch
Session: Opening Ceremony

Time and place: Sunday, September 23, 2007: 9:30—10:00  Gallery

Session Chair: Volodomyr Stepashko

Session: Tutorials

Time and place: Sunday, September 23, 2007: 10:00—14:00  Gallery

Session Chair: Volodomyr Stepashko

10:00 Volodomyr Stepashko  TUTORIAL 1, GMDH-based Inductive Modelling: Theoretical, Structural, and Applied Aspects

The inductive modelling problem is characterized from theoretical, structural, and applied points of view. Main pressing tasks are stressed. The problem of an optimal model construction from a given data sample is investigated. The main goal of the problem solving is the choice of the model structure with minimal variance of the prediction error, or a noise-immunity model. Regulations of the optimal model building are investigated depending on the noise level and sample size. The efficiency of external GMDH criteria in the problem is studied. Typical structures of GMDH algorithms are analyzed: combinatorial, selective (multilayered), and combined. Some ways to enhance the algorithms effectiveness are suggested. GMDH algorithms applications to solving economical, ecological, and other real world problems are discussed.

10:45 Frank Lemke  TUTORIAL 2, Multi-leveled Self-organization

This presentation will outline the concept of multi-leveled self-organization starting from Self-organizing Networks of Active Neurons, KnowledgeMiner's GMDH implementation. Then, the idea of systems of equations will be discussed followed by the introduction of the algorithm's noise sensitivity characteristic as a new, supplemental model validation tool. Finally, an algorithm for high-dimensional state
space modeling will be shown within the framework of an automated knowledge extraction from data. Several examples will illustrate discussed problems.

**12:30 Pavel Kordík**

**TUTORIAL 3, GMDH and FAKE GAME: Evolving ensembles of inductive models**

In this tutorial the state of the art of the GMDH worldwide research will be summarized. Then the idea to evolve ensemble of heterogeneous inductive models (GAME) will be presented. Applications of the GAME algorithm for various real-world data set are to be demonstrated. Finally the concept of Fully Automated Knowledge Extraction (FAKE) using GAME will be explained and possible applications of the FAKE GAME open source software will be illustrated.

**13:15 John Elder**

**TUTORIAL 4, Top 10 Inductive Modeling Mistakes**

Inducing models from data is powerful, but therefore also dangerous. Decisions can be spectacularly wrong if based on bad data or wrong assumptions. The technical modeling task typically dominating an analyst's imagination is surrounded by a "systems engineering" business problem of equal importance and often greater ambiguity. Dr. Elder will illuminate "best practices" for the full cycle of model induction by illustrating their opposite -- mistakes (both simple and subtle), using examples from diverse real-world consulting projects.

**Session: Plenary Session**

Time and place: Sunday, September 23, 2007: 14:15—16:30  Gallery

Session Chair: Frank Lemke

**14:15 Revised GMDH-Type Neural Network algorithms with Self-Selecting Optimum Architecture and Their Applications to Nonlinear System Identification and 3-Dimensional Medical Image Recognition**

Tadashi Kondo

Two kinds of revised Group Method of Data Handling (GMDH)-type neural network algorithms are proposed. One is a multi-layered GMDH-type self-selecting optimum neural network architecture and another is a feedback GMDH-type self-selecting optimum neural network architecture. These revised GMDH-type neural network algorithms have an ability of self-selecting optimum neural network architecture from three neural network architectures such as sigmoid function neural network, radial basis function (RBF) neural network and polynomial neural network. Revised GMDH-type neural networks also have abilities of self-selecting the structural parameters such as the number of layers, the number of neurons in hidden layers and useful input variables. These structural parameters are automatically selected so as to minimize the prediction error criterion defined as Prediction Sum of Squares (PSS) or Akaike’s Information Criterion (AIC). The multi-layered GMDH-type neural
network is applied to the nonlinear system identification problem and it is shown that this algorithm is very accurate and useful for the nonlinear system identification. Furthermore, the feedback GMDH-type neural network is applied to 3-dimensional medical image recognition of the lungs, the pulmonary vessels and the bronchial trees. The recognition results show that the feedback GMDH-type neural network algorithm is useful for the 3-dimensional medical image recognition of the lungs, the pulmonary vessels and the bronchial trees and is ideal for such practical complex problems since the optimum neural network architecture is automatically organized.

15:00 Theoretical and applied aspects of the hybrid immune systems development
Vladimir Litvinenko

Biologically based immune systems used adaptive mechanisms of functioning natural immune systems of vertebrates are described. The review of existing models (negative selection, clonal selection and an immune network) and approaches to development of applied immune systems is made. There are described hybrid systems for the forecasting problems on the basis of clonal selection operators and programming of genes expression, a cooperative synthesis algorithm of an immune network and synthesis radial basis neural networks, wavelet neural networks, and also a fuzzy neural network by means of the clonal algorithm.

15:45 Designing Hybrid Computational Intelligence Methodologies and Group Method of Data Handling for Inductive Modeling of Nonlinear Systems
Godfrey Onwubolu

The group method of data handling (GMDH) and computational intelligence methodology (CIM), which consists of evolutionary computing, fuzzy computing and neurocomputing are two well-known nonlinear methods of mathematical modeling of complex systems. Both methods are explained and a new design methodology which is a hybrid of GMDH and population-based CIM is proposed. The proposed method constructs a GMDH network model of a population of promising CIM solutions. The new hybrid implementation is then applied to modeling and prediction of practical nonlinear datasets as well as data mining datasets and the results are compared with the results obtained by other GMDH-related algorithms. Results presented show that the proposed algorithm appears to perform reasonably well and hence can be applied to real-life prediction and modeling problems as well as data mining problems. One major conclusion resulting from studies carried out in implementing hybrid CIM-GMDH network is that population-based computational intelligence methodologies (genetic programming [GP], genetic algorithm [GA], differential evolution [DE], scatter search [SS], ant colony system [ACS], particle swarm optimization [PSO], etc.) are all candidates of hybridization with GMDH. In the past, GA and GP have been mainly studied for hybridization with GMDH.

Session: Social Event (Welcome reception)
Time and place: Sunday, September 23, 2007: 18:00—21:00 Salaon
Session Chair: Pavel Kordík
Monday September 24th 2007:

<table>
<thead>
<tr>
<th>Track</th>
<th>9:30—11:30</th>
<th>12:30—14:30</th>
<th>15:00—16:30</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Principles &amp; theoretical foundations</td>
<td>High Performance Computing</td>
<td>Inductive Modeling vs. Computational Intelligence</td>
</tr>
<tr>
<td>3</td>
<td>Real-world Applications I</td>
<td>Optimal Complexity of inductive Modeling</td>
<td></td>
</tr>
</tbody>
</table>

Session: Principles & Theoretical Foundations I

Time and place: Monday, September 24, 2007: 9:30—11:30 Room 1
Session Chairs: John Elder & Jan Koutnik

**9:30** The Fuzzy Group Method of Data Handling with Fuzzy Input Variables
Yuriy Zaychenko

The problem of constructing forecasting models with incomplete and fuzzy input data is considered in this paper. For its solution Fuzzy Group Methods of Data Handling (FGMDH) with fuzzy inputs is suggested. The method enables to construct a forecasting fuzzy model using experimental data which are not distinct. The method was implemented as software kit and experimental investigations of were carried out in the problem forecasting stock-prices at the Russian stock-exchange. The comparison of the suggested method with known methods: GMDH and fuzzy GMDH is also presented.

**10:00** Enhanced MIA GMDH algorithm
Petr Buryan

The paper presents a new methodology which is an enhancement of MIA algorithm of self-organizing polynomial Group Method of Data Handling (GMDH). Classical MIA algorithm suffers mainly by quick loss of the layer diversity resulting in almost homogenous layer output which strongly prohibits reaching any improvement in the next layers of the network and has negative impacts on the stability of transfer functions of the nodes (gained by least mean squares method). The impact is mainly non-stable behaviour and loss in quality of the output of the model as a whole. Several specific improved features were therefore applied in order to improve the behaviour of the algorithm. The enhancements described bellow are mainly semi-randomized selection approach to layer pruning, coefficient rounding and thresholding schemes. The usefulness of proposed enhancements is supported by experimental results of time series analyses.

**10:30** Optimization of the Bayesian classifier’s structure using GMDH algorithm to forecast Internet-clients’ preferences
Zhyliaev Sergey, Ryabokon Dmitriy

The paper describes an algorithm allowing to raise accuracy of the Bayesian classifier due to optimization of its structure using the GMDH algorithm MULTI. Comparison of accuracy of proposed algorithm vs. regression algorithms is made on the basis of forecasting Internet-clients’ preferences. The specific character of this problem lies in the fact that Internet-consumers have a number of attributes. Developed algorithm essentially differs from the Bayesian classifier in that it allows to find out few “strong”, statistically stable attributes among large set of properties characterizing object and thus to increase the accuracy of classification.

11:00 The spatial - temporary approach in problems of clusterization
Ludmila Sarycheva

The problem of cluster analysis of the spatial-temporary data is considered. The new algorithm of GeoTime-clusterization takes into account a temporary neighbourhood of objects features (on the basis of the inductive approach) and absolute concepts of objects location (on the basis of the geoinformation approach). Using the actual data of ecological and socio-economic monitoring of the Europe states the experimental comparison GeoTime with known algorithms is carried out by three criteria of clusterization quality.

Session: Principles & Theoretical Foundations II & High Performance Computing

Time and place: Monday, September 24, 2007: 12:30—14:00 Room 1
Session Chairs: Tadashi Kondo & Miroslav Skrbek

12:30 Probability Control Functions Settings in Continual Evolution Algorithm
Zdenek Buk, Miroslav Snorek

The precise setting of all control parameters of evolutionary algorithms is very important because it affects time needed to find solution, quality of final solution or event the ability to find proper solution, and other technical parameters of computation (e.g. memory requirements), etc. In this paper we are presenting some experiences with settings of probability control functions in continual evolution algorithm (CEA). Evolutionary algorithms are typical examples of nature instired methods. We will show that the intuitive approach in exact parameters settings, based on our ideas about the nature processes, is not always the best one and we will show the modifications of control functions in CEA algorithm.

13:00 Influence of sample division on the quality of modeling and forecasting of real processes
Nina Kondrashova

Sample division and a criterion for choice of the best division are important elements in GMDH algorithms. Sample divisions effective in the tasks of approximation, extrapolation and forecasting are considered in the paper. The main attention was placed to quasi-optimal
sample division which enables to enhance the extrapolation and forecasting precision in combination with an adaptive prognosis. Some set of sample division methods allows to choose a proper technology for every task taking into account the object features.

**13:30** Computer tests as an instrument for effectiveness investigation of modeling algorithms
Serhiy Yefimenko, Volodymyr Stepashko

Technology of testing algorithms effectiveness for structural models identification with the use of statistical computers tests is developed. The recurrent bordering algorithm is investigated. Recurrent modifications of well known Gauss and Gram-Schmidt methods are developed. The comparative testing of recurrent methods for solving of linear equations systems for the parameters estimation problem criteria as well as structural identification of models is carried out. Effectiveness of parallel computation are explored with the purpose of extension of modeling possibilities from statistical data. Real world problems are solved for modeling of the ferromolybdenum market price and upper sedimentary layer density of the Caspian Sea bottom

Session: Inductive Modeling vs. Computational Intelligence

Time and place: Monday, September 24, 2007: 15:00—16:30 Room 1
Session Chairs: John Elder & Miroslav Čepek

**15:00** The Combination and Comparison of Neural Networks with Decision Trees for Wine Classification
Rohitash Chandra, Kaylash Chaudhary, Akshay Kumar

This work presents the comparison and combination of neural networks with decision trees on the application of wine classification. Neural networks are first trained and then combined with decision trees in order to extract knowledge learnt in the training process. Artificial neural networks are used for the classification of Italian wines obtained from a region which has three different wine cultivars. Wines are classified according to their respective cultivar using the chemical analysis of the thirteen major chemical constituents. The trained network classifies a sample of wine according to the knowledge the network acquired by learning from previous wine samples. After successful training, knowledge is extracted from these trained networks using decision trees in the form of ‘if-then’ rules. We then use decision trees to train on the same dataset and compare the performance of neural networks, and decision trees in both knowledge extraction from neural networks and classification of wines on their own. Our results show that artificial neural networks perform better when compared to decision trees however, the extraction of knowledge from neural networks do not outperform the performance of decision trees alone. The general paradigm can be applied to other categories of food classification and processing.
**15:30 A Hybrid Approach for Modeling High Dimensional Medical Data**  
Alok Sharma, Godfrey C. Onwubolu  

This work presents the application of hybrid PCA and LDA to modeling high dimensional medical data, which is a real-life problem. For modeling and classifying medical data, we adopted this combination of two stage PCA and LDA procedure which is also known as Fisherface technique. During the training phase we applied this combination for extracting features from medical data. In the classification stage we introduced weighting ratio which is used with the conventional Euclidean distance measure to classify a given sample. For brevity we call this technique the weighted distance Fisherface technique. The presented technique shows promising results for medical data when compared with standard GMDH technique; in the two problems taken from the machining learning databases, the presented approach performed better than the standard GMDH.

**16:00 Age Prediction from Skeletal Indicators using Computational Intelligence Methods**  
Zdenek Buk, Pavel Kordik, Miroslav Snorek  

This paper presents the work of age prediction of human beings from their skeletal indicators using computational intelligence methods, such as feedforward neural networks, learning vector quantization (LVQ) and group of adaptive models evolution (GAME). The anthropology data set we have performed our experiments on, contains a lot of noise, which is characteristic feature of almost all data collected by observation. Goal of this work was to get the best possible results of the age prediction on such noisy data and to compare the results of particular methods.

---

**Session: Real-world Applications I**

Time and place: Monday, September 24, 2007: 9:30—11:30  
Room 2  
Session Chairs: Godfrey C. Onwubolu & Pavel Náplava  

**9:30 Application of GMDH to the Environmental Modeling with Short Samples**  
Vladimir A. Vissikirsky, Volodymyr S. Stepashko, and Ioannis K. Kalavrouziotis  

The paper considers the issues concerning the environmental modeling and analysis of agricultural and forest plant species grown under different controlled and uncontrolled conditions, with short data samples available. The key approach to the analysis of such kind of complex systems consists in the decomposition of a modeled system into subsystems with consistently small number of variables. The paper shows the examples of how to analyze individual subsystems and generalize modeling results for different subsystems to the whole system, by utilizing the GMDH advantages and qualitative analysis techniques. Here, the GMDH has been applied to estimate and analyze features of the tree growth dynamics, heavy metals impact on species cultivated near the motor road, etc. The GMDH generates the sets of models with different structures and accuracies thus enabling the identification of the most relevant structures, the analysis of their similarity and the assumptions about input of different
factors. In particular, it is shown how a pair of linear models of identical structure describing different subsystems can be compiled into a model that describes a higher-level subsystem. Qualitative analysis techniques give the possibility to analyze behavior of individual subsystems as well as the system in whole.

[10:00] The Application of Neural Networks in Prediction Problems
Rohitash Chandra, Godfrey C. Onwubolu

This work presents the application of neural networks to real-life prediction problems. Neural networks are trained to predict three real world application problems given data for training and testing. We pre-process the actual output of the dataset by converting the values to integers and later translate them to binary strings. The sigmoidal output neurons of the feed-forward architecture predicts the output as binary values which are then translated back to integers to further compare the predicted output values of the network with the desired or actual output values from the dataset. The results for two, out of the three problems solved show that neural networks can be used to predict real-life problems similar to other inductive modeling approaches. A neural network can therefore be classified as an inductive modeling approach since it can be used for prediction, given the desired output. Finally, the performance of neural networks are then compared to GMDH.

Nina Kondrashova, Andriy Pavlov, Yaroslav Pavlov

Constructing the forecast models for values of tiol-disulfide ratios in blood samples in several measurement points is considered for decreasing the time of patients’ examination. Models are obtained by the GMDH algorithms. The number of examination measurement points in which models have feasible error is maximized by optimization of initial data sample division, by adaptive forecast, and sequential use of selection criteria. The criteria and results of numeric experiments are given. Application of difference models with an adaptive forecast and algorithm of modeling with the two-stage division of initial sample are effective for the variables forecast.

[11:00] Application of Basic GMDH Algorithm for Environmental System Modelling
Neeraj Kumar, Prasad Modak

This paper discusses the statistical aspects of Group Method of Data Handling (GMDH) algorithm, used for the environmental system modeling. A basic GMDH algorithm is applied for the prediction of air quality modeling. The basic GMDH is a heuristic procedure, which models the input-output relationship of a complex system using a multilayered network structure. Each element in each layer in the network implements a nonlinear function of two inputs. Various experimental aspects of GMDH including Multivariate GMDH models have also been presented. Results obtained, and as compared to that from Autoregressive models, show that the GMDH method is ideal for complex, unstructured systems, with limited data, to obtain a non-linear input-output relationship with a good accuracy. Purpose of this Paper is to discuss statistical aspects of this algorithm and its applicability to environmental systems.
Session: Optimal Complexity of inductive Models I

Time and place: Monday, September 24, 2007: 12:30—14:30  Room 2
Session Chairs: Panos Liatsis & Nikolay Zagoruiko

12:30 Optimization of a Attributes Selection: Minimum of Mistakes Versus FRiS-function
Nikolay Zagoruiko, Irina Borisova, Olga Kutnenko

For an estimation of informativeness of separate attributes or their subsystems it is offered to use average value normalized functions of a rival similarity (FRiS) objects of training sample to the pattern. This criterion differs from criterion in the form of number of correctly recognized objects higher connection with results of recognition of control sample, a greater noise stability, an opportunity to estimate suitability of the chosen attributes and reliability of recognition of control object.

13:00 Inductive Method of Optimal Model Selection by External Error Criterion with Additional Determination Using Bias
Alexey Ivakhnenko, Eugenya Savchenko, Ludmyla Syomina

The problem of optimal model selection often occurs in a real problems of modeling, because there can be a several accurate models. For this case the method of determination of model by additional bias criterion is proposed. At first the criterion of regularity is calculated, then if the optimal model is not possible to select, the additional bias criterion is calculated for the models which fall into interval of uncertainty. The examples of models determination are shown in problems of selection of aircraft surface material, forecasting of glucose level for diabetes patients and prediction of solar activity by Wolf’s numbers.

13:30 Structural identification of interval models of the static systems
Mykola Dyvak, Volodymyr Manzhula, Andriy Pukas

There is the method of structural identification of “input-output” model of static system with interval data described in the paper. This method is based on genetic algorithm of synthesis of model structure and on method of ranking generated structural elements, which allows reducing computing complexity of the developed algorithm.

14:00 Modelling in the Class of Regression Equations Systems
Alexander Sarychev

The problem of search of optimum complexity system of regression equations by principles of the Group Method of Data Handling surveyed. The criterion of quality of a system of regression equations which is system analogue of criterion of the regularity is offered. The criterion is researched in the scheme of repeated observations.
Session: Optimal Complexity of inductive Models II

Time and place: Monday, September 24, 2007: 15:00—16:30  Room 2
Session Chairs: Pavel Kordík & Petr Burian

15:00 Regularization of Evolving Polynomial Models
Pavel Kordík

Black box models such as neural networks are popular because they can deliver reasonably accurate model almost instantly. Sometimes, it is more convenient to use a math model instead of black box model. Math models can be either designed by experts or automatically generated from data describing modelled systems. The disadvantage of generated math models is that they are often too complex to be understood by experts. In this contribution we experiment with regularization of generated models to enable automatic evolution of models that are both enough accurate and understandable. We limit our experiments to models consisting of polynomial transfer units.

15:30 Some Results of the Synthesis of GMDH and Factor Analysis for Inductive Modelling
Yuriy Dzyadyk

It is known that GMDH generally requires building and comparison of \(2^k\) linear models and choice of the best among them (above \(k \approx 2^l\) is a number of all linear arguments selected among basic functions 1, \(t_i\), \(t_it_j\), ..., \(\sqrt{t_i}\), etc, \(i, j, \ldots =1..l\)). Since arguments are correlated, a little alteration of input data often results in a model with absolutely different suites of arguments. We propose two steps for application of factor analysis to GMDH. The first step simply consists in using GMDH in the orthogonal basis of factors. On the second, heuristic step we preliminarily obliterate so called unstable and inessential factors. Some versions of this method are realized by means of Java. It was successfully used for modelling and forecasting of extremely unstable molybdenum prices: monthly prices in 2004–07, and annual prices in 1975–98.

16:00 Adaptive parallel implementation of the Combinatorial GMDH algorithm
Oleksiy Koshulko, Anatoliy Koshulko

The combinatorial algorithm of the Group Method of Data Handling is a compute intensive modeling method well proven for analysis and forecasting of variety of complex systems especially of the so-called "black-box" type. The algorithm is highly dependent on computational power that makes reasonable the use of multiprocessing. To exploit efficiently different kinds of multiprocessor computer systems we propose an adaptive parallel implementation of the combinatorial GMDH algorithm and an example of its usage for the forecasting of the "Top500 Supercomputer's List".

Session: Social Event (Boat on Vltava river)
Time and place: Monday, September 24, 2007: 19:00—22:00
Session Chair: Pavel Kordík
Tuesday September 25\textsuperscript{th} 2007:

<table>
<thead>
<tr>
<th>Track</th>
<th>9:30—11:30</th>
<th>12:30—14:30</th>
<th>15:00—16:30</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Real-world Applications II</td>
<td>Data Mining, TS</td>
<td>OEINN Posters</td>
</tr>
<tr>
<td>5</td>
<td>Time Series (TS)</td>
<td>Optimization &amp; Evolution of Inductive Modeling and Neural Networks (OEINN)</td>
<td>Posters</td>
</tr>
</tbody>
</table>

Session: Real-world Applications II

Time and place: Tuesday, September 25, 2007: 9:30—11:30 Room 1

Session Chairs: Tatyana I. Aksenova & Zdeněk Buk

\textbf{9:30} Comparison of Inductive Modeling Method to Other Classification Methods for Holter ECG

Miroslav Cepek, Václav Chudádek, Milan Petrík, George Georgoulas, Chrysostomos Stylios

In this work we present a study which compares method based on inductive modeling called GAME with several classification algorithms. To compare these methods we will use long-time Holter ECG data. More specifically we focused on the task of classification of normal ‘N’ beats and premature ventricular ‘V’ beats. Some of the tested methods represent the state of the art in pattern analysis, while others are novel algorithms developed by us. All the algorithms were tested on the same datasets, namely the MIT-BIH and the AHA data bases. The results for all the employed methods are compared and evaluated using the measures of sensitivity and specificity.

\textbf{10:00} Reconstruction of Eye Movements Signal using Inductive Model Detecting Saccades

Ales Pilny, Pavel Kordik

This article describes a method for reconstruction of eye movement signals interfered with saccades and post-determination of inherent frequencies in the signal. For healthy patients, a signal of their eye movements should contain the same frequencies as movements generated by special rotating chair. To determine frequencies in eye movements, saccades have to be removed first. This is not an easy task, because saccades can have various shapes. To detect saccades, we use inductive models trained on various saccadic eye movement signals. To remove saccades and to reconstruct the eye movement signal we wrote special script replacing saccades with estimated trend of signal based on the output of the inductive model. When the reconstructed signal is transformed to the frequency domain, it is easy to decide, whether the eye movements signal contains the same frequencies as the original signal of the rotating chair.
GMDH-based Approach for Analysis of Mass Spectra in Clinical Proteomics
Dimitri V. Nowicki, Vladislav Shaposhnik, Ali Bouamrani, Marie Arlottto, François Berger and Tatyana I. Aksenova

Several architectures and algorithms of feed-forward networks and neural associative memories as well as GMDH-based polynomial NNs are tried for proteomic data analysis. The problem of chemotherapy responsiveness prediction by data of mass-spectroscopy is considered to explore potential applications of different neural paradigms for this domain.

Inductive Modeling in Newborn Sleep Stage Recognition
Vaclav Gerla, Miroslav Bursa, Lenka Lhotska, Pavel Kordik, Karel Paul, Vaclav Krajca

This paper addresses automated classification of newborn sleep electroencephalogram (EEG) using inductive classification methods. Newborn EEG plays an important role in determining the maturity level of neonatal brain. Polysomnography (PSG) recording can be classified into four important behavioral states: quiet sleep, active (non-quiet) sleep, wakefulness and movement artifact. Infant sleep significantly differs from adult sleep; we therefore apply methods designed for the problem of differentiation between the described states. The proportion of these states is a significant indicator of the maturity of the newborn brain in clinical practice. In this study we use data provided by the Institute for the Care of Mother and Child in Prague (12 newborn polysomnographic signal; similar postconceptional age; all data are scored by an experienced neurologist). Automated classification is performed by inductive models evolution through ant-colony approach (ACO-DTree algorithm) and the GAME (Group of Adaptive Models Evolution) inductive models. The results are compared with standard method. Using inductive modeling methods produced better results with improved generalization skills of the classifier. The purpose of this study is to facilitate the work of neurologist.

Session: Data Mining and Applications

Time and place: Tuesday, September 25, 2007: 12:30—14:30 Room 1
Session Chairs: Witold Pedrych & Gregory Ivahnenko

Marketing Problems Solution by Different GMDH Algorithms Using Excel Software
Gregory Ivahnenko

In this paper is described the application of different inductive algorithms in Excel environment for solutions to marketing problems. Comparison of data mining algorithms for forecasting of brand sales, segmentation of product launches and their attributes, media analysis is considered. The results show that the GMDH approach can significantly decrease time needed for analysis and improve the quality of knowledge extraction for marketing problems solution.
**13:00 Data Mining using Inductive Modeling Approach**  
Godfrey Onwubolu

The rate at which organizations are acquiring data is getting out of proportion and managing such data so as to infer useful knowledge that can be put to use is increasingly becoming important and challenging. Data Mining (DM) is one such relatively recently technology that has emerged that is employed in inferring useful knowledge that can be put to used from a vast amount of data. This paper proposes a new design methodology which is a hybrid of differential evolution (DE) and Group Method of Data Handling (GMDH) for self-organizing data mining. The new hybrid implementation is applied to the data mining activity of prediction of soil moisture, which is an aspect of hydrology. Experimental results indicate that the proposed approach is useful for data mining technique for forecasting hydrological data.

**13:30 Dataset visualization based on a simulation of intermolecular forces**  
Jan Drchal, Miroslav Snorek, Pavel Kordík

The visualization is an important technique used in many stages of data mining process. This article deals mostly with visualization for preprocessing purposes. The aim of our approach is to visualize distances (Euclidean or others) between data samples. This can be helpful when taking picture of data clustering. In classification tasks it can be used to select outlier for removal. In this paper we present a novel way of such visualization which is based on a physical system simulation. It is inspired by intermolecular forces and employs overall energy minimization. This minimization is done via known unconstrained optimization numerical methods such as Steepest Descent, Conjugated Gradients or Quasi-Newton. The proposed algorithm was originally designed and was found useful when interpreting diversity in evolutionary algorithms. Here, we show its properties on well-known datasets Iris and Ecoli.

**14:00 Application of Agglomerative Hierarchical Clustering Algorithm for Computer Systems Protection**  
Didyk A., Bardachyov Y., Rogalsky F., Mashkov O., Khodakovsky A.

Application of cluster analysis methods for the solving of problems of computer systems attacks detection is considered. The algorithm of agglomerative hierarchical clusterization is offered, allowing effectively to solve a problem of allocation of dangerous areas of computer systems conditions.

**Session: Optimization & Evolution of Inductive Models and Neural Networks I**

**Time and place:** Tuesday, September 25, 2007: 15:00—15:30  
Room 1  

**Session Chairs:** Tadashi Kondo & Jan Drchal

**15:00 Multi-layered GMDH-type neural network self-selecting optimum neural network architecture and its application to nonlinear system identification**  
Tadashi Kondo and Junji Ueno
In this study, a new multi-layered Group Method of Data Handling (GMDH)-type neural network self-selecting optimum neural network architecture is proposed. We call this algorithm as revised GMDH-type neural network algorithm self-selecting optimum neural network architecture. Revised GMDH-type neural network algorithm has an ability of self-selecting optimum neural network architecture from three neural network architectures such as sigmoid function neural network, radial basis function (RBF) neural network and polynomial neural network. Revised GMDH-type neural network also has abilities of self-selecting the number of layers, the number of neurons in hidden layers and useful input variables. This algorithm is applied to the nonlinear system identification problem and it is shown that this algorithm is useful for the nonlinear system identification because optimum neural network architecture is automatically organized.

Session: Time Series

Time and place: Tuesday, September 25, 2007: 9:30—11:00

Session Chairs: Panos Liatsis & Petr Burian

9:30  The Combined Immune Algorithm Based on Clonal Selection
Litvinenko V. I., Bidjuk P. I., Bardachov J. N., Fefelov A. A., Sherstjuk V. G.

A dynamic system identification algorithm is developed using the basic mechanisms of clonal selection and an idea of a new evolutionary computing paradigm – gene expression programming. On the basis of the algorithm developed a computer based system is proposed for making decisions relevant to forecasting of a single variable and multivariate time series. The results of computing experiments achieved with the system developed show high quality of short and medium period forecasts.

10:00 Inductive Modelling of Temporal Sequences by Means of Self-organization
Jan Koutnik

In this paper we present a new self-organizing neural network, which builds a spatio-temporal model of an input temporal sequence inductively. The network is an extension of Kohonen's Self-organizing Map with a modified Hebb's rule for update of temporal synapses. The model building behavior is shown on inductive learning of a transition matrix from a data generated by a Markov Chain.

10:30 Time Series Prediction by means of GMDH Analog Complexing and GAME
Josef Bouska, Pavel Kordik

For time series prediction we can use either parametric or nonparametric models. In this paper we study properties of both approaches for short and medium term prediction intervals. We compare the accuracy of GMDH Analog Complexing as typical nonparametric method and the Group of Adaptive Models Evolution (GAME) as a parametric method. In our study, we
focus on medical data from Motol hospital in Prague and horticulture data from Hort Research New Zealand. Our experiments showed that for short and medium term prediction of periodic data it is better to use parametric method, whereas for irregular data and long term prediction the nonparametric method gives better results.

**Session: Optimization & Evolution of Inductive Models and Neural Networks II**

Time and place: Tuesday, September 25, 2007: 12:30—14:30  Room 2

*Session Chairs: Nader Nariman-zadeh & Jan Koutnik*

**12:30** Combinatorial GMDH algorithm with successive selection of arguments

Samoilenko O.A.

The paper considers the problem of solving tasks with large number of arguments by combinatorial GMDH algorithm. To solve this problem successive selection of the most informative arguments is suggested. Using this algorithm enables to essentially accelerate the retrieval for the best subset of regressors and to solve tasks with considerably larger number of regressors compared with ordinary combinatorial GMDH algorithm of exhaustive search of arguments.

**13:00** Design of Hybrid Differential Evolution and Group Method of Data Handling for Inductive Modeling

Godfrey Onwubolu

The group method of data handling (GMDH) and differential evolution (DE) population-based algorithm are two well-known nonlinear methods of mathematical modeling. In this paper, both methods are explained and a new design methodology which is a hybrid of GMDH and DE is proposed. The proposed method constructs a GMDH network model of a population of promising DE solutions. The new hybrid implementation is then applied to modeling and prediction of practical datasets and its results are compared with the results obtained by GMDH-related algorithms. Results presented show that the proposed algorithm appears to perform reasonably well and hence can be applied to real-life prediction and modeling problems.

**13:30** Pareto Genetic Design of GMDH-type Neural Networks for Nonlinear Systems

Nader Nariman-zadeh, Ali Jamali

In this paper, Genetic Algorithms (GAs) are deployed for multi-objective Pareto optimal design of Group Method of Data Handling (GMDH)-type neural networks that have been used for modelling of a nonlinear system. In this way, GAs with a specific encoding scheme is firstly presented to evolutionary design of the generalized GMDH-type neural networks in which the connectivity configurations in such networks are not limited to adjacent layers.
Multi-objective GAs with a new diversity preserving mechanism are secondly used for Pareto optimization of such GMDH-type neural networks. The important conflicting objectives of GMDH-type neural networks that are considered in this work are, namely, Training Error (TE), Prediction Error (PE) and Number of Neurons (N) of such neural networks. It is shown that the obtained non-dominated Pareto points are inclusive of those which can be found using Akaike’s Information Criterion (AIC) for both training and prediction errors. Moreover, an important trade-off can be discovered by such Pareto optimum approach to the design of GMDH-type neural networks which helps a designer to select a network compromisely.

[14:00] Hybrid Inductive Models: Topology of Model can reveal much about Problem
Pavel Naplava, Pavel Kordik

In this paper we study how much information about a problem we can get from topology of inductive models build to model this problem. Our experiments show that for each problem, combination of various units is the most efficient. Also the performance of optimization methods varies on different problems. When we study the topology of models, we can extract much information about problems modeled.

Session: Optimization & Evolution of Inductive Models and Neural Networks III

Time and place: Tuesday, September 25, 2007: 15:00—15:30 Room 2
Session Chair: John Elder & Gregory Ivahnenko

[15:00] Genetic Selection and Cloning in GMDH MIA Method
Marcel Jirina, Marcel Jirina, jr.

The GMDH MIA algorithm is modified by the use of selection procedure from genetic algorithms and including cloning of the best neurons generated to get even lesser error. The selection procedure finds parents for a new neuron among already existing neurons according to fitness and with some probability also from network inputs. The essence of cloning is slight modification of parameters of copies of the best neuron, i.e. neuron with the largest fitness. We describe the algorithm and show that the procedure is relatively simple. The genetically modified GMDH network with cloning (GMC GMDH) can outperform other powerful methods. It is demonstrated on some tasks from Machine Learning Repository.

Session: Posters

Time and place: Tuesday, September 25, 2007: 14:30—16:30 Atrium

[14:30] Image contrast and its connection with fuzzy logic
Roman Vorobel, Olena Berehulyak

The method for determination of image generalized contrast with power character of its elements modification is considered. The expression for determination of image generalized contrast based on power additive generator is ascertained. Theorems about the contrast
generator and the rule of contrast addition, based on Yager s-norm, are prooved. For the first time the connection between fuzzy logic and contrast determination through Yager fuzzy connectives is demonstrated. The expression for computation of proposed generalized contrast is searched out. The results of investigations are illustrated by examples.

**14:30 The Criterion of Congruence in the Theory of Self-organization**
Natalya Ivakhnenko

The criterion of congruence is one of the selection ones. It selects the best solutions both for clusterization and modeling of processes. It consists of the comparison of two or more clusterizations on two square arrays of specially organized points. Such arrays are named “faces”. It consists of the comparison of both clusterizations on two square arrays of specially organized points. Such arrays are named “faces” of given clusterisation (in Russian – “Litso”). These faces help to compare arrays with various quantity of clusters and various quantity points in them simultaneously, and because this criterion is called as a congruent one. The congruent criterions are called such ones, that take in the consideration two or more special examine qualities of objects simultaneously. For our tasks such objects serve clusters and their contents. How in the case of the algorithms of Self-organizations this is used firstly for finding the multitudes of arguments for two arrays, selecting better ones for electing the criterion. Then knowing a few better multitudes of our arguments, using already knowing procedures in the first part, find the multitudes of arguments for the full our massive Our criterion differs from others and its results of action differs also, because you can compare only your finish results for its using. For the modeling of processes you must hold on the same structures of the algorithm’s steps, but for founding of need multitudes you must not take the function value. And only the end step use the ordinary function analysis. The propose criterion will open area of other selections ones in the future.

**14:30 Comparing NN and GMDH methods for prediction of socio-economic processes**
Oleksandra Bulgakova, Oleksandr Samoilenko

There are considered two different methods for prediction of socio-economic processes: the combinatorial GMDH algorithm and an artificial neural network. These methods are analyzed in the task of modeling of the Ukrainian gross domestic product (GDP) as dependent from input arguments (investments). An analysis and comparison of these methods showed interesting results that gives pre-conditions to use capabilities of neural networks jointly with the GMDH algorithms

**14:30 Rotating Machine Vibration Analysis using Group of Adaptive Models Evolution**
Adam Docekal, Marcel Kreidl, Radislav Šmíd

Modern rotating machines often takes advantage of new designs of used gears and rolling bearings. Usage of these new component enables machine to work quieter, increase their reliability, and lengthen working life. Machine vibrations analysis belongs to important methods used for rotating machine conditions monitoring. Rotating machine vibrations are often processed by Fourier spectrum especially Power Spectrum Density (PSD) analysis. Power Spectrum Density of vibrations measured on rotating machine contains a lot of peaks at many different frequencies. Recognition of peaks produced by an analyzed part of the
machine is necessary for machine monitoring. This could be accomplished by theoretical estimation of frequencies characteristic for the analyzed part, e.g. gears tooth frequency. These theoretical estimations are unknown for new designs of a selected part or even these estimations are often uncertain. Frequencies or bands important for the solved task have to be estimated directly from the measured data too. Group of Adaptive Models Evolution was used for this purpose for tapered rolling bearing analysis and gears vibration analysis.

**Wednesday September 26th 2007:**

<table>
<thead>
<tr>
<th>Track</th>
<th>9:30—11:00</th>
<th>11:00—11:30</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Plenary Session</td>
<td>Closing Ceremony</td>
</tr>
</tbody>
</table>

**Session: Plenary Session**

Time and place: Wednesday, September 26, 2007: 9:30—11:00 Gallery

*Session Chairs: Witold Pedrycz & Nader Nariman-zadeh*

**9:30** Data mining: State-of-the-art and prospects

Nikolay Zagoruiko (Russia, Novosibirsk)

Abstract: The up-to-date status of Data Mining is characterized by the high demand of methods needed for decision of applied tasks and absence of a uniform methodological base for construction of such methods. As a result, there are developed dozens of hard discernible directions (inductive modeling, regularities retrieval, data mining, knowledge mining, knowledge discovering and so on), hundreds methods and thousand algorithms which dozens thousand of users try to understand.

The lecture attempts to systematize an arsenal of Data Mining means in the form of ontology of this subject domain. The approach which can be useful to unify methods of the decision of all primary DM goals is offered. Some general approach can be based on the next hypothesis. Basic action for a user of the classification, recognition, feature selection etc. methods consists in estimation of similarity and distinctions. A function of rival (concurrent) similarity (FRiS-function) reflecting a relative character of the category "similarity" is offered. Advantages of the FRiS-function useful for the construction of decision rules, cluster analysis, choice of informative attributes are shown. Examples of the solving of simulated and real world problems are given.

**10:15** GMDH-PNN and its applications in medicine

Tatyana Aksenova

Group Method of Data Handling (GMDH) is an effective method to identify the functional structure of a model hidden in the empirical data. Iterative GMDH provides good performance in the case of high dimension of data. Feed forward network based on short-term polynomial transfer function whose coefficients are obtained using regression technique is combined with the emulation of the self-organizing activity for the neural network structural
Polynomial Neural Networks (PNN) belongs to the family of Iterative GMDH algorithms and provides the description of a data set with a polynomial model in a parametric form with robust parameter estimation and model validation in presence of outliers. The algorithm implements the iteration procedure without an increase of the power of polynomials or the number of terms. Twice-hierarchical neural network allows improving the stability and provides the convergence of coefficients and structure. The results of computational experiments are presented. The algorithm was applied to study the problems in medicine and pharmacology. Namely the study of association of the clinical symptoms with the temporal patterns of neuronal activities recorded in human patients, prediction of responsiveness of the cancer to the chemotherapy basing on data coming from mass spectrometry and an application for the computer-aid drug design are presented.

**Session: Closing Ceremony**

Time and place: Wednesday, September 26, 2007: 11:00—11:30 Gallery

*Session Chair: Volodomyr Stepashko*