National Academy of Sciences of Ukraine

International Research and Training Centre
for Information Technologies and Systems

Czech Technical University in Prague

PROCEEDINGS

of the 8th International Workshop on Inductive Modelling

IWIM 2016

Kyiv-Zhukyn

2016

The 8th International Workshop on Inductive Modelling (IWIM'2016) was held in Kyiv-Zhukyn on July 18-20, 2016.

The initial Conference ICIM'2002 took place in Lviv, Ukraine, in 2002. Next Conferences and Workshops was held in Kyiv, Ukraine, in 2005, 2008, 2011, 2013, 2014, 2015; in Yevpatoria, Ukraine, 2010; in Prague, Czech Republic, 2007, and in Krynica, Poland, 2009.

The series of conferences and workshops on inductive modelling is the only international forum that focuses on theory, algorithms, applications, and new developments of data mining and knowledge extraction technologies which originate from GMDH as a typical inductive modelling method. Basel on principles of self-organization, inductive modelling has been using in several key areas 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 multileveled self-organization to build models from data.

The motivation of this 4th Conference 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 contribute to solving global challenges.

WORKSHOP ORGANIZERS

- International Research and Training Centre for Information Technologies and Systems of the National Academy of Sciences and Ministry of Education and Science of Ukraine (IRTC ITS)
- Czech Technical University in Prague, Czech Republic (CTU in Prague)

CHAIRS

Volodymyr Stepashko (***Chairman***)
Professor, Head of Department for
Information Technologies and Systems, IRTC ITS
e-mail: stepashko@irtc.org.ua

Pavel Kordík (***Secretary***)
Assistant Professor, Vice Dean
Faculty of Information Technology, CTU in Prague
e-mail: kordikp@fit.cvut.cz

INTERNATIONAL PROGRAM COMMITTEE

Tetiana Aksenova	Grenoble University, France
Hamparsum Bozdogan	The University of Tennessee, USA
He Changzheng	Sichuan University, Chengdu, China
John Elder	Elder Research, USA
Tadashi Kondo	The University of Tokushima, Japan
Pavel Kordík	Czech Technical University, Czech Republic
Frank Lemke	KnowledgeMiner Software, Berlin, Germany
Panos Liatsis	City University London, UK
Nader Nariman-zadeh	The University of Guilan, Iran
Witold Pedrycz	University of Alberta, Canada
Volodymyr Stepashko	IRTC ITS, Kyiv, Ukraine
Miroslav Šnorek	Czech Technical University, Czech Republic

LOCAL ORGANIZING COMMITTEE

Ievgenia Savchenko (Secretary)
Ludmyla Somina
Serhiy Yefimenko
Halyna Pidnebesna
Andriy Pavlov

International Research and Training Centre
for Information Technologies and Systems
of the NAS and MES of Ukraine

ISBN 978-966-02-7964-3

Table of Contents

Adaptive Web-Interfaces Conception Based on Inductive Approach

Oleksandra Bulgakova

Mykolaiv V.O.Suhomlynsky National University, Nikolska str., 24, Mykolaiv, 54030, Ukraine

sashabulgakova@list.ru

Abstract. *This paper presents principles of adaptive human-machine systems and interfaces. Presents the constructing an adaptive interface concept using an inductive approach. These principles in this paper only describes the task of developing an adaptive interface, but did not give its formal description based on criteria, factors and specific information sense.*

Keywords

Interface, inductive modeling, application, adaptive,

Introduction

In recent years, computer modeling and visualization tools have evolved from solving problems in a powerful apparatus for studying various phenomena and problems. They are successfully used in industries such as systems analysis, design automation, organization of the computational tools and computer networks.

With this, changing user requirements and to computer systems, and it raises the question of adaptive information user interaction with the computer.

Interface organization methods

Interface - a combination of software and hardware providing the user interaction with a computer [1].

The interface can be arranged as automatic, interactive and interactive tools for user actions support:

Automatic that run the task to perform, connect it with specific data and perform some maintenance procedures;

The dialogue, which refers to regulated information exchange between man and computer which performed in real time and sent to a joint solution of the specific problem. Each dialogue consists of separate process I / O, which is physically provide a link between the user and computer. The information exchange is carried out by sending the message. In response, the user receives a prompt or help information messages to be answered, orders require actions, error messages and other information.

Interactive. According to [1], interactivity – the ability of information and communication system actively and diversely respond to user actions. Therefore, an interactive interface – the interface that allows on-line effectively intervene in the simulation process to correct some program parameters for improving the particular task results, and helps to reflect the intermediate and final calculation results in the form of both text and graphic information for subsequent visual analysis.

Adaptive. In [2] Adaptation is defined as the process of purposeful parameters and structure system change, which is to determine the criteria for its operation and implementation of these criteria. Mentions about adaptive interface can be seen not only in terms of ergonomics and engineering psychology. For example, in computer science extended the concept of "intelligent interface", the basis on adaptive properties.

Inductive approach to construction adaptive interfaces

Problems of complex systems modeling can be solved using both the logical deductive methods and sorting-out inductive ones. Deductive methods have advantages in the case of simple modeling tasks, if there is known the theory of an object being modeled and therefore it is possible to build a model based on physically based principles using knowledge of processes in an object [3-5]. But these methods can not give satisfactory results for complex systems. In this case, an approach of knowledge extraction directly from the data based on experimental measurements has an advantage. A priori information about the properties of such objects may be only minimal or even absent.

One of the most well-known modeling techniques devoid of problems described above is inductive modeling.

In adaptive interface task: $CR = \{cr_1, ..., cr_{|CR|}\}$ – set of criteria that will change, that is, to adapt to the user. $Fact = \{fact_1, ..., fact_{|Fact|}\}$ – it's a lot of factors that will influence the particular criterion choice.

Based on selected factors which may affect the interface, the interface design criteria will vary. The induction process is that the interface adaptation takes place on the specific observation data, that is, factors - to the total model, which includes criteria set that will be changed (for example, known information about the user's age (age - factor) can be increased size font (for old people) - the criterion.

Thus, it is possible to construct a function that describe the adaptation process.

$$\Phi(cr_1(fact_1,..., fact_k),..., cr_m(fact_k,..., fact_F)) = \theta_0 + \sum_{i=1}^{m} \theta_i cr_i \sum_{F=1}^{k} (fact_1,..., fact_F) +$$

$$+ \sum_{i=1}^{m} \sum_{j=1}^{m} \theta_{ij} cr_i cr_j \sum_{F=1}^{k} (fact_1,..., fact_F) + \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{m} \theta_{ijk} cr_i cr_j cr_k \sum_{F=1}^{k} (fact_1,..., fact_F) + ...$$

where θ is an estimation of parameters for any partial model.

The general principle of interface adaptation mechanism includes several units: determining factors unit, which are generated / saved user factors that may affect the interface change criteria, ie adaptation; formation criteria unit (number of criteria for each factor can be different), the data processing unit, where the program content criteria is processed to result in a interface change.

Conclusion

Presented in the paper concept of building an adaptive interface using an inductive approach, and interface components of the mechanism adaptation.

In this paper only describes the principles of adaptive interface design task, but do not give its formal description based on criteria, measurable factors and specific information sense.

References

[1] User interface – http://ru.wikipedia.org/wiki/ User_interface

[2] Rastrigin LA Adaptation of complex systems. Riga: Zinatne, 1981. 375 p.

[3] Ivakhnenko A.G. Heuristic Self– organization in Problems of Engineering Cybernetics. Pergamon PRESS / Ivakhnenko A.G. // Journal Automatica. – 1970.– №6.– P. 207–219.

[4] Ivakhnenko A.G. Polynomial Theory of Complex Systems / A.G.Ivakhnenko // J. of IEEE Transactions on systems man and cybernetics. – 1971.– №4. – P. 364–378.

[5] Ivakhnenko A.G. Group method of data handling as a rivale of the stochastic approximation method // Soviet Automatic Control, No 3, 1968, pp. 58-72

Effectiveness of Mutation Operators in the Genetic Search for Optimal Model in a Sorting-Out GMDH Algorithm

Olha Moroz

International Research and Training Center for Information Technologies and Systems of the NAS and MES of Ukraine

Address: 40, Glushkov Ave, Kyiv, 03680 GSP, Ukraine

olga_moroz@irtc.org.ua

Abstract. *The paper presents a research of different mutation operator effectiveness in searching-out COMBI-GA algorithm used to find model of optimal structure quickly because of avoiding exhaustive search. The obtained experimental results demonstrate that this algorithm performs well when solving artificial inductive modelling task.*

Keywords.

GMDH, Combinatorial algorithm, Genetic Algorithm.

1 Introduction

GMDH algorithms [[1]] belong to the most effective means for solving inductive modelling problems like structur and parametric identification, forecasting, object and processes modeling from statistical or experimental data samples under uncertainty conditions. These algorithms include two main groups, sorting-out and iterative ones. The main example of the first group of algorithms is combinatorial GMDH algorithm COMBI [[2]], the second is multilayer iterative algorithm MIA GMDH of neural network type.

In the COMBI algorithm is free of main MIA GMDH drawbacks such as high probability of loss of relevant arguments and retention of irrelevant ones. A hybrid structure of COMBI and Genetic Algorithms (GA) for over ten years are developing by foreign scientists [[3]]. However combinatorial algorithm has a significant drawback: for searching the best model by a given external criterion it uses exhaustive search of all possible model structures in a given basic set of functions (e.g. polynomial) and accordingly has exponential complexity. As a result, its usage is unreasonable for problems with more than 30 input variables. So studying the efficiency of hybridization between COMBI and GA is relevant and promising.

Hybrid algorithm COMBI-GA [[4]] is able to largely eliminate these deficiencies using genetic algorithms to search optimal model, which is a powerful tool of direct global search in optimization problems, have a clear structure and easy implementation. The effectiveness of GA significantly depends on the method of coding individuals in a population, the initial population size, genetic operators, fitness functions and stop criteria of the algorithm. Therefore, the detailed investigation of the impact of each factor on the GA performance is an important task.

The most significant impact on the GA performance make a crossover operator, which mostly affects on a variety of individuals in the population, and thus guarantees the solution of the problem. But often GA gets stuck in local optima. This phenomenon usually occurs when crossover operator is poorly chosen under given specific tasks and also when the fitness function is close to the optimum value or all simultaneously.The main mechanism of GA to solve this problem is the mutation operator with successful selection of the intensity and probability.

The aim of this paper is to analyze the effectiveness of two types of mutations, genetic and chromosomal ones, and their influence on speed of COMBI-GA algorithm, depending on their level of intensity and crossover operator probability. In the numerical experiments all other parameters of GA are fixed.

2 Algorithm COMBI-GA

A general definition of the inductive modelling problem may be done as follows. Let us given: a data set of n observations after m input x_1, x_2, …, x_m and one output y variables. The GMDH task is to find a model $y=f(x_1,x_2,\dots x_m, \theta)$ with *minimum variance of prediction error*, where θ is unknown vector of model parameters. The optimal model is $f^*= \arg\min_{\Phi} C(f)$, where $C(f)$ is a given model quality criterion, Φ is a set of models, $f\in\Phi$. This problem consists of a discrete optimization task for finding the best model and continuous optimization task for estimation of parameters.

The two tasks are generally solved by the combinatorial algorithm COMBI GMDH in the following way:

1) transformation of the initial data according to the chosen system of basic functions;

2) generation of the complete set of all possible structures of partial models in the selected basis;

3) use LSM to estimate coefficients of each partial model;

4) calculation the selection criterion for each model;

5) successive selection best models by this criterion and choosing the model of optimal complexity with the minimum criterion value.

General structure of the COMBI-GA algorithm. The developed algorithm [??] generates gradually an incomplete set of most promising partial model structures using genetic principles to search for the optimal complexity model. Generally it has structure like that:

1. Generation of a random set of partial model structures of a given size as an initial population of the COMBI-GA.
2. Computing coefficients of every partial model using least squares method (LSM).
3. Calculation of an external criterion values (fitness function of the GA) for each model, for example, the regularity criterion as typical for GMDH.
4. Current selection of the best partial models (elite selection in GA) or reduction–rejection of worst individuals from the parent and offspring populations. Formation of new population of the same size.
5. Check a stopping criterion for example achieving a given accuracy or number of iterations. Stop if it is fulfilled; otherwise go to the next step.
6. Use of genetic operators (crossover and mutation) with given probability to selected individuals in the population and forming a set of partial model structures for next generation. Go to step 2.

3 Ideas of experiments

In numerical experiments individuls (model structures) of GA initial population is encoded by binary vectors of structural elements 0 and 1, indicating absence or presence of a particular argument in the model respectively. As a fitness function of each individuals we use standard GMDH external regularity criterion based on dividing the data into two samples for training and testing: the first is used to generate the different structures of models and evaluating their parameters, and the second to calculate the accuracy of the generated models.

According to results of crossover comparison study [[5]], we consider here two the best crossover operators – one-point and universal [[6]] that are used to form the offspring population from selected elite individuals. The other GA operators and parameters are fixed. The stop criterion of the hybrid algorithm is achieving given accuracy for the difference of minimum values of the fitness function on two adjacent iterations.

To produce a new offspring by the one-point (OP) crossover, a chromosome point (area between neighboring bits in a row) is randomly selected, the fragment of binary string from beginning of the first parent chromosome to the point is copied, and the rest is copied from the second parent.

Universal crossover instead of crossover point uses an additional binary mask with length equal to that of the chromosome's set. Each offspring is formed by applying the mask to every parent with the some rule (0+0=0, 0+1=1, 1+1=0). The mask can be set randomly.

Gene mutation of an individual is the inversion of a given number of randomly selected bits of chromosome with probability 0.2. A given number of inverted bits forms the level of mutation intensity. Also the paper deals with next gene mutation variation: several genes are selected randomly in a chromosome, which number may be random; only zero values of the selected genes are reversed. The chromosomal mutation is turning whole chromosome or its fragment in 180^0. The level of intensity here is the length of the fragment.

The research was carried out on the test task using laptop with a dual core processor (2.7 GHz). There were artificially generated 50 input arguments by uniformly distributed random numbers generator, and only 25 of the inputs were used as relevant ones. The task of inductive modeling using COMBI-GA is searching the true dependence on the relevant arguments. The optimal model is searching in the class of linear functions.

3 The results of computing experiments

During the experiments it was found that mutation operator in fact make no effect on the COMBI-GA algorithm with universal crossover operator. The optimal model was obtained for 1.4 sec.

For the one-point crossover operator, using mutations can significantly speed up the process of finding the true model for 1.3–1.7 sec. (2.4–5.5 sec. without mutation) Choose the mutation operator and its intensity strongly depends on the probability of crossover operator. In particular, chromosomal mutation (intensity 3 to 15) gives better result (more stable and more accurate) with small crossover probability (0.6–0.8).

If the probability of crossover operator is high (0.8–1), it is better to use gene mutation with a few inverted positions (<10). This provides a better diversity and quick finding the optimal model.

Corresponding results are shown in Fig. 1 and Fig. 2. The pair of modified gene mutation and one-point crossover gives best COMBI-GA effectiveness. Fig. 3 shows mutations influence on the convergence of the COMBI-GA algorithm.

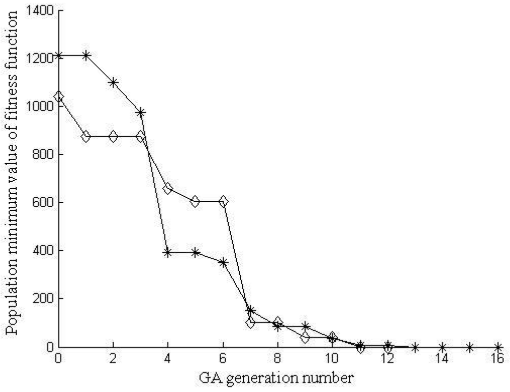

Fig. 1 COMBI-GA convergence without using (*) and with using (◊) mutation, crossover probability 0.9.

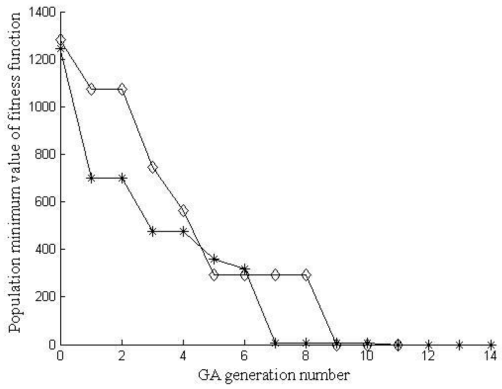

Fig. 2 COMBI-GA convergence without using (*) and with using (◊) mutation, crossover probability 0.7.

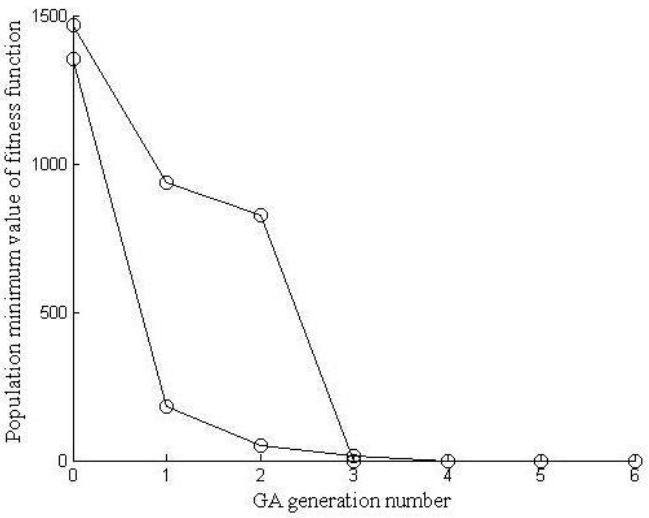

Fig. 3 COMBI-GA convergence using new variant of gene mutation for one-point crossover with different probability 0.7 (3 generations) and 0.9 (6 generations)

4 Conclusion

The use of GA is very effective to search true model in sorting-out GMDH algorithms for solving inductive modelling tasks with large number of input variables (much more than 30) during really small time.

A mutation operator is able to accelerate calculating speed of the algorithm depending on the kind of task, crossover operator and its probability. The best way to accelerate is to use the new bioinspired gene mutation operator, which is less dependent on crossover probability and has good influence on effectivenes of the COMBI-GA algorithm.

In the future, it would be reasonable theoretically substantiate the results obtained in practice when solving inductive modelling tasks using the COMBI-GA algorithm.

References

[1] Ивахненко А.Г., Степашко В.С. Помехоустойчивость моделирования. – Киев: Наук. думка. (1985). 216 с.

[2] Степашко В.С. Комбинаторный алгоритм МГУА с оптимальной схемой перебора моделей. Автоматика. № 3 (1981). С. 31-36.

[3] Мороз О.Г., Степашко В.С. Огляд гібридних структур МГУА-подібних нейронних мереж та генетичних алгоритмів. Збірник наукових праць "Індуктивне моделювання складних систем. Випуск 7. (2015). С. 173-191.

[4] Мороз О.Г., Степашко В.С. Перебірний алгоритм МГУА з генетичним пошуком моделі оптимальної складності. Міжнародна наукова конференція «Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту». Збірник матеріалів. (Залізний Порт, 26-30 травня 2016). Херсон: Вид-во ПП Вишемирський В.С. (2016). С. 297-299.

[5] Мороз О.Г Порівняння ефективності операторів кросинговеру при застосуванні генетичного пошуку кращих моделей в перебірному алгоритмі МГУА. VI Всеукраїнська школа-семінар молодих вчених і студентів «Сучасні комп'ютерні інформаційні технології» (ACIT-2016). Збірник матеріалів. Тернопіль: Вид-во ТНЕУ. (2016). С. 77-79.

[6] Гладков Л.А., Курейчик В.В., Курейчик В.М. Генетические алгоритмы. Москва.: ФИЗМАТЛИТ. (2006). 320с.

To the Issue of Econometric Modeling in Fish-Farming Technologies

[1]V.V. Osypenko, [2]V.P. Osypenko, [3]V.I. Lytvynenko

[1]Dr. Sc. in Engineering, As. Prof., National University of Life and Environmental Sciences of Ukraine
vvo7@ukr.net
[2]PhD in Biology, Senior Research Fellow, Institute of Hydrobiology of National Academy of Sciences of Ukraine
vosypenko@ukr.net
[3]Dr. Sc. in Engineering, Prof., Kherson National Technical University
immun56@gmail.com

Abstract. *The subject of this article refers to the direction of intelligent computer-integrated technologies and decision support systems for the optimal control of fish-breeding technological processes, created on the basis of modern information technology with the use of basic principles of inductive modeling techniques and application of systems analysis as well. Given the system predictive econometric model, that links the predictive integrated efficiency of aquaculture with other relevant indicators on all fisheries technological cycle after stocking of ponds to the realization of fish.*

Keywords

econometric modeling, fish-farming, inductive modeling, aquaculture, indicators

From the available literature sources, it is known that intensive technologies of industrial cultivation of aquatic living resources (ALR), including fish – is a system combination of manufacturing operations, methods and techniques focused on maximum utilization of the physiological properties of fish under the given climatic conditions throughout the vegetation period of its development. Such technologies are above all things aimed at the efficient use of water biological resources as well as the most cultivated ALR during of the entire ontogenesis at the expressed astatic of variety of environmental factors [1]. Apparently, the production in this case is clearly prevails over issues of economic type.

On the other hand, purely economic benefits should not prevail over the protection of the natural environment, compliance of the hydro-chemical and hydro-biological indicators in optimal water, that is, comfortable ranges of parameters for ALR.

From the standpoint of system-economical production of ALR, it is necessary conditions of productivity in open continental waters, is not only compliance with specific technologies aquaculture. To such technologies is important to enroll sanitation, integrated land reclamation of "water-bed" [1, 2] and so on.

A particular direction of optimal maintaining of water exchange is the modern automated process control systems (APCS). One of the priority tasks that should be resolved within the APCS is support of both hydro-chemical and hydro-biological indicators in optimal, that is comfortable range of water indicators, such as aquatic living resources in fattening ponds [3, 4].

A special place in cost price of modern intensive aquaculture technologies, including industrial fish-farming in open water, is the most costly part - feeding of fish and costly balanced feed. Therefore, the intensification of technology is important and relevant.

The main system criterion in the outlined direction of research, in our opinion, it would serve the some economic prototype that would to maximize the income (profitability) from the sale of live fish and fish-products, provided at the same time the undeniable compliance with its high quality at the moment of the maximum peak of selling.

It is clear that the problem, which has been put here, should include the tasks of system marketing as a subset of a more general class of system information-analytical research (SIAR) [5, 6].

Hence, from the foregoing, logically faces the task of system econometric forecasting of growing aquatic biological resources profitability at the moment of their future realization under current real or in some way given conditions. In this case, it is clear that besides biological and other parameters it would be worth still involve certain other factors that are described in the available literature.

Therefore, without claiming for completeness and highest relevance, the primary basis for further information and econometric studies in this direction would look like follow.

I. The natural indicators:

- yield of larvae or fry of fish from one female;
- planting fish density per one hectare of pond of all age categories;
- mass ratio grown fish to planted for growing;
- average weight of fish been grown of all age categories;
- yield from wintering yearling fish and older age groups;
- gross production of fish;
- fish productivity of fattening ponds and other water bodies in fish farm;
- the fish production per employee;
- natural food base of fish: phytoplankton and zooplankton, zoobenthos;
- the industrial returning of fish when of fish stocking of natural water bodies.

II. Current values of economic indicators:

- the specific cost of fixed assets in 1 quintal of marketable fish;
- unit cost of growing in 1 quintal of marketable fish without feed;
- the share the cost of feed on 1 quintal of cultivation of marketable fish;
- current prime price of 1 quintal of marketable fish;
- current cash sales revenue of possible of fish mass (options);
- the current selling price of 1 quintal of marketable fish;
- current a net profit from the sale of fish including taxes;
- current profitability of fish farming and fisheries;
- current breeding efficiency.

III. The predicted value of economic indicators at the moment of future product sales:

- the forecasted cost of fixed assets in 1 quintal marketable fish;
- the forecasted specific cost of growing on 1 quintal marketable fish without of feed;
- the forecasted specific cost of feed on 1 quintal of cultivation of marketable fish;
- forecasted prime price of 1 quintal of marketable fish;
- forecasted selling price of 1 quintal of marketable fish;
- reasonable forecasted sales volume;
- forecasted a net profit from the sale of fish including taxes;
- forecasted profitability of fish farming;
- forecasted sales revenue of possible of fish mass (options);
- forecasted breeding efficiency.

IV. The current factors of environmental conditions of water biological resources:

- chromaticity of water and water clarity;
- dissolved oxygen, mg / L;
- pH value of water;
- alkalinity, mg eq. / L;
- the total salinity and others.

Reference

[1]. Baltadzhi R.A. Technology reproduction of herbivorous fish in inland waters of Ukraine / R.A. Baltadzhi //. Kiev. 1996. – 65 p. [ukr]

[2]. Andryushchenko A.I. Aquaculture artificial reservoirs. Part II. Industrial aquaculture / A.I. Andryushchenko, N.I. Vovk. – K .: NUBiP, 2014. – 586p. [ukr]

[3]. Osypenko V. P. Seasonal and spatial changes in the content and molekular mass distribution of carbohydrates in the surface water. Hydrobiological Journal. – 2014 – V. 50, № 5 – P. 89-98.

[4]. Vasylchuk T.A., Osypenko V.P., Yevtukh T.V. The peculiarities of the migration and distribution of main groups of organic matter in the water of the Kyiv reservoir depending on the oxygen regime. Hydrobiological Journal. – 2011. – Vol. 47, № 2. – P. 97-106. [ukr]

[5]. Osypenko V.V. Intelligent systems for optimum control of fish-farming technologies on continental waters. Part I. Methodological Aspects / V.V. Osypenko // Fish industry of Ukraine. 2006. – №2. – S.43-52. – 2006. – №2. – P.43-52. [ukr].

[6]. Osypenko V.V. The concept of information-analytical decision support systems in intensive aquaculture technologies / V. Osypenko // Scientific and methodical journal behalf of Petro Mohyla complex "Kyiv-Mohyla Academy" – Mykolaiv, 2012. – Vol. 173, No.163 – P. 67-73. [ukr]

Application of GMDH for Modelling of Influence of Different Factors on Chlorophyll Concentration in Phytoplankton in Kakhovka Reservoir

Halyna Pidnebesna[1], Alevtyna Kureyshevych[2]

[1] International Research and Training Center of Information Technologies and Systems, Kyiv, Ukraine

[2] Institute of Hydrobiology, Kyiv, Ukraine

pidnebesna@ukr.net, alischuk@rambler.ru

Abstract. *This paper studies the influence of various natural and anthropogenic factors on the chlorophyll concentration, which is the main photosynthetic pigment of phytoplankton. The study was conducted using GMDH of Kakhovka reservoir data.*

Keywords

chlorophyll content, phytoplankton, GMDH, Kakhovka reservoir

Introduction

Studying of the aquatic ecosystems and forming the biological productivity and water quality in the increasingly eutrophication and pollution is an important task of modern ecology.

High level of anthropogenic load on water bodies in Ukraine is conditioned by high population density, the influence of industry, agriculture, and ineffective system of sewage entering the reservoir. In this connection, it is necessary to ascertain the response of planktonic algae to environmental changes. Study of the fundamental basis of phytoplankton forming and global impact of environmental factors on this process in terms of anthropogenic impact are important for the development of the theory of biological productivity of reservoirs, scientifically based management functioning ecosystems and predicting water quality.

This paper studies the influence of various natural and anthropogenic factors on the chlorophyll concentration - the main photosynthetic pigment of phytoplankton, which is a measure of the intensity of its vegetation and the level of primary production. The study was conducted by analyzing of long-term data Kakhovka reservoir using the combinatorial GMDH algorithm.

Description of research

The reaction of planktonic algae on the influence of natural and anthropogenic factors can be studied using the concentration of chlorophyll *a.* Chlorophyll *a* is the main photosynthetic pigment of phytoplankton. It is a measure of the level of phytoplankton vegetation and primary products. The ecological status of the water bodies and biomass of planktonic algae can be assessed by the concentration of the chlorophyll per unit volume of water.

Mentioned concentration of chlorophyll depends on many factors such as hydrological and meteorological conditions of light and temperature, season, time of day, the species of algae, their supply of nutrients, physiological state populations. [1]

In the Dnieper reservoirs in summer the direct dependence is observed between long-term dynamics of average chlorophyll concentration and phytoplankton biomass. This dependence appears the most in the largest reservoirs of the Dnieper regulating its flow, particularly in Kiev, Kremenchug and Kakhovka.

In this work we fitted the linear model of the chlorophyll concentration. It includes the following factors:
x_1 – total amount of the dissolved inorganic nitrogen, mg N / l (factor 1);
x_2 – dissolved inorganic phosphorus P mg / l (factor 2);
x_3– N/P – the ratio of nitrogen to phosphorus, the relative unit (factor 3);

x_4 – water temperature, t ° C (factor 4);
x_5 – the volume of water flow, м³×109 / month (factor 5);
x_6 – total solar radiation, MJ / m² × month (factor 6).

Description of the study

The Combinatorial GMDH algorithm was used to build the process models. Models were constructed using different numbers of factors (from 1 to 6) based on the data from the years 1980-93 [2]. We selected the most appropriate model according to the regularity criterion AR for each data division options. We studied three options for division of the sample on learning set X_A and testing set X_B:

- 7/6 – the learning set $X_{A,}$ contains 7 data points for 1980-1986 years, testing set X_B contains 6 data points from 1987 to 1993 years,
- 8/5 - the learning set $X_{A,}$ contains 8 data points for 1980-1987 years, testing set X_B contains 5 data points from 1988 to 1993 years,
- 9/4 – the learning set $X_{A,}$ contains 9 data points for 1980-1988 years, testing set X_B contains 4 data points from 1989 to 1993 years,

Data point means all measured data in August of each year.

We selected three best models as the result of modelling. These models are presented in Table 1:

Table 1.

The model	The regularity criterion	The coefficient of multiple correlation R
$y_1 = 21.59x_1$	15,46	0,6751
$y_2 = 18.23x_1 + 96.25x_2$	13,37	0,7417
$y_3 = 16.1x_1 + 0.81x_4 – 1.56x_5 – 0.002x_6$	25,23	0,8212

The value of the coefficient of multiple correlation suggests a strong relationship between inputs and independent variable.

Conclusion

The analysis of the obtained results showed that all chosen models include variable x1 - mineral form of nitrogen. So, the most important factor of influence to the content of chlorophyll in phytoplankton Kakhovka reservoir in summer is the variable x1. Phosphorus is growing faster than nitrogen in the water, that is the reason why nitrogen is a factor which is limiting the development of phytoplankton. It was observed in the Kakhovka reservoir in the 1980-90 biennium.

For Kakhovka reservoir using the model with four factors, y=f(x1, x4, x5, x6) is completely sufficient.

The model includes the following factors:

- the concentration of nitrogen in the water,
- the ratio of the concentrations of nitrogen and phosphorus (N / P),
- the flow of water through the turbines of hydropower
- total solar radiation, respectively.

These results were obtained for Kakhovka reservoir, the current sampling statistics, and summer season, which is characterized by an intense "bloom" of blue-green algae water.

References

[6] Kureyshevych A.V., Pidnebesna H.A., Stepashko V.S. Analysis of dependence of chlorophyll in phytoplankton of Kakhovka reservoir on combined factors by of multiple regression. // Inductive modelling of complex system. – K.: IRTC ITS, 2015, # 7. – P. 147-153.

[7] Pidnebesna H.A., Kureyshevych A.V. Application of GMDH to model the impact of various factors on the content of chlorophyll in phytoplankton Kakhovka reservoir. // Preceedings of ISDMCI-2016. Kherson . – P. 303-305.

Study of the Deterministic Component of the Solution Error of Discrete Ill-Posed Problems for Truncated Singular Value Decomposition and Random Projection Methods

Elena G. Revunova

Dept. of Neural Information Processing Technologies, International Research and Training Centre for Information Technologies and Systems, prospect Academica Glushkova 40, Kiev, Ukraine

helab@i.com.ua

Abstract. *We have shown that the deterministic component of error obtained by the random projection method can be represented as the sum of the deterministic component of error obtained by the truncated singular value decomposition method and another term.*

Keywords

ill-posed inverse problems, truncated singular value decomposition, random projection.

1 Introduction

Discrete ill-posed problems (DIP) often occur when signals are reconstructed based on the results of indirect measurements [1, 2, 3]. The stable solution of DIP can be obtained, for example, by truncated singular value decomposition (SVD) [4, 5, 6, 7] or random projection [8, 9, 10, 11] methods.

The studying of the solution error for these methods is actual. In this paper we have taken steps to study the connection of the deterministic component of the solution error obtained by the truncated singular value decomposition and the random projection methods.

2 Solving of DIP using the methods of truncated singular value decomposition and random projection

The problem of estimating a signal vector $\mathbf{x} \in \Re^n$ for the model $\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\varepsilon}$, where we know the matrix $\mathbf{A} \in \Re^{n\times n}$ and the vector of measurements $\mathbf{y} \in \Re^n$ ($\mathbf{y} = \mathbf{y}_0 + \boldsymbol{\varepsilon}$), if the singular values of the matrix $\mathbf{A}$ gradually decrease to zero, belongs to the class of discrete ill-posed problems [7].

Stable solution of DIP can be obtained by using the truncated singular value decomposition [4, 5, 6, 7] or the random projection method [8, 9, 10, 11]. The vector of estimation $\mathbf{x}^*$ based on truncated singular value decomposition is obtained by the following linear model [6]:

$$\mathbf{x}^*_{k\ SVD} = \sum_{i=1}^{k} \mathbf{v}_i s_i^{-1} \mathbf{u}_i^{\mathrm{T}} \mathbf{y}\,, \tag{1}$$

where $\mathbf{u}_i \in \Re^n$ are left singular vectors, $\mathbf{v}_i \in \Re^n$ are right singular vectors and s_i are singular values of the matrix $\mathbf{A}$. Note that $\mathbf{A} = \mathbf{A}_k + \mathbf{A}_n$, where $\mathbf{A}_k = \sum_{i=1}^{k} \mathbf{u}_i s_i \mathbf{v}_i^{\mathrm{T}}$, $\mathbf{A}_n = \sum_{i=k+1}^{n} \mathbf{u}_i s_i \mathbf{v}_i^{\mathrm{T}}$.

For the solution based on random projection both sides of the given equation are multiplied by a matrix $\mathbf{G}_k \in \Re^{k\times n}$: $\mathbf{G}_k\mathbf{Ax} = \mathbf{G}_k\mathbf{y}$, and then the estimation vector $\mathbf{x}^*$ is obtained by the following linear model [8]:

$$\mathbf{x}_{k\ G}^{*} = \sum_{i=1}^{k}\mathbf{h}_i\mathbf{g}_i^{\mathrm{T}}\mathbf{y}, \tag{2}$$

where $\mathbf{g}_i \in \Re^n$ is a column of the matrix $\mathbf{G}_k = [\mathbf{g}_1,...,\mathbf{g}_k]$, $\mathbf{G}_k \in \Re^{k\times n}$, elements of this matrix are realizations of a standard Gaussian random variable; $\mathbf{h}_i \in \Re^n$ is a column of the matrix $(\mathbf{G}_k\mathbf{A})^{+} = [\mathbf{h}_1,...,\mathbf{h}_k]$.

The expected squared error of the $\mathbf{x}$ recovery is calculated as $e(k) = \mathrm{E}\{\|\mathbf{x}-\mathbf{x}_k^*\|^2\}$, where $\mathrm{E}\{\cdot\}$ is expectation with respect to noise in the vector of measurements y and $\mathbf{x}_k^*$ is the recovered signal vector obtained by (1) or (2) with $\mathbf{y}$ obtained by some noise realization.

The expected squared error for the truncated SVD obtained in [4, 5] is

$$e_{SVD}(k) = \|(\mathbf{A}_k^{+}\mathbf{A}_k - \mathbf{I})\mathbf{x}\|^2 + \sigma^2\, trace\,(\mathbf{A}_k^{+\mathrm{T}}\mathbf{A}_k^{+}) \tag{3}$$

and its components are

$$e_{SVD\ d}(k) = \|(\mathbf{A}_k^{+}\mathbf{A}_k - \mathbf{I})\mathbf{x}\|^2 \text{ and } e_{SVD\ s}(k) = \sigma^2\, trace\,(\mathbf{A}_k^{+\mathrm{T}}\mathbf{A}_k^{+}), \tag{4}$$

where $e_{x\ d}$ is the deterministic component of the error and $e_{x\ s}$ is the stochastic component.

The expected squared error for the random projection method obtained in [8] is

$$e_G(k) = \|(\mathbf{B}_k^{+}\mathbf{B}_k - \mathbf{I})\mathbf{x}\|^2 + \sigma^2\, trace\,(\mathbf{B}_k^{+\mathrm{T}}\mathbf{B}_k^{+}),\ \mathbf{B}_k = (\mathbf{G}_k\mathbf{A}) \in \Re^{k\times n}, \tag{5}$$

and its components are

$$e_{G\ d}(k) = \|(\mathbf{B}_k^{+}\mathbf{B}_k - \mathbf{I})\mathbf{x}\|^2,\ e_{G\ s}(k) = \sigma^2\, trace\,(\mathbf{B}_k^{+\mathrm{T}}\mathbf{B}_k^{+}). \tag{6}$$

Numerous experimental studies [8, 9] of the dependence of expected squared error estimation on the number k of model components for the truncated SVD and for the random projection methods showed closeness of these dependencies.

In this paper we study closeness of $e_{SVD\ d}(k)$ and $e_{G\ d}(k)$ analytically.

3 The study of deterministic component of error for the truncated singular value decomposition and for the random projection methods

Let us study the connection between the deterministic component of error for the truncated SVD and for the random projection methods. $\mathbf{B}_k$ can be represented as the sum: $\mathbf{B}_k = \mathbf{G}_k\mathbf{A} = \mathbf{G}_k\mathbf{A}_k + \mathbf{G}_k\mathbf{A}_n$. We consider $\mathbf{G}_k\mathbf{A}_n$ as the disturbance of the matrix $\mathbf{G}_k\mathbf{A}_k$ and represent $\mathbf{B}_k^{+}$ by the Stewart formula [12]:

$$\mathbf{B}_k^{+} = \mathbf{C}_k^{+} + (\mathbf{G}_k\mathbf{A})^{+}\mathbf{P}_k(\mathbf{I} - \mathbf{C}_k\mathbf{C}_k^{+}) + (\mathbf{G}_k\mathbf{A})^{+}\mathbf{P}_k\mathbf{E}_k\mathbf{D}_k\mathbf{C}_k^{+} - \mathbf{R}_k^{\perp}\mathbf{D}_k\mathbf{C}_k^{+}, \tag{7}$$

where $\mathbf{C}_k = \mathbf{G}_k\mathbf{A}_k$, $\mathbf{E}_k = \mathbf{G}_k\mathbf{A}_n$, $\mathbf{D}_k = \mathbf{C}_k^{+}\mathbf{C}_k$, $\mathbf{R}_k^{\perp} = \mathbf{I} - (\mathbf{G}_k\mathbf{A})^{+}\mathbf{G}_k\mathbf{A}$.

Taking into account that $\mathbf{C}_k\mathbf{C}_k^{+} = \mathbf{I}_k$, $\mathbf{P}_k = \mathbf{G}_k\mathbf{A}(\mathbf{G}_k\mathbf{A})^{+} = \mathbf{I}_k$, $\mathbf{E}_k\mathbf{C}_k^{+} = \mathbf{O}$, $(\mathbf{I} - \mathbf{C}_k\mathbf{C}_k^{+}) = \mathbf{O}$, where $\mathbf{O}$ is a zero matrix, we obtain the following equation

$$\mathbf{B}_k^+ = \mathbf{C}_k^+ - \mathbf{R}_k^{\perp}\mathbf{C}_k^+ . \tag{8}$$

Using (8) and taking into account that $\mathbf{B}_k = \mathbf{C}_k + \mathbf{E}_k$, the expression for the deterministic error component can be written as:

$$e_{G\ d}(k) = \left\|(\mathbf{B}_k^+\mathbf{B}_k - \mathbf{I})\mathbf{x}\right\|^2 = \left\|((\mathbf{C}_k^+ - \mathbf{R}_k^{\perp}\mathbf{C}_k^+)\mathbf{B}_k - \mathbf{I})\mathbf{x}\right\|^2 = \left\|(\mathbf{C}_k^+\mathbf{B}_k - \mathbf{I})\mathbf{x} - \mathbf{R}_k^{\perp}\mathbf{C}_k^+\mathbf{B}_k\mathbf{x}\right\|^2 ,$$

$$e_{G\ d}(k) = \left\|(\mathbf{C}_k^+(\mathbf{C}_k + \mathbf{E}_k) - \mathbf{I})\mathbf{x} - \mathbf{R}_k^{\perp}\mathbf{C}_k^+\mathbf{B}_k\mathbf{x}\right\|^2 = \left\|(\mathbf{D}_k + \mathbf{C}_k^+\mathbf{E}_k - \mathbf{I})\mathbf{x} - \mathbf{R}_k^{\perp}\mathbf{C}_k^+\mathbf{B}_k\mathbf{x}\right\|^2 . \tag{9}$$

Using the SVD of $\mathbf{A}_k$, the $\mathbf{D}_k$ matrix can be represented as

$$\mathbf{C}_k^+\mathbf{C}_k = (\mathbf{G}_k\mathbf{A}_k)^+\mathbf{G}_k\mathbf{A}_k = (\mathbf{G}_k\mathbf{U}_k\mathbf{S}_k\mathbf{V}_k^{\mathrm{T}})^+\mathbf{G}_k\mathbf{A}_k = (\mathbf{S}_k\mathbf{V}_k^{\mathrm{T}})^+(\mathbf{G}_k\mathbf{U}_k)^+\mathbf{G}_k\mathbf{U}_k\mathbf{S}_k\mathbf{V}_k^{\mathrm{T}} . \tag{10}$$

Since $(\mathbf{G}_k\mathbf{U}_k)^+\mathbf{G}_k\mathbf{U}_k = \mathbf{I}_k$

$$\mathbf{D}_k = (\mathbf{S}_k\mathbf{V}_k^{\mathrm{T}})^+\mathbf{S}_k\mathbf{V}_k^{\mathrm{T}} = \mathbf{A}_k^+\mathbf{A}_k . \tag{11}$$

Note that after (10, 11), the matrix $\mathbf{G}_k$ (with the elements that are realizations of a standard Gaussian random variable) is not used in the representation of $\mathbf{D}_k$. This allows us to represent the deterministic error component of the random projection method as a sum including the deterministic error component of the truncated SVD method:

$$e_{G\ d}(k) = \left\|(\mathbf{A}_k^+\mathbf{A}_k - \mathbf{I})\mathbf{x} + \mathbf{C}_k^+\mathbf{E}_k\mathbf{x} - \mathbf{R}_k^{\perp}\mathbf{C}_k^+\mathbf{B}_k\mathbf{x}\right\|^2 , \tag{12}$$

$$e_{G\ d}(k) = \left\|(\mathbf{A}_k^+\mathbf{A}_k - \mathbf{I})\mathbf{x}\right\|^2 + \left\|\mathbf{C}_k^+\mathbf{E}_k\mathbf{x} - \mathbf{R}_k^{\perp}\mathbf{C}_k^+\mathbf{B}_k\mathbf{x}\right\|^2 + 2\left\langle(\mathbf{A}_k^+\mathbf{A}_k - \mathbf{I})\mathbf{x}, \mathbf{C}_k^+\mathbf{E}_k\mathbf{x} - \mathbf{R}_k^{\perp}\mathbf{C}_k^+\mathbf{B}_k\mathbf{x}\right\rangle . \tag{13}$$

$$e_{G\ d}(k) = e_{SVD\ d}(k) + z(k); \tag{14}$$

where $z(k) = \left\|\mathbf{C}_k^+\mathbf{E}_k\mathbf{x} - \mathbf{R}_k^{\perp}\mathbf{C}_k^+\mathbf{B}_k\mathbf{x}\right\|^2 + 2\left\langle(\mathbf{A}_k^+\mathbf{A}_k - \mathbf{I})\mathbf{x}, \mathbf{C}_k^+\mathbf{E}_k\mathbf{x} - \mathbf{R}_k^{\perp}\mathbf{C}_k^+\mathbf{B}_k\mathbf{x}\right\rangle$. The expression (14) provides a background for studying the value of the difference $|e_{G\ d}(k) - e_{SVD\ d}(k)|$, for example, by estimating the upper bound of $z(k)$. Figures 1-4 show the examples of dependencies of $e_{SVD\ d}(k)$ and $e_{G\ d}(k)$ for the Phillips, Delves, Carasso and Baart problems [13].

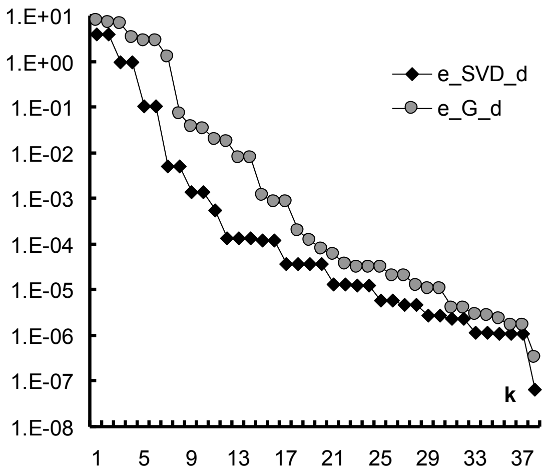

Fig.1. Dependencies of $e_{SVD\ d}(k)$ and $e_{G\ d}(k)$ for the Phillips problem

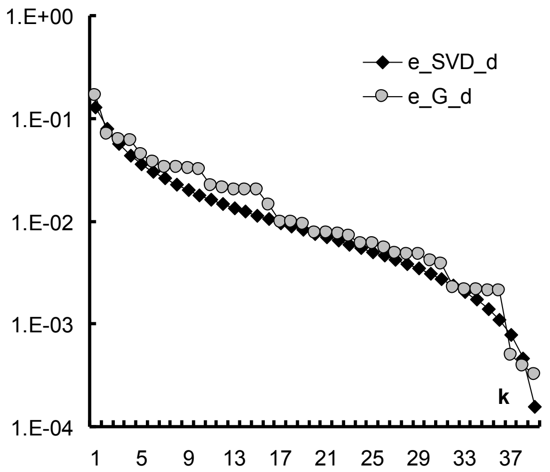

Fig.2. Dependencies of $e_{SVD\ d}(k)$ and $e_{G\ d}(k)$ for the Delves problem

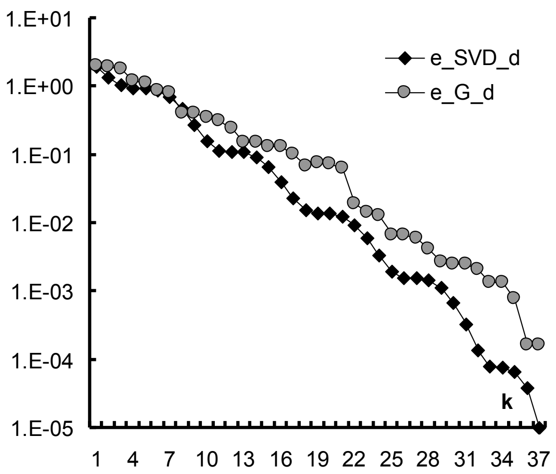

Fig.3. Dependencies of $e_{SVD\ d}(k)$ and $e_{G\ d}(k)$ for the Carasso problem

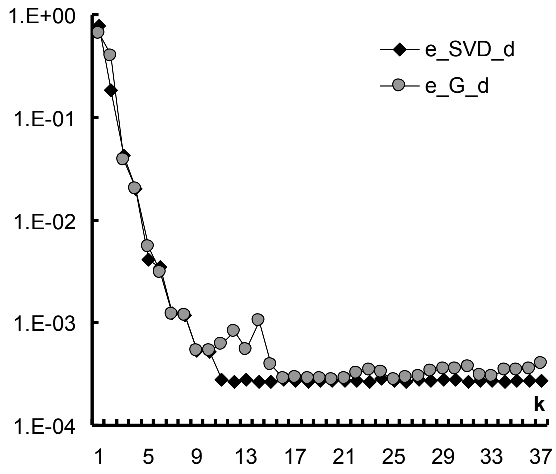

Fig.4. Dependencies of $e_{SVD\ d}(k)$ and $e_{G\ d}(k)$ for the Baart problem

4 Conclusion

In the previous papers we experimentally investigated the dependence of the mean squared error of signal recovery on the number k of components in the linear model that represents the solution vector, for both the truncated singular value decomposition and the random projection methods. The experiments had shown that the dependences are close and motivated the analytical investigation of the reasons for such a closeness started in this paper.

We have shown that the deterministic component of error obtained by the random projection method can be represented as the sum of the deterministic component of error obtained by the truncated singular value decomposition method and another term. The direction of further analytical study is to get an upper bound for the additive term.

References

1. *Хмелевский В.К.*, *Бондаренко В.М.* Электроразведка. – М.: Недра, 1999. – 438с.
2. *Забулонов Ю.Л.*, *Коростиль Ю.М.*, *Ревунова Е.Г.* Оптимизация решения обратной задачи по восстановлению функции плотности распределения поверхностных загрязнений // Сборник научных трудов ИПМЭ НАН Украины "Моделирование и информационные технологии". – 2006. – С. 77 – 83.
3. *Rachkovskij D.A.*, *Revunova E.G.* Intelligent Gamma-Ray Data Processing for Environmental Monitoring // Intelligent data analysis in global monitoring for environmental and security. – 2009. – P. 124-145
4. *Revunova E.G.*, *Tyshchuk A.V.* A Model Selection Criterion for Solution of Discrete Ill-Posed Problems Based on the Singular Value Decomposition, The 7th International Workshop on Inductive Modelling (IWIM'2015), Kyiv-Zhukyn on July 20-24, 2015, p.43-47.
5. *Ревунова Е.Г.*, *Тищук А.В.* Критерий выбора модели для решения дискретных некорректных задач на основе сингулярного разложения // Управляющие системы и машины. - 2014. - № 6. - С. 3-11.
6. *Hansen P.C.* The truncated SVD as a method for regularization. BIT 27, (1987), 534–553.
7. *Hansen P.C.* Rank-deficient and discrete ill-posed problems. Numerical Aspects of Linear Inversion. – SIAM, Philadelphia, 1998. – 247 p.
8. *Revunova E.G.*, *Rachkovskij D.A.* Using randomized algorithms for solving discrete ill-posed problems // International Journal Information Theories and Applications. – 2009. – 16, N 2. – P.176–192.
9. *Rachkovskij D.A.*, *Revunova E.G.* Randomized method for solving discrete ill-posed problems // Cybernetics and Systems Analysis. – 2012. – 48, N 4. – P. 621-635.
10. *Revunova E.G.* Analytical study of the error components for the solution of discrete ill-posed problems using random projections // Cybernetics and Systems Analysis – 2015 – V. 51 – Issue 6 – P. 978-991.
11. *Revunova E.G.* Model Selection Criteria for a Linear Model to Solve Discrete Ill-Posed Problems on the Basis of Singular Decomposition and Random Projection // Cybernetics and Systems Analysis. – 2016 – 52, N4. – P. 647-664.
12. *Stewart G.W.* On the perturbation of pseudo-inverses, projections and linear least squares problems // SIAM Review. – 1977. – 19, N 4. – P. 634-662.
13. *Hansen P.C.* Regularization Tools: A Matlab package for analysis and solution of discrete ill-posed problems // Numer. Algorithms. – 1994. – 6. – P.1-35.

The Methods of Human Development Index Data Analysis and Modeling

Savchenko E., Tutova O.

International Research and Educational Center of Information Technologies and Systems of NAS of Ukraine, pr. Academika Glushkova, 40, Kyiv, 03680, Ukraine

savchenko@irtc.org.ua, sir_ludovick@yahoo.com

1. Introduction

Macroeconomic data often contain uncertainties; also they can be fuzzy and noisy. Inductive approach is one of the approaches that give possibility to receive acceptable results on small data samples in case of presence of noise and uncertainty in data.

This approach helps to solve problems of data analysis, modeling and forecasting. Inductive modeling is based on principles of induction, i.e. general conclusion relying on some facts. The group method of data handling (GMDH) is one of the main methods of inductive modeling. Combinatorial GMDH algorithm is the most popular among them. It doesn't demand the convergence proof because it completely looks through all possible variations of models. This algorithm solves a problem of structural-parametric identification, i.e. it gives a possibility to define not only the parameters of a model, but its optimal structure too.

The aim of this report is to outline the methods of solving the problems of human development index data analysis, modeling and forecasting based on the combinatorial GMDH algorithm and correlation-regression analysis. These methods help to improve the decision-making process in the field of the social and economic development of Ukraine both on the regional level and to compare the country with other nations. They also help to estimate the situation in the field of the human potential development both in Ukraine and in the world as a whole.

2. Human development index

The rapid development of scientific and technological progress demands a new quality of human resources. Structural changes in economics became the basis for appearance of the contemporary concept of human development.

The general model for analysis of investments in human capital is based on the assessment of the expediency to invest in education, on the analysis of differences in income, and also on the demand for education services. This theory develops within the framework of the neoclassical trend in economical theory. Also it is used to explore such fields as education, health care, family relations and other fields of non-market activity. The human capital consists of acquired knowledge, skills, motivation and energy that a human being possesses, and which can be used during a certain period of time with the purpose to produce goods and services.

The human capital is a form of capital because it is a source of future income. At the same time it is an aggregate of knowledge, skills, habits, strains of a man [1]. The human capital is included into the category of capital for three reasons:

1) expenses for upbringing, education and training are real expenses;
2) the product of this work increases the national wealth:
3) the expenses incurred by a person that increase this product also will increase the national value.

The advocates of the human capital theory developed the quantitative methods to analyze the efficiency of investments in human capital. Education, medical care, job coaching, migration, giving birth to children and taking care of them, and their monetary return to their family and society are the factors of the human capital reproduction. Mainly these researches focus on human abilities and differences in income caused by different levels of investments in their reproduction. The quantitative result of these researches is the human development index (HDI).

HDI includes three main components of human life: life span, knowledge and the basic income to provide people with good standards of living. Life span and knowledge provide the process of giving the opportunities for people, their abilities; the income is a way to measure how people put into practice their abilities [2].

The values for all three components of HDI are aggregated in one composite index with help of the geometrical mean. HDI promotes to visual comparison of how different countries function.

3. The methods of human development index data analysis, modeling and forecasting

The development of these methods began from solving the applied problems of modeling and searching for the interdependencies in data [3-6].

The methods of solving of these problems were generalized; and the methods of human development index data analysis, modeling and forecasting were developed on its basis. The scheme of these methods is shown on the Fig. 1.

A **data sample** contains data about human development factors for countries or for a chosen region.

Recovery of omissions in data. There are many various ways how to recover omitted data. These ways depend on a form of this omission and the availability of sufficient information for its recovery. In the proposed methods the combinatorial GMDH algorithm is used to recover omissions.

In order to recover every omission its own data sample is composed and its own model is built because as many models should be built as there are omissions in data.

Data scaling. Taking into consideration that usually data are measured in different units and ranges, the procedure of data normalization or scaling is applied in order to bring data together in the same format. For that the values of every factor is divided on maximum value of this factor.

Data sample expansion. With the purpose of the possible increasing of a model's accuracy the method of data sample extension by introducing of new additional factors into it is used. These new factors can be products, sums, and other functions of initial factors [7].

Data analysis. This block includes two ways of analysis – with help of correlation analysis and GMDH. Correlation analysis is used to analyze the extent of impact of every factor on the output value and to select the informative factors. It is made by computing the pair correlation coefficient between a factor and an output value. One can come to conclusion about the dependencies in data using the combinatorial GMDH algorithm for data analysis. It is possible based on the developed models to analyze what factors were included in the model and with what coefficients.

Model construction is carried out by two methods:

1) Regression models are built. The coefficients are determined with help of the least-squares method. And the model structure is given;
2) Models are developed with help of the combinatorial GMDH algorithm with the adjusted exterior criteria. At the same time not only the parameters of the model are defined but its whole structure.

Depending on what of exterior criteria is given as the main criterion for the model selection the models with different characteristics can be obtained. The exterior regularity criterion AR is the main exterior criterion which is responsible for the model accuracy. The unbiasedness criterion is the criterion that describes the condition that the best models obtained on different parts of the sample should have minimum differences. In the GMDH theory there are the unbiasedness criteria of solutions CB and errors BS [8].

Comparison of models on the independent sample. The obtained models are compared on a some independent (examination) data sample that wasn't used to obtain them. The results of the research confirm the well known in the field of inductive modeling fact that regression models give a good result only on the data on which they were obtained unlike the GMDH models that give rather good results on new data too.

Analysis of received models and elaboration of recommendations. The models are analyzed from the point of view of both received parameters – how every of the parameters included into the model affects a given output value (the value of the coefficient with it and its sign (plus or minus) – and the accuracy of a model. Also one can analyze how the change of one or several parameters (their increase or decrease) will influence on the output value (by how many percent this output value will change).

Based on the analysis of models the recommendations about interrelations among parameters can be elaborated for the specialists in this field. On the grounds of the analysis of model coefficients and their signs (plus or minus) the contribution that every factor makes into the output value can be evaluated.

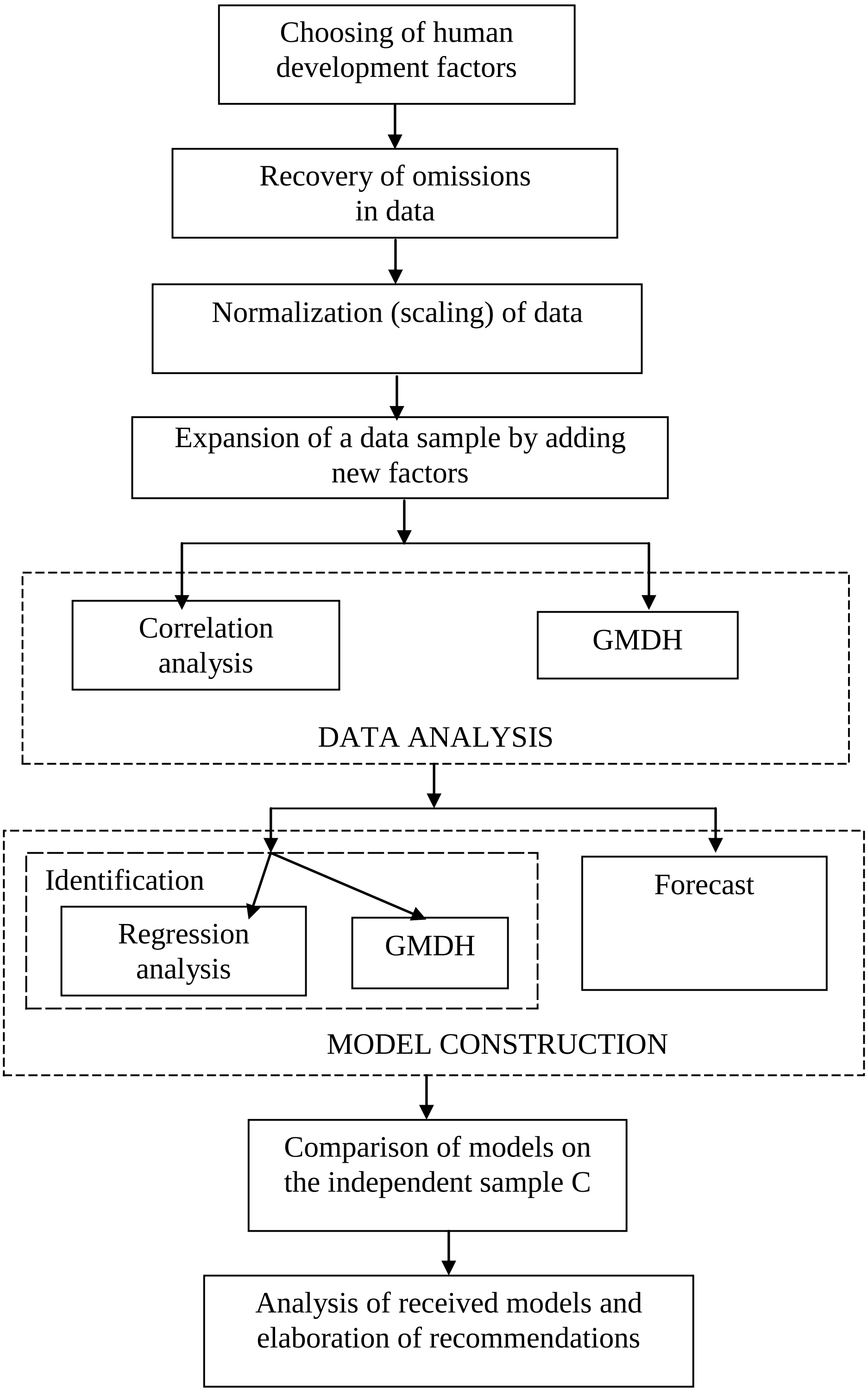

Рис.1. The flow-block of solution for the problem of data analysis, modeling and forecasting

Summary

Based on the inductive approach and correlation and regression analysis taking into consideration particular features of the problems of research in the field of the human development level of countries of the world and regions of Ukraine the methods of data analysis, modeling and forecasting were developed. Using this method one can solve the problems to recover omitted data, to select the informative arguments, to build the models for identification and forecasting, and also to extend the modeling space by adding new arguments, for example, the products and sums of initial arguments.

On the basis of this method the policy-makers in the social and economic field can develop recommendations to possibly correct some aspects of human development. It will give opportunity to improve the decision-making process in this field.

References

1. Тутова О.В. Концептуальні основи формування людського капіталу // Економіко-математичне моделювання соціально-економічних систем. - 2012. - Вип. 17. - С. 119 -129.

2. Tutova O. V. Conceptual Fundamentals of Assessment of a Human Development Level // Управляющие системы и машины. - 2014. - № 6. - С. 65 -71.

3. Savchenko Ie., Tutova O. Use of GMDH for investigation of impact of non-income components on HDI // Індуктивне моделювання складних систем. – 2012.–Вип. 4. – С.28-37.

4. Tutova O., Savchenko Ie. Modeling of Impact of Macroeconomic Indicators on the Growth of National Income // Proceedings of 4th International Conference in Inductive Modelling ICIM 2013. – Kyiv: IRTC, 2013. – P. 315-319.

5. Тутова О.В., Савченко Є.А. Вплив демографічних і соціальних показників на валовий регіональний продукт в областях України // Індуктивне моделювання складних систем. – 2015. – Вип. 7. – С. 271-281.

6. Савченко Е.А., Тутова Е.В. Прогнозирование валового регионального продукта областей Украины на основе индуктивного подхода // Управляющие системы и машины. - 2016. - № 3. - С. 84-92.

7. Савченко Е.А. Предварительная обработка данных в задаче индуктивного моделирования // Управляющие системы и машины. – 2015. – №2. – С.82 – 87.

8. Савченко Е.А., Степашко В.С. Анализ селективных свойств критериев МГУА при их последовательном применении // Моделювання та керування станом еколого-економічних систем регіону. Збірник праць. – Київ: МННЦ ІТС, 2008. – С. 199–210.

On stepwise estimation of parameters of the nonlinear Solow model of economic dynamics

Pavlo Stepashko, Volodymyr Stepashko

International Research and Training Centre for Information Technologies and Systems of the NAS and MES of Ukraine; Glushkov Ave, 40, Kyiv, 03680, Ukraine

pstepashko@gmail.com stepashko@irtc.org.ua

Abstract. *The paper considers the problem of estimating the parameters of a first-order nonlinear differential equation describing the dynamics of a separate sector of the economy* (*Solow model*). *Since the parameters are included in the equation nonlinearly, an approach is proposed that allows one to gradually estimate some of them on the basis of statistical data, and at each stage the problem turns out to be linear*.

Keywords
Economy dynamics, Solow model, nonlinear differential equation, parameter estimation

5 Single-sector model of economic dynamics

In this model, the state of a separate (autonomous) sector of the economy is typically characterized by 5 variables [1]:

Y – the volume of the final product fully used in the sector;

C – volume of consumption fund (gross non-production expenses);

S – volume of saving fund (gross expenses for development);

L – volume of labor resources (labor);

K – volume of fixed assets (capital).

Resources K and L are fully used, then $Y = F(K, L)$ is the production function.

The product goes to consumption and saving (non-production and production expenses): $Y = C + S$.
These components are fixed: $S = sY$, $0 < s < 1$, $s = const$ is unchanged rate of saving. Obviously, $S = sF(K, L)$, and accordingly $C = (1-s)Y$.

Gross saving S is spent on net capital gain $\dot{K}$ and on the recovery fund μK of outgoing fixed assets, so $S = \dot{K} + \mu K$, $0 < \mu < 1$, $\mu = const$ – rate of deductions for amortization.

Dynamics of labor resources is described by typical Malthusian growth model): $\dot{L} = gL$, where $g = const$ is growth rate of labor.

These are the main assumptions of the *Solow model*. Everything is determined here by the dynamics of the volume of fixed assets (capital) $\dot{K}$, in addition to labor resources. Therefore, it is necessary to get the equation for the dynamics of capital gains per unit of labor, for which all values are normalized per unit of labor.

In this case, we use the fact that the production function is homogeneous: $F(\lambda K, \lambda L) = \lambda F(K, L)$. Entering values $k = \frac{K}{L}$ as capital-labor ratio and $y = \frac{Y}{L}$ as output per capita (labor productivity), we obtain that the production function will depend on one variable: $y = F(\frac{K}{L}, 1) \overset{\Delta}{=} f(k)$, so this is the dependence of labor productivity on capital assets. Function $f(k)$ is a strictly monotonously growing and concave (with an increase in capital, the growth rate of product volume is slowing down).

With this in mind, an equation for the dynamics of capital ratio is derived, that is, the basic equation of the model of economic dynamics is obtained in the following form:

$$\dot{k} = sf(k) - (\mu + g)k, \ k(0) = k_0 . \tag{1}$$

This is a nonlinear model; evidently that the strictly monotonically increasing over k and concave function $sf(k)$ intersects the linear function $(\mu + q)k$ at some point k^*, so when $sf(k) = (\mu + g)k$, this model has a stationary mode that meets the condition $\dot{k} = 0$.

It is obvious that the stationary solution of this model depends on the parameters of the equation, the initial conditions, and also on the specific variant of the production function. As a rule, the most known *Cobb-Douglas* production function is used: $Y = F(K, L) = a K^{\alpha} L^{(1-\alpha)}$, $0 < \alpha < 1$, where α is the share of capital in the product, $1-\alpha$ the share of labor, a – coefficient characterizing the level of development of a particular economy.

Then the normalized Cobb-Douglas function takes the form $f(k) = a k^{\alpha}$, that is, the basic model of economic dynamics becomes a nonlinear differential equation of the first order:

$$\dot{k} = s a k^{\alpha} - (\mu + q) k \quad \Big| \quad k_0 . \tag{2}$$

From that, it is easy to analytically determine a stationary solution $k^* = \sqrt[1-\alpha]{\frac{sa}{\mu + g}}$.

As one can see, the parameters s, a, α, μ, g are included in the equation in a nonlinear way, which complicates the task of evaluating them for specific conditions. It should be noted that a model of this type can be applied to a separate enterprise, an economic sector or a country economy as a whole [2].

Before proceeding to the problem of estimating the parameters of equation (2), let us consider what variants of solutions such a model can have.

6 Typical solutions of the Solow model

Since equation (2) has a stationary point k^* (which is determined by the specific given values of its parameters), all its solutions with any initial conditions are approaching that point. To determine the nature of possible solutions of this nonlinear equation we use the WolframMathematica system.

Let the following values of the equation parameters be given (2): $s = 0.6, a = 2.5, \alpha = 0.5, \mu = g = 0.1$. Using the NDSolve function, we find numerical solutions $k(t)$ of such an equation under different initial conditions.

It is found that this equation has three typical variants of solutions, which correspond, for example, to the value of initial conditions: $k_0 = 1$ (logistic function of time); $k_0 = 30$ (growing exponent); $k_0 = 90$ (declining exponent). Appropriate solutions are shown on the graph fig. 1.

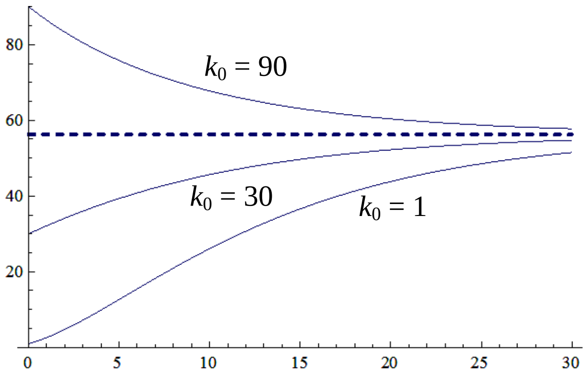

Fig. 1

7 Estimation of parameters of the model

Consider now the problem inverse to the previous one: let a discrete set of n numerical values $k(t)$ be given from the solution of equation (2), obtained numerically for a certain value of the initial condition; it is necessary to find such values of the equation parameters that correspond to this given solution. This is the task of identifying the parameters of the model (2). Note that the choice of a particular graph does not matter, because they are all received for one set of parameters.

So let as have a discrete set of values $k(t) \equiv k_t$ with a certain step on time, in the simplest case Δt=1: $t = 1, 2, \ldots, n$. This set corresponds to some difference approximation of equation (2); for example, assuming that $\dot{k}_t \approx k_{t+1} - k_t$, we obtain the following difference equation:

$$k_{t+1} = sak_t^{\alpha} + (1 - \mu - g)k_t \,,\; t = 1, 2, \ldots, n \,. \tag{3}$$

Entering the notation $c_1 = sa, c_2 = 1 - \mu - g, v_{1t} = k_t^{\alpha}, v_{2t} = k_t$, $w_t = k_{t+1}$, we get a linear dependence $w_t = c_1 v_{1t} + c_2 v_{2t}, t = 1, 2, \ldots, n$, generally applicable for applying the least squares method (LSM) for the evaluation of generalized intermediate parameters c_1 and c_2 . But parameters s, a, α are included in the equation in a nonlinear way, and μ, g are components of one coefficient c_2 . Therefore, it is impossible to apply here the LSM directly.

However, if we take into account that in practical problems of modeling of real economic processes we can obtain statistics on time-varying quantities Y_t, K_t, L_t, $t = 1, 2, \ldots, n$, we can offer a sequential evaluation of unknown parameters in three stages:

1. Based on the difference approximation of the equation for dynamics of growth of labor resources $\dot{L} = g\,L$, get an estimate of the parameter g by LSM; let us consider it given from now, i.e. $g = 0.1$.
2. Calculating the normalized values for labor productivity y_t and capital ratio k_t , based on n values of normalized production function $y_t = a\,k_t^{\alpha}$ determine the parameters a and α . To do this, we first find the logarithm of this function:

$$\ln y_t = \ln \alpha + \alpha \ln k_t \,.$$

Entering these symbols: $b_1 = \ln \alpha, b_2 = \alpha, x_{1t} \equiv 1, x_{2t} = \ln k_t$, $z_t = \ln y_t$, we can apply the standard LSM program LeastSquares for estimating unknown parameters of the linear dependence $z_t = b_1 x_{1t} + b_2 x_{2t}$. In particular, using the values k_t obtained after the solution of equation (2) with n=10, one can calculate the value y_t and round-off them up to two decimal points (if this is not done, the result of further calculations will obviously be the exact given values a and α).

Then we use the obtained values of vectors

$$z = (\ln y_1, \dots, \ln y_n)^T,\ x_1 = (1, \dots, 1)^T,\ x_2 = (\ln k_1, \dots, \ln k_n)^T$$

as the input for the LeastSquares function. As a result of the calculations, we obtain the following coefficients: $b_1 = 0{,}91616$ i $b_2 = 0{,}500027$, then the required parameters are: $b_2 = 2{,}4997 \approx 2{,}5$, and the index of the exponent of production function (the share of capital in the product) $\alpha = 0{,}500027 \approx 0{,}5$. Consequently, we have obtained sufficiently precise results of the evaluation of parameters a and α.

3. Knowing the parameter α, on the basis of the difference approximation (3) of equation (2) it is possible to form data for the evaluation of the intermediate parameters of the algebraic equation $w_t = c_1 v_{1t} + c_2 v_{2t}$, $t = 1, 2, \dots, n$, according to (3). For this, given $n = 10$ values of the variable k_t, we first form the vector of the dependent variable w, then the vector of the second independent variable v_2 and, finally, we calculate the vector of the first independent variable v_1.

4. Let us use thus obtained values of these vectors

$$y = (k_2, k_3, \dots, k_n)^T,\ v_1 = (k_1^{\alpha}, k_2^{\alpha}, \dots, k_{n-1}^{\alpha})^T,\ v_2 = (k_1, \dots, k_{n-1})^T,$$

each of which has a length $l = n - 1$, as input to the LeastSquares function. As a result of the calculations, we obtain the following values of the coefficients of the algebraic equation $w_t = c_1 v_{1t} + c_2 v_{2t}$: $c_1 = 1{,}6501$ and $b_2 = 0{,}7633$.

Obviously, the required parameters of equation (2) are now calculated as $s = \dfrac{c_1}{\alpha}$, $\mu = 1 - c_2 - g$, where α and g are known; as a result, we obtain the following values of estimates for these parameters: $s = 0{,}6601$; $\mu = 0{,}167$. The inaccuracy of these estimates is due to the too large step of the difference approximation, which is equal to 1.

8 Conclusion

The problem of estimating the parameters of a nonlinear differential equation that describes the dynamics of an autonomous economic sector (Solow model) is considered in the paper. Because five unknown parameters s, a, α, μ, g of this model are included in the equation in a nonlinear way, an approach is proposed that allows to estimate them according to statistical data in three phases: first, the growth rate of labor resources g from the corresponding equation of dynamics, then parameters a (level of development of a particular economy) and α (the share of capital in the product) from the normalized production function, and, finally, the rate of saving and the rate of amortization μ from the equation of capital change dynamics (2). It is important that at each stage the task is linear, and this allows applying a standard LSM program.

The efficiency of the proposed approach is shown in the test case, when the parameters included in the model (2) were given in advance and used to obtain its solution and appropriate data sample. Both for solving the equation (2) (direct problem) and estimating its parameters (inverse problem), the Wolfram Mathematica system was used.

The described approach can be used to identify a model of this type based on statistical data on the functioning of a particular enterprise, economy of a particular sector or a country as a whole.

References

1. Solow, R.M. A Contribution to the Theory of Economic Growth // The Quarterly Journal of Economics, 1956. February Vol. 70, No. 1. P. 65-94.

2. Romer, David (2011). The Solow Growth Model. Advanced Macroeconomics (Fourth ed.). New York: McGraw-Hill. P. 6–48.

Development of Information Technology for Solution of Casting Production Monitoring Tasks

Tokova O., Savchenko Ye.

International Research and Educational Center of Information Technologies and Systems of NAS of Ukraine, pr. Academika Glushkova, 40, Kyiv, 03680, Ukraine

len327@ukr.net, savchenko@irtc.org.ua

Abstract. *Development of information technology for monitoring tasks solution in the foundry industry based on an inductive approach is shown. The technology is designed to solve the problem of searching dependence of the casting quality on its speed and cooling mode as well as select in the process of experimentation the best casting, by comparing it with a standard cooling curve.*

Keywords

Information technology, inductive modeling, casting,
casting production,

9 Introduction

The purpose of this operation is development of an information technology for the decision of tasks of monitoring in foundry production which is aimed at improvement of quality of the final product. Depending on composition of casting, the modes and speed of its cooling, the sequence of adding of the necessary components in alloy it is possible to receive the final product of any quality. Quality of a product is estimated by the expert by means of different express-analyzes. Improvement of quality of castings is the principal job of foundry production which will influence not only mass purchase of products of casting, and also will provide economic stability and overall performance of the enterprises. Foundry production considerably defines economic development of the country in general.

Unfortunately for today's time foundry production in Ukraine falls into decay because of absence of financing, high cost of raw materials which is ordered more often behind a cord than get in the middle of the country. Absence of the necessary equipment, is computer technique and the necessary software products on productions doesn't give the chance to our founders to compete with oversea vendors. Unfortunately on today's time a casting production in Ukraine falls into a decay on reason absences of financing, by costliness of raw material which is more frequent ordered abroad what obtain in the middle of country. Absence of necessary equipment, a computer technique and necessary software products on productions is not given by possibility to compete our foundry hands with oversea producers.

Use of computer technologies for data analysis and simulation of processes in foundry production will be essential to facilitate process of receiving casting (from his fusion to complete consolidation), will decrease appearance of defective details on a production, the same a requirement will disappear in processing of wares, and the process of outpouring will become more economic advantageous and cheap.

Thus the task of construction of computer technology is in-process put with a comfortable, clear for an user interface for the decision of tasks of the optimum mode selection of cooling for the receipt of founding of the set quality, and also search from the set base of information for founding of unknown fusion of the nearest to him standard by comparing of curve of his cooling to the standards of this base for the estimation of his chemical composition.

Problem of upgrading casting products to the international standards and expansion of markets of its sale as in the middle of country so in the world remains relevant. In spite of variety of developed software for a casting production, there is a requirement in development comfortable in the use and main accessible on a cost product.

2 Problem statement

Foundry production call processes of receiving shaped products (castings) by a flood of molten metal or plastic in the empty form which reproduces the form and the sizes of future detail. After metal hardening in shape receive casting or a detail which give in then to machining (abrasion) and are used in different spheres. Castings widely apply in mechanical engineering, metallurgy, construction and other [1]. Depending on the mode of cooling of casting it is possible to receive the final product of different necessary quality [2]. Quality of such product is estimated by the expert by means of different express-analyzes. For example, finite casting shall have dispersible compact-grained structure. Requirements to the final product can be set or the customer, or are defined according to data which are set in bases which contain the international quality standards for each type of a product.

In spite of the fact that many software products [3-5] which meet requirements of workers in branch of foundry production, the task of creation convenient in use, and, above all a product, available at cost, are at the moment developed, remains urgent. We will direct the review of the existing software products, their features and the functional capabilities (Tab. 1) below.

Tabl.1

Opportunities and shortcomings of the existing software

№	***Name of product***	***Possibilities***	***Failings***
1	**SolidCast** **Developer**: Finite Solutions Inc., USA	1. Metal filling (fast and normal) 2. Metal hardening: placement of thermal nodes and shrinkable defects 3. Automatic optimization of LPS	1. Inconvenient interface; there is a possibility of the translation of "Menu" into Russian. 2. Inconvenient output of results of calculations for the visual analysis; 3. There is no domestic database with materials and alloys; 4. In the database of properties, there are constant values on all interval of temperatures; 5. Long-time computer calculation 6. Problems with STL import
2	**Magmasoft** **Developer:** Magma GMBH (Aachen, Germany)	Magmasoft models the forms completed with metal, crystallization, calculation of tension and deformation, structure, distribution of a perlite and ferrite, and it is a lot of other useful opportunities. The program recognizes the following types of casting in case of simulation. In basic MAGMASOFT standard option allows to model casting in sandy – clay forms and in the chill mold. For simulation of other types of casting it is necessary to buy additional modules.	1. High cost 2. Very heavy for understanding interface without support of Russian 3. Absence of the Ukrainian (domestic) database of metals and alloys; 4. Big hour costs of introductions this.
3	***CASTCAE*** **Developer:** CT-Castech Inc. Oy, Espoo (Finland)	Casting under high pressure; 2. Casting under low pressure; 3. Casting on the melted models; 4. Casting in the chill mold; 5. Casting in the sandy form;	CASTCAE is a program of initial level. It is possible to use this program for casting not of difficult castings. In recent years has no information on an output of new versions of this program.
4	***JSCast*** **Developer: Osaka, Japan.**	1. Casting on the melted models; 2. Casting in the chill mold; 3. Casting in the sandy form; 4. Casting under high pressure; 5. Metal filling; 6. Metal hardening: placement of thermal nodes and shrinkable defects.	1. Small choice of opportunities of the program; 2. Absence of the Ukrainian (domestic) database of metals and alloys.

5	***AnyCasting*** **Developer:** AnyCasting Co., LTD (Korea)	The program simulates: Metal filling; Metal hardening: placement of thermal nodes and defects Tension; Structure; In case of simulation the program distinguishes the following types of casting: Casting on vacuum plate form; On the melted models; Casting in the chill mold; Casting in the sandy form; Casting under low pressure; Casting under high pressure;	There is no Russian interface (translation); Absence of the Ukrainian (domestic) database of metals and alloys.

The carried-out review showed that a row of software products is developed for the solution of the main problems which arise in the course of production of castings, but use of these products is impossible in connection with high technical requirements to the computer equipment, the high cost of these software products and absence of the domestic database of metals and alloys. Therefore at Institute of metals and alloys there was a need of developing convenient in use and a software product, available at the price, for the decision of tasks of data analysis and simulation.

3 Technology of data analysis and simulation of processes in foundry production

Constant observation or monitoring in the course of foundry production will give the chance to founders to control process of formation of casting for preventing of formation of low-quality products.

In general monitoring is a procedure of execution of the current observation, monitoring, estimation, the analysis and prediction of processes [6]. In foundry production it is possible to define such tasks of monitoring:

- collection of the experimental data and observation (monitoring) over casting process;
- estimation of a current status of process of casting;
- the analysis of temperature dependences of cooling of a melt in the course of casting;
- express forecast of quality of casting;
- display of results of monitoring.

In technology which is developed, planned the decision all five tasks of monitoring [7-9].

Unit 1. Receipt of founding. This block characterizes the stages of process of casting, from preparation of alloy to the receipt of founding. Authors do not decide a task on these stages, but this block is included for possibility of interference with the process of cooling of founding and correction of chemical composition of founding. Quality of the got founding is estimated an expert and founding of incongruous quality can be again molten.

Unit 2. Data handling and simulation. In this unit such tasks are decoupled.

The formalizations this about the cooling modes. It is applied when it is necessary to transfer the description of the mode to the digital form. For example, the mode is described so: "vibration during 20 min + 5 min. evacuation". Variable value "vibration" will be 20, the evacuations variable - 5.

Previous processing of selection this. The experimental data can be processed, for example normed or scaled if there are passes they can be renewed, or by means of correlation analysis variables which have the greatest value of the module of coefficient of conjugate correlation of a variable with the initial value can be selected.

Creation of process models of cooling on the basis of GMDH. Models behind a combinatorial algorithm of MGUA are built, using external criteria of accuracy.

Assessment behind the given criteria of accuracy and adequacy of model. Quality of the received model can be estimated on selection which wasn't used in case of simulation – examination selection.

Visualization of results. Results are brought in the form of diagrams and models out of the accounting of their accuracy.

Application. By means of technology which is developed the search of temperature dependence of cooling of cylinder casting from Sch20 task cast iron task in a sandy casting mold is solved from the parameters of the mode of its cooling for the subsequent optimization of this mode. The models of dependence of the cooling temperature on its time for different cooling modes have been constructed [9].

The technology which is developed contains such main units. On Fig. 1 it is shows the flowchart of technology which is developed.

Block 1. *Receipt of founding.* This block characterizes the stages of process of casting, from preparation of alloy to the receipt of founding. Authors do not decide a task on these stages, but this block is included for possibility of interference with the process of cooling of founding and correction of chemical composition of founding. Quality of the got founding is estimated an expert and founding of incongruous quality can be again molten.

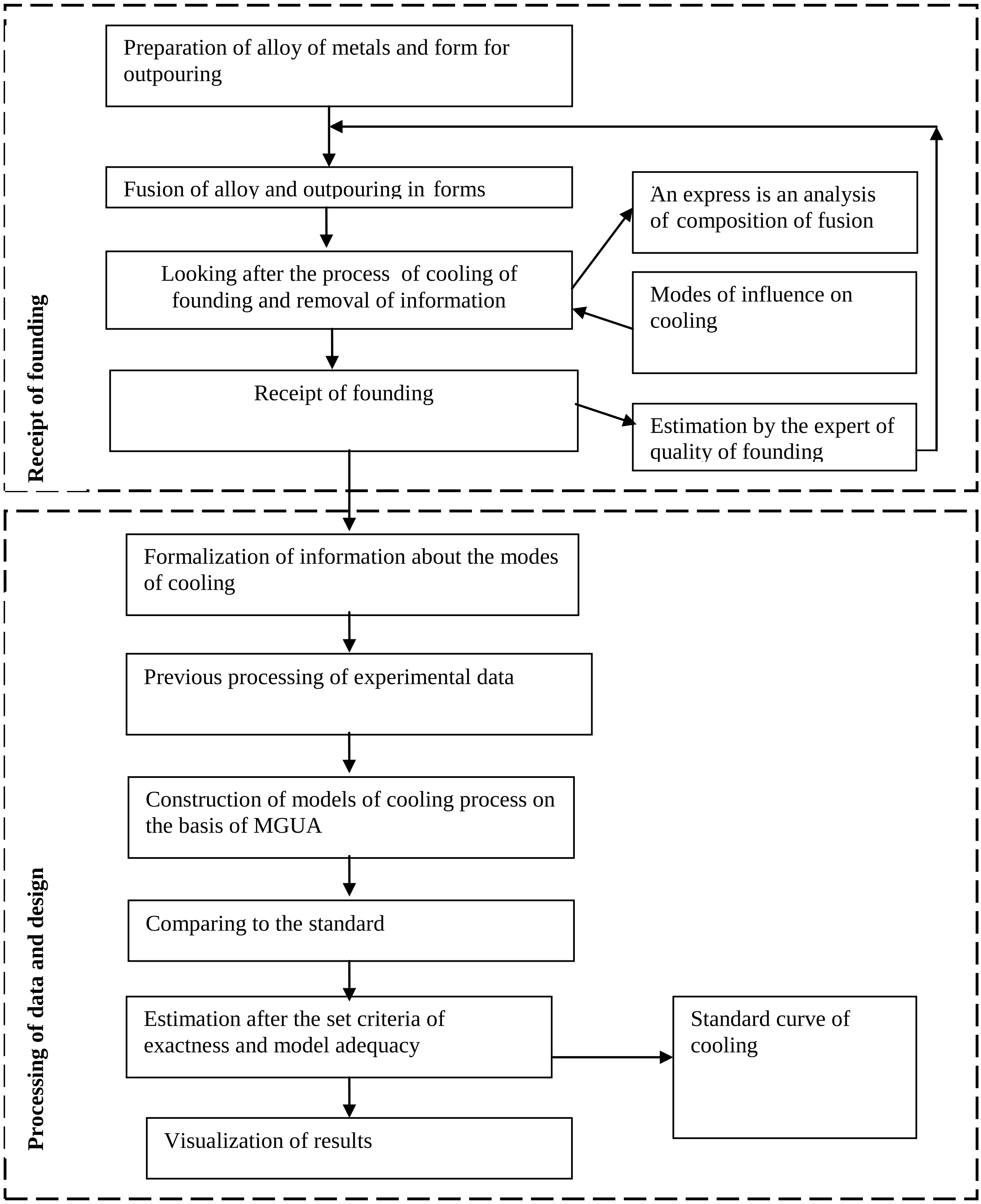

Fig. 1. Diagram of technology of the decision of tasks of monitoring of processes of foundry production

Intended for the grant of recommendations specialists in industry of casting production in relation to the analysis of present raw material, having a special purpose selection of constituents of founding, and also support of making a decision at the choice of the modes of cooling with the purpose of receipt of founding of the set quality.

The technology is developed on the basis of MS Excel as a part of three main subsystems for:

1) target to a p_dbor of components of casting;

2) the analysis of the available raw materials and recommendations concerning its application;

3) supports of a choice of the mode of cooling. Creation of the domestic database of metals and alloys is planned.

Technology is developed on a base MS Excel in composition three basic subsystems for:

1) having a special purpose selection of constituents of founding;

2) analysis of present raw material and recommendations in relation to its application;

3) supports of choice of the cooling mode. Creation of domestic base of these metals and alloys is planned.

4 Conclusion

Having carried out the review of the existing software it is possible to draw the following conclusions:

1. All programs are very expensive, besides each of programs needs constant addition by additional components, continuous updating of the old version of the program.

2. Programs are developed by foreign experts therefore they are written in language of the developer that not always understood to our Ukrainian expert, and also have the difficult interface therefore require clarification and detail to instructing.

3. Considering that at many domestic enterprises rather decayed equipment is installed (computers, instruments for measurements) which requires or completely up-dating of operating systems, video cards, or generally complete changeover of the old equipment, on new therefore there are a lot of software developed on improved the computer can't be used on domestic.

4. Above and above the considered means use a method of finite differences therefore in castings of responsible assignment such programs aren't suitable for elimination of shrinkable porosity because the applied mathematical method badly works in case of thin-walled castings when thickness of walls becomes comparable with slot pitch.

5. Also a lack of the existing programs is absence of the domestic database of metals and alloys, and it in turn complicates operation of the founder.

Reference

[1] Basics of technology of the most important industries. Part 2: the manual / Ed. I.V. Chentsova, V. V. Vashchuk, Minsk: Higher School, 1989, 199 p.

[2] Sanacheva G.S., Stepanova T.N., Baranov V.N., Gubanov I Yu. Technology of Foundry Production: Uch. - Krasnoyarsk: SFU, 2008 - 86 p.

[3] http://www.magmasoft.com

[4] http://www.dial-engineering.ru/magma.html

[5] http://www.ruscastings.ru/work/168/2130/2984/2988

[6] Ivakhnenko A.G., Stepashko V.S. Noise immunity modelling, Kiev: Naukova dumka, 1985, 215 p.

[7] Ivakhnenko O.G., Ivakhnenko G.O. Inductive methods of forecasting the analysis of folding economical systems, Economist, 1998, No. 5, P. 88-97.

[8] Aystrahanov D.D., Pugachova M.V., Stepashko V.S. et al. Conceptual basis of statistical monitoring, K .: IVTS Derzhkomstatu Ukrayiny,2003, 343 p.

[9] Savchenko, Ye.A., Kravchenko O.V. Application of the inductive approach for modeling the cooling process of casting according to experimental data. Inductive modeling of complex systems. Inductive modeling of folding systems, K.: MNNTS ISTS, Vyp. 6, 2014, P. 126-136.

Qualitative Approach to Environmental Modeling Based on GMDH Models

V. Vissikirsky

International Research and Training Centre of Information Technologies and Systems of the NAS and MES of Ukraine, Akademik Glushkov Prospect 40, 03680 Kyiv, Ukraine

vlanvis@gmail.com

Abstract. *The paper considers the main aspects of the GMDH application together with deductive qualitative assessment techniques, which provides a consistent basis for the development of integrated tools intended to cover the modeling, analysis, and decision-making tasks in the field of environment and agriculture.*

Keywords

Environmental modeling, species irrigation, processed wastewater, qualitative approach, GMDH, decision making support.

1 Introduction

Wide utilization of wastewater and sludge for agricultural purposes requires a thorough investigation of the behavior of various species in biological, chemical, and physical aspects under different wastewater supply conditions. The problem involves many factors and parameters, such as the impacts of treatments on soil and flora, survival and mortality rates of species, their growth, silvicultural issues (roots and crown development, competition), wood properties, assimilation behavior for different chemical elements, etc.

The investigation process is virtually an iterative sequence of empirical procedures intended for the identification and selection of species with stable behavior and properties "good" enough to utilize them for agricultural purposes under cultivation with processed wastewater. These procedures involve the analysis of the problem at different degrees of knowledge detalization and certainty. The investigator may have a detailed look at a single direct effect of a chemical substance on a species, or consider a chain of interactions by trying to generalize and find out a similarity or distinctiveness in the behavior between different species.

Therefore, under uncertain multifactor conditions the search by using only conventional statistical methods becomes a trial-and-error procedure. Two main features of this process shall be emphasized: empirical estimation of complex nonlinear functional relationships between interacting objects in ecological network, and qualitative reasoning about such relationships to reveal stable characteristics of the objects and regularities of their behavior.

2 Peculiarities of the problem

More deep understanding of the behavior of species requires its investigation in different aspects and under different experimental plans, which at the same time substantially complicates the analysis of empirical results by using conventional statistical techniques. Important aspect of investigation is the analysis of how the concentrations in different species change and what would be the reaction of species when the concentrations in irrigated water increase up to the high toxic level.

The proposed solution to the development of computer-based intelligent tools is intended for the empirical modeling of species' behavior and qualitative reasoning with imprecise information about this behavior.

The studies aimed to estimate characteristics of different plant species irrigated with processed wastewater, can be carried out with use of GMDH methods that allow us to analyze various aspects of plant species behavior: mechanical properties of trees (axial compression and bending) depending on concentrations of chemical elements contained in the wastewater [1], growth dynamics of trees irrigated with wastewater (development of tree height and mortality over time) [2, 3], etc. After that, qualitative analysis techniques can be applied to assess plant species behavior under different conditions: irrigation of forest species with wastewater and sludge [4].

3 Characterizing the approach

Here, the key approach to resolve the problem consists in the decomposition of a modeled system into subsystems, to provide the modeling and analysis of subsystem(s) that have subset(s) of input variables consistent with volume of sample available. The objective of the GMDH modeling is to obtain the regression models describing different relationships within the chain of interactions "soil - roots - leaves", as well as to provide further qualitative analysis of involved factors. Qualitative assessment provides initial generalization of regularities in change of heavy metals concentrations in soil, roots, and leaves, depending on distance from the road. It also allows us to determine the level of certainty / uncertainty of the models and to direct future experiments to obtain more detailed information.

The set of generated GMDH models gives the possibility to analyze the system from different viewpoints, according to the model accuracies and structures. In the decomposition of a modeled system into subsystems, the related sets of GMDH models can be analyzed to identify the most relevant structures and make assumptions about input of different factors, assess the uncertainty level regarding each factor input, as well as to make decision on how to design the next-stage experiments to obtain more information about system behavior. This also makes it possible to compare the sets of GMDH models describing different subsystems by their structural similarity / difference, hence, merge the models into a model that describes a higher-level subsystem.

In addition, this provides the basis to develop simple reasoning procedures for qualitative assessment of system behavior at different stages of long-term experiments.

4 Conclusions

Qualitative analysis gives the possibility, in terms convenient for the end user, to analyze behavior of individual subsystems as well as of the system in whole. It is intended to facilitate clear understanding of the experimental results, easily classify different irrigation/plant species cases and identify the best ones. In general, the application of GMDH methods together with deductive qualitative assessment techniques provides a consistent basis for the development of integrated tools intended to cover the modeling, analysis, and decision-making tasks in the field of environment and agriculture.

References

[1] Kalavrouziotis I. K., Stepashko V. S., Vissikirsky V. A., Drakatos P.A.: Group Method of Data Handling (GMDH) application for modelling of mechanical properties of trees irrigated with wastewater. *International Journal of Environment and Pollution*, p. 589-601, 18 (6), 2002.

[2] Kalavrouziotis I. K., Stepashko V. S., Vissikirsky V. A., Drakatos P.A.: Group Method of Data Handling (GMDH) in modelling of growth dynamics of tress irrigated with wastewater. *International Journal of Environment and Pollution*, p. 350-364, 21 (4), 2004.

[3] Vissikirsky V.A., Stepashko V.S., Kalavrouziotis I.K., Drakatos P.A.: Growth Dynamics of trees Irrigated with wastewater: GMDH Modeling, Assessment and Control Issues. *Instrumentation Science and Technology*, p. 229-249, 33 (2), 2005.

[4] Kalavrouziotis I. K., Pantazis V. N., Vissikirsky V. A., Stepashko, V. S.: Irrigation of Forest Species with Wastewater and Sludge: Experiments, Modeling, and Qualitative Assessment, *Instrumentation Science and Technology*, p. 727–742, 34, 2006.

Dynamics Modeling of Multidimensional Interrelated Processes

Serhiy Yefimenko

International Research and Training Centre of Information Technologies and Systems of the NAS and MES of Ukraine, Akademik Glushkov prospect 40, 03680 Kyiv, Ukraine

syefim@ukr.net

Abstract. *The paper considers theoretical grounds of recurrent computing applying for dynamics modeling of multidimensional interrelated processes. The combinatorial GMDH algorithm is used for vector autoregressive (VAR) modelling. The procedure of structural and parametric identification of VAR models is proposed.*

Keywords

Inductive modelling, VAR, recurrent parameters estimation, combinatorial GMDH algorithm

5 Introduction

An approach to the problem of structural and parametric identification of vector autoregressive models, when parameters for every model are estimated independently, is considered. The drawback of the approach is that the parameters of separate models for interrelated processes are interdependent. The paper uses an algorithm with selection not one but a few best models for every process to eliminate the shortcoming.

We use the Group Method of Data Handling (GMDH) [1] as one of the most effective inductive modelling methods. When solving the problem of vector autoregressive modelling, the run-time is one of the most important criteria of efficiency. The most effective way to achieve the high performance is recurrent parameter estimation. In this case it is more expedient to use the high-speed methods of parameters estimation based on the recurrent algorithms for solving of linear equations systems.

2 Vector Autoregressive Modelling

The vector autoregressive (VAR) model generalizes the model of autoregression to multidimensional case. It is built by the stationary time series. It is the system of equations, in which every variable (component of multidimensional time series) is linear combination of all variables in the previous time points. The order of such model is determined by the order of the lags.

In the general case for *m* time series and *k* lags the model will be the system of *m* equations (in matrix form):

$$X(t) = \sum_{j=1}^{k} \Theta_j X(t-j), \tag{1}$$

where Θ_j, $j = \overline{1,k}$ – matrices of model parameters (3) of size $m \times m$.

3 Procedure of Structural and Parametric Identification of VAR Models

The general models structure in the form of a system of *m* difference equations is determined as a result of the sequence

of such operations:
1. Data array of $m \cdot k$ arguments is composed under the number of interrelated processes *m* and lags *κ*.
2. Maximal complexity for restricted search is defined. COMBI algorithm with sequentially complicated structures of models on the basis of recurrent computing is used for modelling. For every time series the best *F* (by the value of the regularity criterion [2]) models are selected. Overall $F \cdot m$ models are passed to the next step.
3. The sorting of $G = F^m$ possibilities of systems of models is carried out. The best model (by the value of the systemic integral criterion of vector models quality) is selected. The value of the criterion is calculated on the given part of initial data set in the prediction mode of the process for the given steps number n_S:

$$B = \sum_{i=1}^{n_S} \sum_{j=1}^{m} (x_{ij} - x_{ij}^*)^2 , \tag{2}$$

where x_{ij}^* is the result of step-by-step integration (recursive computations) of the system of *m* equations.

4 Combinatorial GMDH Algorithm with Recurrent Parameters Estimation

The combinatorial algorithm is used for VAR modelling by exhaustive search of all possible variants and finding the best model for every time series containing the most informative subset of input arguments (regressors). It consists of such main units:

- data conversion according to a basic class of models (linear in parameters);
- forming models of different complexity;
- calculation values of external quality criteria for all models being formed;
- selection of the best models.

For linear object with *m* input, all possible models are compared in the process of exhaustive search. Total quantity of all generated models of the type

$$\hat{y}_v = X_v \hat{\theta}_v, \; v = 1, \ldots, 2^m - 1 \tag{3}$$

is 2^m–1. Decimal number *v* corresponds to binary number d_v in (3). Unit elements of d_v indicate inclusion regressors with corresponding numbers in the model, whereas zero elements signify exclusion.

Due to the exponential growth of 2^m as function of arguments amount, it is advisable to use algorithms recurrent in the number of parameters in structural identification problems for the parameters estimation.Efficient recurrent modifications of classic Gauss and Gramm-Schmidt algorithms were offered in [3]. The modification, in a nutshell, is as follows. The matrix $H_s = X_s^T X_s$ of the size $s \times s$ is reduced to superdiagonal form by computing only elements $h_{i\,s}^s$, $i = \overline{2, s-1}$, $h_{s\,i}^s$, $i = \overline{2, s}$, and $g_s = X_s^T y$ at every step s, $s = \overline{1, m}$ during the direct motion. The elements of the nested matrix H_{s-1} of size $(s-1)\times(s-1)$ (reduced to superdiagonal form on the previous step) remain changeless. So only "bordering elements" are computed on step *s*.

5 Conclusion

The technique of recurrent computing in COMBI GMDH algorithm for building discrete prediction models of complex multidimensional interrelated processes is proposed. Software tools for modelling of complex multidimensional interrelated processes on the basis of high-performance combinatorial algorithm in the class of discrete dynamic VAR models are developed.

References

[5] Madala, H.R., Ivakhnenko, A.G.: Inductive Learning Algorithms for Complex Systems Modeling. CRC Press, New York (1994).
[6] Farlow, S.J.: Self-Organizing Methods in Modelling: GMDH Type Algorithms. New-York, Bazel: Marcel Decker Inc., (1984)
[7] Stepashko, V.S., Efimenko S.N.: Sequential Estimation of the Parameters of Regression Model. Cybernetics and Systems Analysis 41(4), 631–634 (2005).

The Significant Features for Search Engines Ranking Results

Viacheslav Zosimov

Mykolayiv V.O. Sukhomlynsky National University, 54030, Ukraine, Mykolayiv, Nikolska str., 24

zosimovvv@bk.ru

Abstract. *The article describes an experiment to promote a web-resource in search results. During the experiments, the effect of quantitative indicators ranking features the position of web resources in the ranking of search engines was studied. Ranking signs for the experiments were selected using a generalized iterative algorithm for the group method of data handling.*

Keywords

Search engine rankings, GMDH, web resources, information search.

Introduction

Today it is difficult to imagine a person who would not use search engines on the Internet at least once a day. The requirements to serch engines grows depending on the development of the World Wide. Their algorithms had been changing. Continuously growing number of web resources forced the search engines developers to look for ways to speed up its algorithms. But the data processing speed is not the only or the most important problems for developers. The main objective is to create the optimal search results ranking model. It must respond to several requirements:

- respond the information needs of all users;
- provide high value completeness and accuracy of the search;
- effectively filter out search engine spam, sites clones;
- have a complex internal structure to minimize the possibility of external influence on search results.

Building the ranking model

The Google search engine [1] has been selected as a base for experimentation as the most demanded and technically equipped. It is necessary to build actual search result ranking model to get more accurate results of the experiments. For this procedure was used the technique described in [2]. The criteria described in [2] were used as ranking features.

To construct the ranking model it was used the inductive approach, namely a generalized iterative algorithm for group method of data handling [3].

Ranking model:

$$y = 0{,}76x_2 + 1{,}84x_3 + 0{,}15x_4 + 0{,}008x_7 - 1{,}9x_{11} + 0{,}012x_{22} - 14{,}8x_{27} -$$
$$- x_{28} - 14{,}33x_{35} + 2{,}4x_{40} - 0{,}3x_{41} - 8{,}8x_5x_6 + 0{,}26x_{14}x_{15}^2 + 0{,}04x_{37}x_{38}x_{39}$$

Experimental verification of the ranking features importance.

To prove the importance of getting in the ranking model features it was performed an experiment for web resource promotion. On each stage of the experiment it was artificial overstatement of certain performance parameters of the model and estimated position on Google search results.

Web-resours: mebilite.kiev.ua
Keywords: kitchens Kyiv, custom-made furniture Kyiv, wardrobes coupe Kyiv.

Web-site positions in search results before optimisation:
kitchens Kyiv - 46
custom-made furniture Kyiv - 51
wardrobes coupe Kyiv -72
Stage 1. Adding keywords in the title and meta tags.

Table 1 illustrates web-site positios for six weeks after the change.

Table 1. Schedule change positions during the six weeks after the change.

Keywords	Optimization length					
	1 week	2 weeks	3 weeks	4 weeks	5 weeks	6 weeks
kitchens Kyiv	35	32	28	26	25	25
custom-made furniture Kyiv	46	40	35	27	23	22
wardrobes coupe Kyiv	57	50	44	36	36	35

Site positions growth almost stopped on the sixth week. To further increase the positions it is necessary to optimize more parameters.

Stage 2. Placement the external links on web-resours.

Table 2 illustrates web-site positios for six weeks after the change.

Table 2. Schedule change positions during the six weeks after the change.

Keywords	Optimization length					
	1 week	2 weeks	3 weeks	4 weeks	5 weeks	6 weeks
kitchens Kyiv	25	20	18	17	15	14
custom-made furniture Kyiv	22	20	20	16	13	13
wardrobes coupe Kyiv	35	30	26	24	21	20

Stage 3. Writing the unique text with keywords for the home page.

Table 3. Schedule change positions during the six weeks after the change.

Keywords	Optimization length					
	1 week	2 weeks	3 weeks	4 weeks	5 weeks	6 weeks
kitchens Kyiv	14	11	8	6	6	6
custom-made furniture Kyiv	13	8	7	7	5	4
wardrobes coupe Kyiv	20	20	16	15	10	9

Conclusion

The experiments of web resources promotion, confirm the accuracy of the Google search engine ranking model obtained by the generalized iterative algorithm GMDH. Experiments proved that the gradual increase of the main features values, based on the ranking model, allows you to achieve high results in web resources promotion of.

References

[8] http://www.google.com - главная страница поисковой системы Google

[9] Zosimov V., Stepashko V., Bulgakova O. Inductive building of search results ranking models to enhance the relevance of the text information retrieval. –Proc. of the 26th Intern. Workshop "Database and Expert Systems Applications, 1-4 Sept., Valencia, Spain / Ed. by Markus Spies at al. – Los Alamitos: IEEE Computer Society, 2015. – 316 p. / – P. 291-295. – ISSN: 1529-4188.

[10] Stepashko V., Bulgakova O., Zosimov V. Performance of Hybrid Multilayered GMDH Algorithm. – Proceedings of the III International Workshop on Inductive Modelling IWIM-2011, 5-9 July 2011, Kyiv-Zhukyn, Ukraine. – Kyiv: IRTC ITS NANU, pp 109-113, 2011.