ITHEA

International Journal INFORMATION THEORIES & APPLICATIONS

2009 Volume 16 Number 1

International Journal
INFORMATION THEORIES & APPLICATIONS
Volume 16 / 2009, Number 1

IJ ITA is official publisher of the scientific papers of the members of the ITHEA® International Scientific Society

IJ ITA welcomes scientific papers connected with any information theory or its application.
IJ ITA rules for preparing the manuscripts are compulsory.
The **rules for the papers** for IJ ITA as well as the **subscription fees** are given on *www.ithea.org* .
The camera-ready copy of the paper should be received by http://ij.ithea.org.
Responsibility for papers published in IJ ITA belongs to authors.
General Sponsor of IJ ITA is the **Consortium FOI Bulgaria** (www.foibg.com).

International Journal “INFORMATION THEORIES & APPLICATIONS” Vol.16, Number 1, 2009

Printed in Bulgaria

Edited by the **Institute of Information Theories and Applications FOI ITHEA®**, Bulgaria,
in collaboration with the V.M.Glushkov Institute of Cybernetics of NAS, Ukraine,
and the Institute of Mathematics and Informatics, BAS, Bulgaria.

Publisher: **ITHEA®**
Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org, e-mail: info@foibg.com

ISSN 1310-0513 (printed) **ISSN 1313-0463 (online)** **ISSN 1313-0498 (CD/DVD)**

PREFACE

The **International Journal on Information Theories and Applications** (IJ ITA) has been established in 1993 as independent scientific printed and electronic media.

IJ ITA major topics of interest include, but are not limited to:

INFORMATION THEORIES

- *Artificial Intelligence*
  - *Computer Intellectualization*
  - *Intelligent Networks and Agents*
  - *Intelligent Technologies*
  - *Knowledge Discovery and Engineering*
  - *Knowledge Acquisition and Formation*
  - *Distributed Artificial Intelligence*
  - *Models of Plausible Reasoning*
  - *AI Planning and Scheduling*
- *Bioinformatics*
- *Business Informatics*
- *Cognitive Science*
- *Decision Making*
- *Education Informatics*
- *General Information Theory*
- *Hyper Technologies*
- *Information Models*
- *Intellectualization of Data Processing*
- *Knowledge-based Society*
- *Logical Inference*
- *Natural language Processing*
- *Neuroinformatics*
- *Philosophy and Methodology of Informatics*
- *Quality of the Programs*
- *Software Engineering*
- *Theory of Computation*

APPLICATIONS

- *Business Information Systems*
- *Communication Systems*
- *Computer Art and Computer Music*
- *Hyper Technologies*
- *Intelligent Information Systems*
- *Multimedia Systems*
- *Programming Technologies*
- *Program Systems with Artificial Intelligence*
- *Pyramidal Information Systems*
- *Very Large Information Spaces*

During the years, IJ ITA became as well-known international journal. Till now, including this volume, ***715*** papers have been published. IJ ITA authors are widespread in ***47*** countries all over the world: *Armenia, Azerbaijan, Belarus, Brazil, Belgium, Bulgaria, Canada, China, Czech Republic, Denmark, Egypt, Estonia, Finland, France, Germany, Greece, Hungary, India, Iran, Ireland, Israel, Italy, Japan, Jordan, Kyrgyz Republic, Latvia, Lithuania, Malaysia, Malta, Mexico, Moldova, Netherlands, Poland, Portugal, Romania, Russia, Scotland, Senegal, Serbia and Montenegro, Sweden, Spain, Sultanate of Oman, Turkey, UK, Ukraine, and USA.*

Volume 16/2009 of the IJ ITA contains ***29*** invited papers.

The great success of IJ ITA belongs to the whole of the ITHEA® International Scientific Society. We express our thanks to all authors, editors and collaborators who had developed and supported the International Journal on Information Theories and Applications.

More information about the IJ ITA rules for preparing and submitting the papers as well as how to take out a subscription to the Journal may be obtained from www.ithea.org .

XV-th anniversary of the International Prize "ITHEA"!

The **International Prize "ITHEA"** for outstanding achievements in the area of information theories and applications has been established in 1995.

For fifteen years, 41 leading scientists of the ITHEA International Scientific Society have been awarded by the "ITHEA" Prize.

Congratulations to Prof. Anatoliy Danilovich Krissilov who has been awarded by the "ITHEA" Prize for 2009 year.

Awarded Scientists till 2009:

1995	*Sandansky*	*K. Bankov, P. Barnev, G. Gargov, V. Gladun, R. Kirkova, S. Lazarov, S. Pironkov, V. Tomov*
1996	*Sofia*	*T. Hinova, K. Ivanova, I. Mitov, D. Shishkov, N. Vashchenko*
1997	*Yalta*	*Z. Rabinovich, V. Sgurev, A. Timofeev, A. Voloshin*
1998	*Sofia*	*V. Jotsov*
1999	*Sofia*	*L. Zainutdinova*
2000	*Varna*	*I. Arefiev, A. Palagin*
2001	*St.Peterburg*	*N. Ivanova, V. Koval*
2002	*Primorsko*	*A. Milani, M. Mintchev*
2003	*Varna*	*T. Gavrilova, A. Eskenazi, V. Lozovskiy, P. Stanchev*
2004	*Varna*	*B. Kokinov, T. Vamos*
2005	*Varna*	*M. Dobreva, L.F. de Mingo, N.Zagoruiko, A.Zakreuskii*
2006	*Varna*	*J. Castellanos, G. Totkov*
2007	*Kiev*	*V. Donchenko*
2008	*Varna*	*Vladimir Lovitskii, Orly Yadid-Pecht*
2009	*Varna*	*A. Krissilov*

Krassimir Markov
IJ ITA Founder and Editor in chief

COGNITION HORIZON AND THE THEORY OF HYPER-RANDOM PHENOMENA

Igor Gorban

Abstract: *In the generalized paper, materials of physic-mathematical theory of hyper-random phenomena oriented to description of statistically unstable physical phenomena are presented. Cognition questions of the world are discussed from position of this theory. Different measurement models are researched. The hypothesis of statistical unpredictability and the hypothesis of hyper-random setting up of the world are analyzed. From these hypotheses follow that cognition horizon is restricted by boundaries of unpredictable changes of researched physical phenomenon and of statistical conditions of their observations.*

Keywords: *hyper-random phenomena, setting up of the world, measurement models, cognition.*

ACM Classification Keywords *G.3 Probability and Statistics*

1. Introduction

How our world set up is, how is a human looking for it, what is cognition process? These are the fundamental questions that have been disturbing the humanity for a long time. Probably, it is impossible to find exhaustive answers to these questions. Here we have been encountering with the problem, well known in philosophy that it is impossible to rigorously prove any natural science theory.

There are in the basis of every natural science theory some statements that although well matched with a lot of experimental data however cannot be rigorously proved. These statements are accepted on faith and are used as absolutely reliable source until they will be revised. In philosophy science these statements are known as fundamental abstract objects [Степин, 1999]. The example of such statements is classic lows of Newton's mechanics seemed unshakeable for a long time. However lack of correspondence between results of a number new experiments and classic physical lows discredited assertion that the world can be sufficiently described by these lows and led to forming new lows that became the basis of quantum mechanics and Einstein's theory of relativity.

Mark, mathematics contains unproved elements – axioms and postulates too.

The study base of the world is statements-hypotheses that well consistent with experimental data. They are accepted on faith because are not proved and play rule of axioms. They lie inherently of any theory.

The main demands for the system of basis hypotheses are the consistency and the mutual independence. For the hypotheses of natural sciences it is requested accordance with experimental data too. Hypotheses of natural sciences may be generalized type, for instance the hypothesis that in the base of the universe is random principles and therefore the world is sufficiently described by stochastic models (this standpoint is a prevalent hypothesis now) or more concrete type one, for instance that the world is described by Newton's lows.

Replacement of real objects, relationships, and operations by definite hypotheses (axioms) essentially reduces cognition process and supplies stable base for study of the world, however it creates insuperable (within the bounds of accepted hypotheses) obstacles for penetration to the essence of physical phenomena. There is paradoxical situation: the cognition is impossible without systems of hypotheses but existence of such systems hold back understanding of basis of the universe.

The multitude of research results of any real phenomenon obtained on the base of different systems of axioms can be regarded as a complex of projections on a number of abstract planes. The more systems of hypotheses and consistent theories created on them, more projections and dipper penetration to the base of the world. Therefore the building of new theories that gives possibility to view on the known facts from new points of view supplies the progress of science.

The science is developed by extension manner, if hypotheses system is fixed. For intensive development new alternative hypotheses and theories are requested. Humanity progress is unthinkable without creation of new theories.

Any mathematical theory is an abstract one. It stays such type before it is not used for solving of practical tasks. Correct using of mathematical theory in different areas of physics, technique, social science, and other ones is possible only when experimental data that conform of adequate description of researched phenomena by corresponding mathematical models are existed. For instance, for correct using of probability theory, experimental data that conform of adequate description of physical events, magnitudes, processes, and fields by random (stochastic) models are requested.

It is not possible to create mathematical models those absolutely identical to real physical phenomena. Even if such models exist it is impossible to prove their adequacy because the accuracy of any measurement is finite.

It is possible to estimate the adequacy of the models to real research objects when there are number experimental data. Different methods can be used for this aim. In any case, it is impossible to obtain the absolutely accurate answer. Therefore, only the *adequacy hypothesis* is accepted for the model if the estimate of adequacy is high.

An adequacy hypothesis is a physical hypothesis that open door for correct using of the mathematical theory in practice. The mathematical theory becomes the physic-mathematical theory thanks to this hypothesis.

For instance, the mathematical probability theory based on the mathematical axiom system only after acceptance of addition *hypothesis that real phenomena may be described adequately by stochastic models* becomes the physic-mathematical theory.

Wide application of any mathematical theory points that there is an adequacy hypothesis and also there is the agreement that visual environment (or its part) is created on the principles of this hypothesis. In particular, occurring everywhere using of the probability theory means that the *hypothesis of random* (*stochastic*) *principles of visual environment* is accepted.

The thesis of stochastic character of visual environment was considered as unquestionable one. However a number facts point that this thesis is a fallacy one.

Recent methods and models of probability theory were developed mainly for statistically uniform (for statistical stable) phenomena samples of which are described by no changed distributions. Just statistically stability of physical phenomena was the bench mark for founders of probability theory.

The hypothesis about adequate description of physical phenomena by stochastic models is well matched with experimental data on small temporal, space, or temporal-space observation intervals. However, it is unjust on large intervals.

One from the most essential argument against the hypothesis of stochastic character of visual environment is *absence of global statistical stability* of real physical events, magnitudes, processes, and fields [Горбань, 2007]. There are not absolutely stable phenomena in the real physical world. Exceptions may be only world physical constants such as light velocity, gravitation constant, and so on. Mark, it is impossible to prove their existence by experiment ways because the accuracy of real measures is limited.

Finite accuracy of any physical measurement is the second weighty argument against the hypothesis of stochastic character of the world [Горбань, 2007]. Classic probability theory and mathematical statistics deal with consistent estimates that lead to true values when the sample volume tends to infinity. Errors would be not limited if physical world is adequately described by stochastic models of consistent type. But it is not so.

The searching of effective means of adequate description of real world has produced numbers of new theories. As a rule, they have interdisciplinary character and are touched with mathematics, informatics, physics, philosophy, and other sciences.

The theories created on the paradigms of probability theory [Колмогоров, 1936], fuzzy-technology [Zadeh, Kacprzyk, 1992, Batyrshin, Kacprzyk, Sheremetov, Zadeh, 2007, Бочарников, 2001], neural network [Hagan, Demuth, and Beale, 1996], dynamic chaos [Cronover, 2000, Дыхне, Снарский, Женировский, 2004, Анищенко, Вадивасова, Окрокверцхов, Стрелкова, 2005, Гринченко, Мацыпура, Снарский, 2005, Sharkovsky, Romanenko, 2005], interval data [Шокин, 1981, Алефельд, Херцбергер, 1987, Кузнецов, 1991, Shary, 2002, Kreinovich, Berleant, Ferson, Lodwick, 2005, Ferson, Kreinovichy, Ginzburg, Myers, 2003], and others are related to them. New physic-mathematical theory of the theory of hyper-random phenomena [Горбань, 2007] is in this list too.

The aim of the paper is the discussion of cognition boundaries of the reality in the aspect of the hypothesis of hyper-random setting up of the world advancing in the theory of hyper-random phenomena.

The paper consists of nine sections. Brief description of mathematical and physical bases of the theory of hyper-random phenomena is presented in Section 2. Unformalized, physical, and mathematical models and also processes of knowledge, reasoning, and cognition are submitted in Section 3. Measurement questions of physical quantities are researched in Section 4. The mathematical basis of measurement, the problem of forming of adequate estimates and models are considered hear. Modern approaches used for estimation of measurement accuracy are discussed in Section 5. Different measurement models are described in Section 6. The hypothesis of statistical unpredictability and the hypothesis of hyper-random setting up of the world, that are the base of the theory of hyper-random phenomena are discussed in Section 7. These results give possibility to understand why accuracy of any physical measurements and cognition of the world are limited. The results demonstrating the nonlinear character dependence of measurement accuracy from the volume of data smoothing are presented in Section 8. Last Section 9 is devoted to conclusion.

2. Mathematical and Physical Bases of the Theory of Hyper-random Phenomena

The theory of hyper-random phenomena has two components – mathematical and physical ones. Mathematical part is an advancement of classic probability theory and mathematical statistics. Physical part is based on the hypothesis of statistical unpredictability and the hypothesis of hyper-random setting up of the world. According to the first one, there are unpredictable phenomena in the world that not depend from any events occurred before. According to the second one, it is possible to describe sufficiently the real physical world by hyper-random models.

The basis hypotheses using in the theory of hyper-random phenomena include Kolmogorov axiom system of probability theory and the hypothesis of hyper-random setting up of the world.

The concept of randomness and random phenomenon (event, quantity, process, or field) are understood by different researcher in different ways. Mark, all initial conceptions, used in science, including randomness concept, are accepted by agreement. It was written a lot about this [Тутубалин, 1972, Тутубалин, 1993, Алимов, Кравцов, 1981, Кравцов, 1989].

In the probability theory, Kolmogorov set-theoretic definition of the random event [Колмогоров, 1936] is used usually. Strictly mathematically, a random event is described by the probability space $(\Omega, \Im, P)$, where Ω is a sample space with elements $\omega \in \Omega$, $\Im$ is a sigma algebra of events, P is a probability measure defined on the sigma algebra $\Im$. Random event is defined in such manner even in the International Standard [International standard, 2006].

According to less strict but more illustrative statistical definition (by P. von Mises [Mises, 1964, Гнеденко, 1961]), the probability $P(A)$ of a random event A is a limit of the frequency $p_N(A)$ of its occurrence during experiments in equal conditions when a number of the experiments N tends to infinity: $P(A) = \lim_{N \to \infty} p_N(A)$.

With low values of N frequency $p_N(A)$ can vary, but with increasing of N it gradually stabilizes and with $N \to \infty$ it tends to a definite quantity $P(A)$.

It is known algorithmic definition of the randomness proposed by Kolmogorov in the middle of 60th years of past century. It basis on the analyzing of algorithmic complexity of the program, that translates a known sequence to the researched one [Колмогоров, 1987].

It is known other approaches too. On practice, as a criterion of the randomness it is used often by physicists a lot of different empirical, semi-empirical, semi-formalized, and even non-formalized criterions such as a fall down correlation, a continues spectrum, a irreproducibility, a non-repeatable observation, a non-controllability, a unpredictability, and so on [Кравцов, 1989].

In current paper, Kolmogorov set-theoretic definition of the random event is used. Random quantity is a deterministic numerical function defined on a sample space Ω of random events $\omega \in \Omega$. Random function is a deterministic numerical function of independent argument, value of which under fix value of argument is a random quantity.

When one says about a hyper-random phenomenon (as a mathematical object) a set of the condition random phenomena (events, quantities, or functions) depending from the condition $g \in G$ is implied.

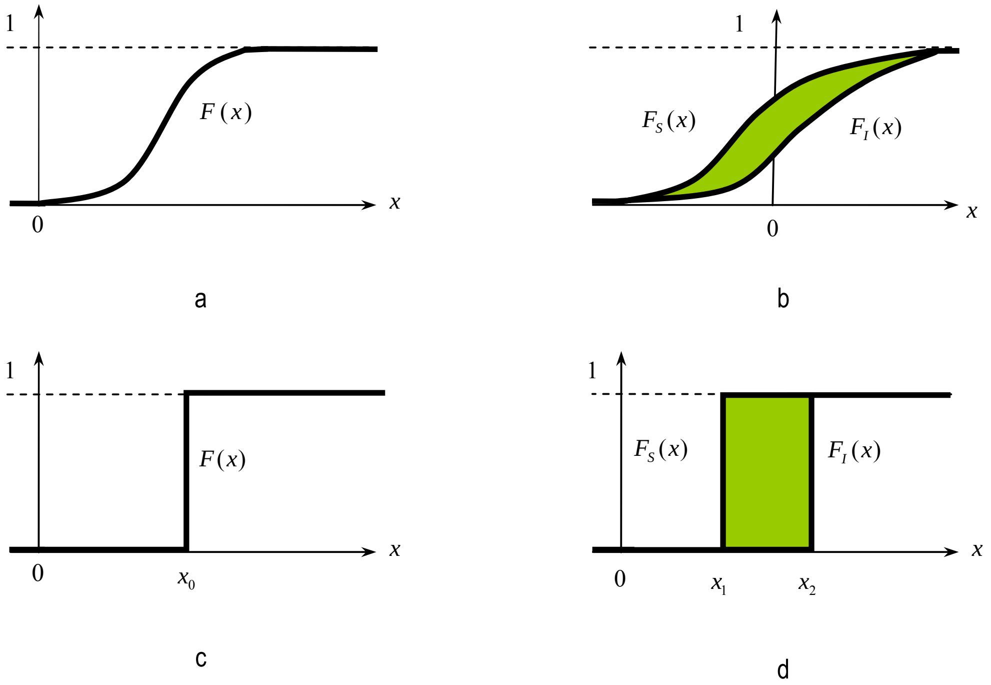
Hyper-random phenomena can be described by means of tetrad $(\Omega, \Im, G, P_g)$ [Горбань, 2005, Gorban, 2006, Горбань, 2007] where Ω is a space of elementary events $\omega \in \Omega$, $\Im$ is a sigma algebra of events, G is a set of conditions $g \in G$, P_g is a probabilistic measure of the subset of events depended on a condition g. Thus, the probabilistic measure is defined for all subsets of the events and all possible conditions $g \in G$ while the measure for conditions $g \in G$ remains undefined.

By using less strict approach, conditions g cab be regarded as statistical conditions that associated with definite distributions and a hyper-random event A can be treated as the event which frequency of occurrence $p_N(A)$ does not stabilize with increasing in the number of experiments N and has no limit, when $N \to \infty$.

As probability distribution fully characterizes random phenomenon, as set of condition probability distributions fully characterizes hyper-random phenomenon. For instance, a random quantity X is fully described by any distribution function $F(x)$ (Fig. 1, *a*) and a hyper-random one $X = \{X / g \in G\}$ – by set of conditional distribution functions $F(x/g)$, where g is an element of the set of statistical conditions G.

Expectation, variance, and other moments less fully characterize random quantity. Mark, if one or some moments of random quantity are known (or given), it is implied that there is definite (may be unknown but single) distribution low that fully describes examined random quantity.

Hyper-random quantity is described less fully by upper $F_S(x)$ and lower $F_I(x)$ boundaries of the distribution function (Fig. 1, *b*), by their central and crude moments (expectations of boundaries m_{Sx}, m_{Ix}, variances of boundaries D_{Sx}, D_{Ix}, and so on), by boundaries of moments (expectation boundaries $m_{ix} = \inf_{g \in G} m_{x/g}$, $m_{sx} = \sup_{g \in G} m_{x/g}$, variance boundaries $D_{ix} = \inf_{g \in G} D_{x/g}$, $D_{sx} = \sup_{g \in G} D_{x/g}$, where $m_{x/g}$ and $D_{x/g}$ are accordingly expectation and variance of condition random quantity X/g) and others.



a

b

c

d

Fig. 1. Presentation of different type quantities by distribution functions.

When any hyper-random quantity is researched, it is implied that there is definite (may be unknown but single) set of condition probability distributions that fully describes of it.

Mark, that a random quantity is a particular case of a hyper-random quantity in which boundaries of the distribution function are coincided: $F_S(x) = F_I(x) = F(x)$.

Determinate quantity x_0 can be regarded in rough as a particular case of a random (or hyper-random) quantity that has in the distribution function a unit step in the point x_0 (Fig. 1, *c*).

Interval quantity [Шокин, 1981, Алефельд, Херцбергер, 1987] characterized by interval boundaries x_1, x_2 can be offered as infinite number of hyper-random quantities with unit step boundaries of the distribution function $F_S(x)$, $F_I(x)$ in the points x_1, x_2 (Fig. 1, *d*).

A special case of the interval quantity $[x_1, x_2]$, when x_1 tends to minus infinity and x_2 – to infinity, is fully ambiguity quantity that can be considered as "chaotic" one not given in to any description.

So, a hyper-random quantity is a generalized concept of determinate and random quantities and all set of hyper-random quantities with definite distribution boundaries is a generalized concept of interval and ambiguity quantities.

All pointed quantities can be interpreted as same type objects with different determinism level. Determinate quantity has the highest determinism level. Determinism in the random quantity is on the level of statistical distribution low, in the hyper-random quantity is on the level of concrete conditional distribution functions, and in the interval quantity is on the level of determinate boundaries of interval. Ambiguity quantity has not any determinism elements.

First papers devoted to hyper-random theory were published in 2005 [Горбань, 2005]. For passed years the theory was developed in many directions [Горбань, 2006, Gorban, 2006, Горбань, 2007, Горбань, 2008, Gorban, 2008]. The classic principles of probability theory and mathematical statistics were used [Колмогоров, 1936, Гнеденко, 1961, Тутубалин, 1972, Королюк, 1985, Боровков, 1986] on the basis of it.

The theory of hyper-random phenomena covers hyper-random events, hyper-random quantities, and hyper-random functions. Results firstly obtained for one dimension case were generalized on multidimensional case too. Number new definitions and conceptions were inputted in particular convergence, continuity, differentiability, and integrability of hyper-random quantities and functions and also stationary, ergodic, and Markov hyper-random processes. Particularities of different hyper-random quantities and functions were researched. Conception of hyper-random sample was inputted, properties of estimates of hyper-random quantities and processes were researched, in particular their convergence.

It was devoted [Горбань, 2007] that in general hyper-random estimates are not consistent ones, i.e. their errors do not follow to zero when the volume of the sample tends to infinity. This circumstance is very important. Hyper-random estimates substantially differ from analogues random ones by this particularity.

To obtain image about physical aspect of the theory let us examine a classic stochastic example, from which the learning of probability theory begins – the “toss-up” coin game. It researched in detail by a lot of mathematics. Pearson made $24,000$ tossing of the coin [Тутубалин, 1972]. The heads was obtained $12,012$ times. The same type experiments were led by Bernoulli, Laplace, and other mathematicians. The task offered not trivial for them.

It is assumed generally that results of the experiments are random and have specific probabilities: the probability of heads is $P_h = 0.5$, and the probability of tails is $P_t = 0.5$.

Is the model described correct? It seems, ex facte, there are no reasons to doubt its adequacy. Even Pearson’s experiment proved this.

However, it is true at first sight. Is it not possible the probabilities P_h, P_t be different? It is easy to show that with some training and having the controlled initial position of the coin one can learn to toss it so the dropout rate of a side will be within the range of some fixed value larger than 0.5, and the dropout frequency of the other one – within the range of the value less than 0.5. Indices can change in either side when conditions of tossing vary.

In this case, the results of the experiments can be treated as a hyper-random event. Thus, hyper-random model taking into account possible changes of probabilities for heads and tails describes real situation more adequately than the random one which assumes fixed values of these probabilities.

3. Cognition of the World

The environment is a complicated system containing numberless interconnected elements – objects. Every object is characterized by a lot of properties determining its peculiarities. Some properties of different objects can be coincided or be closed on definite criterion. Objects with identical or near properties are grouped in our mind in classes of similar objects.

Every object can belong to several classes. Classes can be included in other classes. Properties that proper to similar objects are regarded as regularities or laws. Regularities and laws are objects too. As other objects they can form classes. Regularities and laws define specific peculiarities of connections of objects of the class with other objects of the same class and with objects of others classes.

Classes, for instance, are the multitude of physical objects that obey to lows of classic mechanics, all known lows of classic mechanics, objects sizes of which are less or larger to definite magnitude, plants or animals of definite type, people of definite nationality, confession, race or age, citizen of the country, and so on.

Systematization (forming the system of interconnected object classes) and classification (distribution of objects to classes) are the main actions in cognition of the world. They include detection new objects, examine and description of their properties, replenishment of existed classes by new objects, creation of new classes, and removal of outdated classes from the system.

As a result of systematization and classification, unformalized models that give integral presentation about similar classes are formed in the mind.

Unformalized models are not identical to examined objects. They carry information about multitude of similar objects included to class and are theirs averaged image data.

Everybody tries to attribute every new object, as a rule subconsciously, to different unformalized models. The models to that the object mostly accords are associated with the object and contrary the object is associated with these models. As a result, a multitude unformalized models are formed. This multitude is interpreted by person as integrated image of the object. A set of images according to different objects creates subjective image of the person about surrounded world.

The world is continuously changed; objects and their models are changed too. Changing of the models is connected not only by changing of the objects but revision of classification criteria. When a criterion is chosen, essential role plays acquired experience, environment, and many others factors.

Persons are differed each other and conditions of their life are differed too. Therefore unformalized models of the objects, images, and world conceptions are different for different persons.

Real objects are perceived by a human by sense organs as estimates that can be regarded as objects too.

Estimate depends from a lot of subjective and objective factors (conditions). The estimate is differed from the real object as by imperfection of our sense organs and observation means as different noise effects. When conditions are changed (for instance, sight become weaker, more or less perfect observation means are used, noise characteristics are changed) the estimate is changed too.

A model and an estimate are different conceptions. If the model of the object is the generalized image about the multitude of similar objects, the estimate is more or less distorted image of the single object.

On the base of estimates the average estimate and estimate model are formed.

A multitude of different estimates corresponding to single object or some objects from definite object class creates in human mind the unformalized model of the estimate. This model as an object model is continuously changed. Changes are called by changing of objects, estimates, and classification criteria of objects and estimates. Unformalized models of estimates are the basis for forming of unformalized models of objects.

Unformalized models of estimates and unformalized models of objects are different category however as a rule they are identified in human mind.

Unformalized models of objects and estimates can be formalized by physical models, taking into account the most essential their peculiarities, and be described by mathematical models.

A set of physical models can be built for every multitude of real objects and every multitude of estimates. Every physical model can be described by the multitude of different mathematical models.

Part objects that belong to the examined object's class cannot be described by physical model on the phase of formalization. Together, objects not belonged to the examined class can be included in the physical model. As a rule, when mathematical model is built, all objects of the examined class that belong to the physical model are described by mathematical model.

Similar objects of any class, their estimates, examined object x, its estimate x^*, models of the object (unformalized N_x, physical P_x, and mathematical M_x) and also the models of estimates (unformalized N_{x^*}, physical P_{x^*}, and mathematical M_{x^*}) are described schematically in Fig. 2.

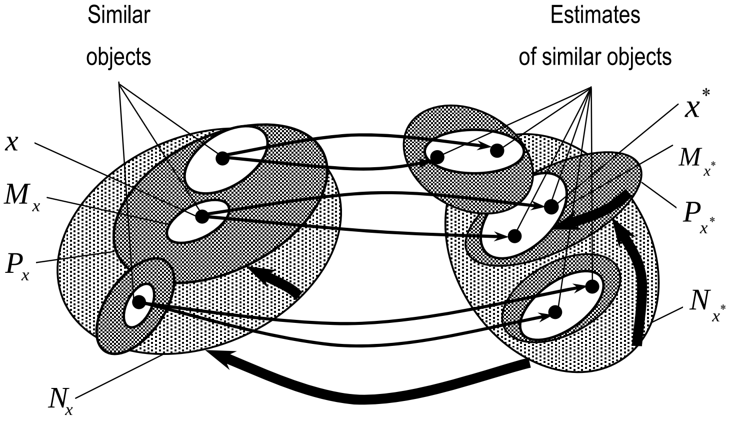


Fig. 2. A scheme of the models and the estimates.

Knowledge is a multitude data in any area. A lot of knowledge about the world has been accumulated over millenniums of civilization development. Inflow tempo of new information increases continuously. It becomes more and more difficult to orientate in information flow. Well-tested mechanisms of data systematization, classification, and generalization help to solve the problem.

Human knowledge is not a number of odd data. It is a system of systematized, classified, and generalized data that is a system of models.

Human possibilities in perception and processing of information are limited. The cognition mechanism based on the model system defends human organism from information overload.

Human knowledge is formed on the basis of the knowledge of separate persons by their systematization, classification, and generalization.

Person knowledge and human knowledge are consisted from formalized and nonformalized models. As a rule, models in exact science are formalized and in humanities are nonformalized.

Person image about the world, his world-view is a set of models forming knowledge and person's estimates of these models on a multitude criteria for instance risk, reliability, significance, novelty, accordance to definite regulations, and so on.

Mark that estimates of models and models of estimates are absolutely different categories. If criteria are depended, estimates on different criteria are linked. In spite of person's individuality, knowledge and estimates of different humans in definite object area may be similar. Persons with the same views, joint interests, the same religion, belonged to the same ethnos, and obtained the same education are humans with similar models and analogues person's estimates of models on the multitude of different criteria. Subjective estimates of models are basis for collective estimates.

Models and estimates, as subjective as collective ones, are relatively stable objects that are slowly changed in time under the influence of external factors. Therefore the world-view may be regarded as the system of stable models with stable estimates.

Models can have different level of generalization. On the basis of low level models, creation of more high level derivative models is possible.

Models may be divided to two classes: to models of structural elements and to links among them. The last are the models of real lows of the nature.

Synthesis of new models of structural elements, finding new links, forming new estimates, and their revision are the essence of the thought.

Curiously enough in the first view, creation of derivative models of structural elements, finding of new links, forming of new estimates, and their revision are possible owing to stability of models of structural elements, links among them, and their estimates.

Thought process depends from the degree of stability of initial knowledge: the more stability the more levels of derivative models of structural elements, links, and estimates may be formed.

As show a lot of biology researches, a human substantially inferiors to many animals in possibilities of perception of information, processing of information, random-access storage (i.e. in possibility to form initial models), however owing to definite "conservatism" (thought inertness), he considerably exceeds animals in finding of links among models and in synthesizing of derivative models.

The models of structural elements, the links, and the estimates are changed under the influence of different factors with the lapse of time. New models, links, and criteria of estimation are created, specifications of elements, exclusion of outdated and unclaimed ones, changing of them to new elements are occurred, and also forming and revision of estimates are taken place.

Cognition is a process of knowledge acquisition and comprehension of the lows of the objective world. Cognition by a person and by the humanity may be divided on two phases. The first phase is forming of initial knowledge (initial set of the models) and estimates. The second one is actualization (revision, specification) of existed knowledge and estimates.

In the first phase, nonformalized models play a special role; in subsequent phase, as nonformalized as formalized ones are important.

Both phases are teaching. Person's teaching may be without or with outside help. Ways for knowledge transmission are different: by communication of the persons or by indirect ways with using auxiliary means such as books, computers, video techniques, and so on. Described model of forming knowledge is presented schematically in Fig. 3.

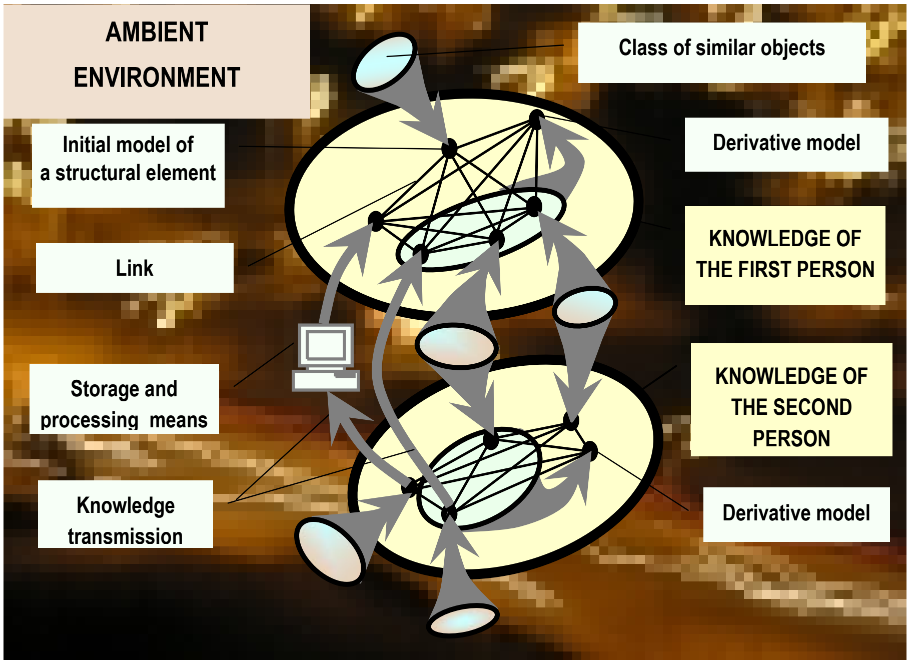

AMBIENT ENVIRONMENT
Class of similar objects
Initial model of a structural element
Derivative model
KNOWLEDGE OF THE FIRST PERSON
Link
Storage and processing means
KNOWLEDGE OF THE SECOND PERSON
Knowledge transmission
Derivative model

Fig. 3. A scheme of the process of knowledge forming.

Cognition is a complex compound process demanding the forming of the estimates, creating of the models, comparison, confrontation, systematization, classification, detection, and many other operations. In the basis of them is measurement. Cognition facilities are defined by measurement precision. Let us stop on this question at grate length.

4. Measurement Principles

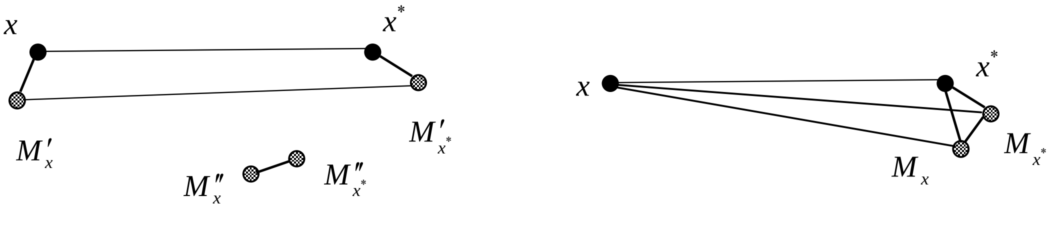
All nonformalized, physical, and mathematical models of objects and estimates are objects. To characterize numerically the closeness of the objects the metric space is demanded. Such space is a multitude S for any elements of witch x, y there is a real function $\mu(x, y)$ that is the metrics or distance that satisfy to following postulates: $\mu(x, y)=0$ if and only if $x = y$; $\mu(x, y) \le \mu(z, x)+\mu(z, y)$ (inequality of triangle), where x, y, z are any elements of the multitude S. The metrics is a nonnegative quantity and $\mu(x, y) = \mu(y, x)$.

It is known that there are some multitudes for which it is impossible to create a metric space. For instance, it is impossible to build any metric space for the multitude of real functions defined in the finite range. Therefore it is impossible to create a metric space that includes the model of the real object and the models of all its possible estimates. If to restrict the class of considered functions, for instance, to continuous real ones, it is possible to create metric space. In this case, it is possible to define the distance for all mathematical objects described by such functions.

Different metrics generates on a definite multitude S the different metric spaces. The variation distance row according to elements $x, y, z, \ldots$ depends from the metrics. So in two different metric spaces created on the multitude including the model of real object and the models of its estimates, the nearest to the model of real object may be different models of estimates.

It may be case when for a whole class of metric spaces given on the definite multitude the same estimate model is a nearest one for the model of real object. Reference quantity (conventional unit) is settled by the metrics. A distance is compared with this unit. Mark that Euclidean metrics usually is used in the metrology.

If the model of the object M'_x is near to the object x, the model of the estimate M'_{x^*} is near to the estimate x^* ($\mu(x, M'_x) \sim 0$, $\mu(x^*, M'_{x^*}) \sim 0$), or the model of the estimate M''_{x^*} is near to the model of the object M''_x ($\mu(M''_x, M''_{x^*}) \sim 0$) it is not ensure that the estimate x^* and its models M'_{x^*}, M''_{x^*} are near to the object x (Fig. 4, *a*). The estimate x^*, the model of the estimate M_{x^*}, and the model of the object M_x may be near to each other however far from the object x (Fig. 4, *b*).



a b

Fig.4. Mutual disposition of the object, the estimate, and their models.

Often on the practice, for confirmation of the theory the closeness of theoretical results to experimental data are inspected. Mark, that low difference of the results often regarded as irrefutable reason that the theory is correct, in real is not such one. The experimental result is the estimate and the theoretical result is the model of another estimate. The closeness of these results indicates that the results close to each other but not points that they close to the researched object. The adequacy question of the theoretical model remains open before the closeness of the experimental result to the researched object is not proved.

True (undistorted) information about real object x is inaccessible. It is impossible to know how close are the estimate x^*, its model M_{x^*}, and the object's model M_x to the object x.

Therefore the building task of adequate estimates and models has not exact solution. In future, under adequate estimates and models it will be implied ones that are in the round of the corresponding examined objects.

Coincidence of the estimate, the estimate model, or the object model with the real object practically is impossible. There are a lot of reasons for this. The most essential ones are:

- taking into consideration under forming of the physical models not all factors defining the state of the real object and its estimate;
- forming of the object's model on the base of the estimate's model;
- unmatched parameter changes of the real object and its estimate;
- “delaying” of physical models from the current state of the object and the estimate;
- “delaying” of mathematical models from physical models;
- statistical unpredictability of the real object and its estimate;
- influence of different noises that leads to distortion of the estimate, the estimate model, and the object model;

- statistical unpredictability of noises;
- imperfection of physical and mathematical models, and others.

The distance between the estimate and the object (physical quantity, process, or field) is the measurement error. The error is caused by inadequate perception of the object and by noises. Inadequate perception may be called as by objective as subjective reasons. Noises are connected as a rule with objective factors.

Different character of reasons that calls difference of the estimate and the real object requires different approaches for taking into consideration and compensation of error components. As a rule, the same strategy is used: data accumulation and their averaging.

Compensation of inadequate perception of the object called by objective reasons is based on forming of a number estimates obtained by different ways, compensation of inadequate perception called by subjective reasons is based on multiple estimation of the object by different persons, and compensation of noise influence in statistically stable conditions – on repeated estimates obtained in these statistical conditions.

Results of multiple estimation of the object by different ways and persons are used for obtaining of the average estimate. This estimate although as a rule is nearer to the real object than the most part of initial estimates but not coincide with the real object. This is so not only because the data volume is limited and the average mean is not optimal. The main reason is that it is impossible to watch changes of characteristics of the object and noises. As a result, measurement precision always is limited.

Mark, measurement accuracy is a qualitative category quantitatively characterized by error or uncertainty of measurement.

5. Bias Conception and Uncertainty Conception

In the metrology two approaches are used to characterize measurement precision. One of them is based on bias conception and another one (that has reputation of more progressive) – on uncertainty conception.

In bias conception, systematic and random errors are regarded. Systematic error is that one which remains fix or changes on definite low when multiple measurements are led. Random error is that one which changes by a random manner when multiple measurements are led.

Random error usually explains by time or space random changes of quantities, that influent on the result of measurement, and systematic error – by deviation of parameters or measurement conditions from ideal ones.

Random error may be reduced by statistic processing of number results of measurements and systematic error – by taking into account of known dependences of measurement results from parameters influencing on the results.

If the systemic error not changes from measure to measure (this fact is accepted, as a rule by default) the systematic error coincides with expectation of the cumulative error. In this case, expectation of the random error equals to zero.

The error of the estimate Θ^* of the measured quantity θ are usually characterized or by the systematic error ε_0 (expectation of the error) and the standard deviation σ_{θ^*} of the estimate Θ^* or by the confidence interval $I_\gamma(p) = [\Theta^* - \varepsilon_0 - \varepsilon, \Theta^* - \varepsilon_0 + \varepsilon\ ]$ according to the definite confidence probability $\gamma = P(\left|\theta^* - \varepsilon_0 - \theta\right| \le \varepsilon)$ that the absolute deviation of the random quantity $\Theta^* - \varepsilon_0$ from the measured quantity θ is not more than the definite volume ε.

In some cases, the systematic error can be compensated partly by special measurement methods that give possibility to reduce it without error detection. It is known many such methods, in particular the substitution method, the error compensation on sign method, opposition method, symmetric surveys one and others.

The random error can be reduced by multiple estimation of the quantity and averaging of the obtained data.

In uncertainty conception, two types evaluations of uncertainty (A and B) are regarded. Type A evaluation of uncertainty is a method of evaluation by the statistical analysis of series of observations and type B evaluation of uncertainty is a method of evaluation by means of other manner [Guide, 1993].

Uncertainty of the measured quantity θ is characterized by type A standard uncertainty $u_{\mathrm{A}\theta}$, by type B standard uncertainty $u_{\mathrm{B}\theta}$, combined standard uncertainty $u_{\theta} = \sqrt{u_{\mathrm{A}\theta}^2 + u_{\mathrm{B}\theta}^2}$, and expanded uncertainty $U_{\theta} = ku_{\theta}$ (were k is a coverage factor) that in case of absence of $u_{\mathrm{B}\theta}$ component is interpreted as uncertainty according to confidence probability γ [Guide, 1993].

Dividing of errors on random and systematic is caused by nature of their formation and manifestation in the process of measurement, and dividing of uncertainty on A and B types – by evaluation methods.

Underline, although these approaches have essential differences, both concepts are based on the proposition that errors can be called only by deterministic and random factors. Other types of factors, in particular of hyper-random type are not taken into account. Seemingly this orthodox position requests the revision.

Let us consider the example related to the metrology – the precise measurement of diameter of the cylinder of circular section [Горбань, 2007]. The completely trivial problem appears quite complicated under in-depth analysis. To make a detail of absolutely circular section is impossible. Its section will always differ from ideal circle: firstly, because of ellipsoidal or any other deviation from ideal circular form, and, secondly, due to roughness of the surface. It should be also born in mind that different sections by the cylinder axis are differed. Therefore, the true size of the detail, even without considering temperature and a number of other factors, which can be ignored further in order to simplify the speculations, is different in different measurements.

In this case, because of complex shape of the detail section the concept of the diameter is not acceptable. Considering this, the problem should be defined as the measurement task of average size of sections.

The physical model of the measurable value should to consider the deviation from the ideal circular shape, roughness, and difference in sections along the axis. For mathematical description of physical model, in principle, both common random and hyper-random mathematical models can be used.

The random model is based on an assumption that probabilistic characteristics of the measurement results are constant. In reality, this is not true. Within small local areas they can be approximately constant, but in general they can considerably depend on a direction along which the measurement is taken, and on the section under consideration. Therefore, hyper-random model, considering the variability of distribution functions, better describes the measurable value than the random model.

Any measurements are carried out under impacts of various obstructive factors (obstacles). In the examined problem, such is the clogging of surface of the detail. Dust and dirt on the surface accumulates unevenly. Within small local areas clogging has random type, but in whole, because of difference of distribution laws for different areas, it is of a hyper-random type.

There are no absolutely precise measurement instruments in the world. Neither a trammel, nor a micrometer or any other measurement instrument can measure the section dimensions with infinitely high precision. Reasons are different: roughness of the detail surface, various instrumental errors etc. The unifying feature of them is that they are of hyper-random type. Therefore, for mathematical modeling of a measurement instrument the hyper-random model is preferable.

Hence, it follows that a generalized physical model, considering in whole real features of measurable value, obstacles and measurement instruments, is described more adequately by a hyper-random mathematical model than the random one.

Mark, the task of measurement of physical quantity in conditions of statistical instability is not knew [Тутубалин, 1972]. A lot of mathematicians and physicians discussed the problem. In the middle of former age number (Kolmogorov, Pearson's chi-square, omega-square, and others) testing nonparametric statistic hypotheses methods that give possibility to evaluate statistical homogeneity of the sample were developed.

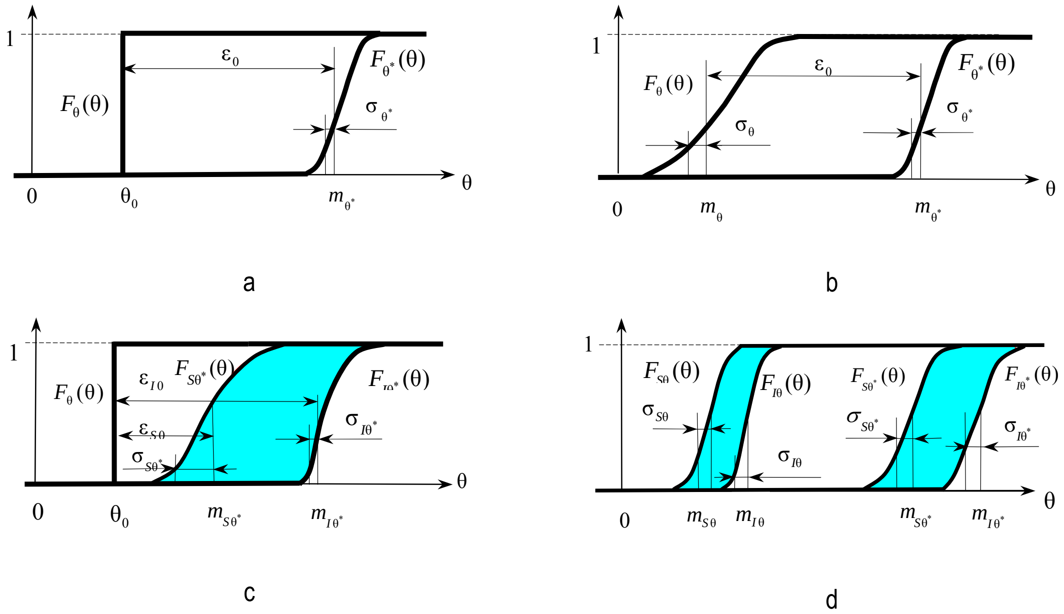
Examine now the classic measurement model used in metrology and other ones [Горбань, 2006, Gorban, 2006, Горбань, 2007, Gorban, 2008].

6. Measurement Models

In the traditional classic notion it is supposed that the measured physical quantity θ is not changed for measurement time and the result of measurement (estimate Θ^*) is changed from measure to measure on a random low, so there is a statistical steadiness of the estimate. Taking into account that the measured quantity is considered as a deterministic one and the result of the measurement is a random one (Fig. 5, *a*), this measurement model may be called as a deterministic – random type.

A measured quantity can be a random type too (Fig. 5, *b*). According measurement model may be called as a random – random type.

When a measured quantity has a deterministic character and its estimate has a hyper-random character (Fig. 5, *c*) we obtain a deterministic – hyper-random model of the measurement. It may be other types of the models. A most usual model is a hyper-random – hyper-random one. In this model a measured quantity and its estimate are variables of hyper-random types (Fig. 5, *d*).



a

b

c

d

Fig. 5. Different measurement models: deterministic – random (a), random – random (b), deterministic – hyper-random (c), and hyper-random – hyper-random (d) models.

Parameters and characteristics describing measurement accuracy are different for different measurement models. For a deterministic – random model (Fig. 5, *a*) such ones are the distribution function $F_{\theta^*}(\theta)$ of the estimate, the expectation m_{θ^*} of the estimate, its bias $\varepsilon_0 = m_{\theta^*} - \theta_0$, the standard deviation σ_{θ^*} of the estimate, and other ones.

For a random – random model (Fig. 5, *b*) they are the distribution functions $F_\theta(\theta)$, $F_{\theta^*}(\theta)$ of the measured quantity and the estimate accordingly, the expectations m_θ, m_{θ^*} of the measured quantity and the estimate correspondently, the bias $\varepsilon_0 = m_{\theta^*} - m_\theta$ of the estimate, the standard deviations σ_θ, σ_{θ^*} of the measured quantity and the estimate, and others.

For a deterministic – hyper-random model (Fig. 5, *c*) they are the upper $F_{S\theta^*}(\theta)$ and lower $F_I(x)$ boundaries of the distribution function of the estimate Θ^*, the expectations $m_{S\theta^*}$, $m_{I\theta^*}$ of the boundaries of the estimate, the biases $\varepsilon_{S0} = m_{S\theta^*} - \theta_0$, $\varepsilon_{I0} = m_{I\theta^*} - \theta_0$ of the boundaries of the estimate, the standard deviations $\sigma_{S\theta^*}$, $\sigma_{I\theta^*}$ of the boundaries of the estimate, and others.

For a hyper-random – hyper-random model (Fig. 5, *d*) they are the upper $F_{S\theta}(\theta)$ and lower $F_{I\theta}(\theta)$ boundaries of the distribution function of the measured quantity, the upper $F_{S\theta^*}(\theta)$ and lower $F_{I\theta^*}(x)$ boundaries of the distribution function of the estimate Θ^*, the expectations $m_{S\theta}$, $m_{I\theta}$ of the boundaries of the measured quantity, the expectations $m_{S\theta^*}$, $m_{I\theta^*}$ of the boundaries of the estimate, the standard deviations $\sigma_{S\theta}$, $\sigma_{I\theta}$ of the boundaries of the measured quantity, the standard deviations $\sigma_{S\theta^*}$, $\sigma_{I\theta^*}$ of the boundaries of the estimate, and others.

7. Hypothesis of Statistical Unpredictability and Hyper-random Hypothesis of World Organization

Estimates according to different measurement models have different properties. A estimate of expectation of a random quantity obtained by calculating of sample mean is consistent estimate. The variance of this estimate is reduced proportionally to sample volume. Therefore in case of a deterministic – random model, a random measurement error of any constant magnitude can be theoretically reduced to zero when the sample volume tends to infinity.

On practice, possibilities of reducing of the error are limited by the longitudes of intervals where the measured magnitude practically is not changed and statistical conditions of its observation are remained invariable.

Since these longitudes of intervals are always finite quantities the real estimates are not consistent ones. Therefore it is impossible to obtain the unlimited accuracy, even under unlimited data volume. The accuracy is defined by the longitudes of space-time intervals of stability of the researched object and statistical conditions of its observation. With approaching to the highest measurement accuracy the effectiveness of averaging reduces. This shows up in slowdown of the reducing rate of the measurement variance, when the sample volume rises.

Maximum number of the running data of the sample for that the deterministic – random measurement model stays the adequate one can be estimated by known estimation methods for sample homogeneity [Королюк, 1985, Леман, 1979, Горбань, 2003].

When there is a large sample volume the inadequate description of real estimates by random models leads to conclusions not conforming to experimental data. This becomes especially clear when high accuracy measures are led.

In every concrete case, physicists give different explanations why it is impossible to achieve of the extremely high accuracy (for instance, by Brownian motion of molecules). These explanations as a rule are correct. However, great number of them and their variety look from the side as subconscious desire to save classic random model of world organization.

It seems more natural *to accept that in the real world there are statistically unpredictable phenomena* limiting the accuracy of measurement and the possibility of cognition.

Statistical unpredictable hypothesis may be regarded as the physical model of one of fundamental lows of the nature. Its mathematical model is the *hyper-random hypothesis of world organization* assumed that *any physical phenomena can be described adequately by hyper-random models.*

According to hyper-random hypothesis of world organization *all real physical quantities, processes, and fields* (may be excepting only small number of world physical constants) *have hyper-random character* [Горбань, 2006, Горбань, 2007, Gorban, 2008].

This hypothesis concerns to different real phenomena including quantities, functions, and fields that are measured, current noises, and measurement errors. Every estimate is forming as an estimation result of the mixture of the measured quantity, process, or field and the noises prevented for observation.

If noises are of hyper-random type, even in case, when the measured quantity is a constant (world constant) the result of measurement is of hyper-random type.

Hyper-random character of the estimate is shown up firstly in unpredictable drift of the shift. Since the drift is changed in unpredictable manner it is impossible to compensate it.

So it is clear why all estimates of real quantities, processes, and fields are not consistent and potential accuracy of any measurements is limited. The boundary of accuracy defends not only the number of measurements and their random dispersion but mainly unsteady character of probability characteristics of measured quantity and noises.

In fact, hyper-random hypothesis recognizes existence of unpredictable physical phenomena that defines the strategy of world development.

Philosophy idea, based on the existence of unpredictable phenomena, suggested and discussed by many philosophers. But it was not developed to the level of formalized mathematical models that give possibility to obtain strict logical conclusions. The novelty and specific of hyper-random theory consist mainly in that it formalizes this idea and proposes mathematical apparatus for solving different practical tasks.

Hyper-random estimates are not consistent ones while parameters used to describe these estimates are of consistent type. When a sample volume tends to infinity the value of hyper-random estimate Θ^* unpredictably changes in the range $[m_{S\Theta^*}, m_{I\Theta^*}]$. In many cases, hyper-random estimates due to their unpredictability are more prefer for description of real objects than random ones.

To illustrate specific particularities of developing approaches look at concrete task.

8. Example

Let in uncertainty statistical conditions the measured quantity, the estimate, and the sample are adequately described correspondently by hyper-random quantities Θ, Θ^*, $\vec{X}$. The random sample according to statistical condition $g \in G$ is presented by additive mixture of random quantity Θ / g and random homogenous noise described by a vector $\vec{V} / g$ the components of which are independent and have expectations $m_{v/g}$ and variance $\sigma^2_{v/g}$. The noise does not depend from measured quantity and its variance is in the range of [σ^2_{iv} , σ^2_{sv}]. Statistical condition changes so slowly that sample forming conditions may be accepted as practically invariable ones.

It is necessary to estimate the measured quantity θ/g in uncertainty condition $g \in G$ and the measurement accuracy.

If we have N running samples $x_n / \theta, g$ it is possible to form for uncertainty condition $g \in G$ the estimate $\theta^*/\theta, g = \frac{1}{N}\sum_{n=1}^{N} x_n/\theta, g$.

The middle error squared is $\Delta^2_{z/g} = m^2_{v/g} + \frac{\sigma^2_{v/g}}{N}$. This quantity can be estimated by the following inequality:

$$|\varepsilon|^2_i + \frac{\sigma^2_{iv}}{N} < \Delta^2_{z/g} < |\varepsilon|^2_s + \frac{\sigma^2_{sv}}{N} \qquad (1)$$

were $|\varepsilon|^2_i = \inf_{g \in G} m^2_{v/g}$, $|\varepsilon|^2_s = \sup_{g \in G} m^2_{v/g}$ are squares of low and upper boundaries of the module of the estimate shift.

When $N \to \infty$, expression (1) becomes the following one: $|\varepsilon|^2_i < \Delta^2_{z/g} < |\varepsilon|^2_s$.

Clear, it is reasonable to increase the sample volume before this gives perceptible rising of accuracy. As it follows from expression (1) for hyper-random estimates there is a critical sample volume N_0 from that rising of the volume of processing data is not reasonable. In this sense hyper-random estimates are similar to interval estimates [Шокин, 1981].

By using right part of inequality (1) the critical sample volume N_0 can be estimated in the following manner: $N_0 > \frac{10\sigma^2_{sv}}{|\varepsilon|^2_s}$. It follows from this inequality that with decreasing of upper boundary of the module of the estimate shift $|\varepsilon|_s$ and increasing of upper noise variance σ^2_{sv} the critical sample volume rises. If upper boundary of the module of the shift is comparable with upper boundary of noise standard deviation the critical sample volume N_0 is in the region of dozen samples.

Described approaches may be useful for modeling not only physical quantities but also physical processes and fields. They give possibility to find and explain facts unnoticed or inexplained on the ground of traditional approaches.

9. Conclusion

Cognition possibilities are limited. This is called not only by the limited human possibilities in perception and processing information. Main cause consists in unpredictable phenomena occurred in the world.

A lot of data point that nowhere there is absolute statistic stability. In ambient space all objects and processes are changed unpredictably, sooner or later. Researches show that statistical stability of physical phenomena on that founders of classic probability theory and mathematical statistics are oriented can have in real world only local character. In the best case it occurs on small temporal, space, or temporal-space intervals of observation.

The absence of global statistical stability of physical phenomena does not permit to obtain consistent estimates the errors of which tend to zero when the sample volume tends to infinite. This gives rise to essential doubt on adequacy of universally recognize hypothesis that the real world is organized on random principles.

The hypothesis of statistic unpredictability of phenomena that is the physical model of one of the fundamental lows of nature and the hypothesis of hyper-random world organization that is the mathematical model of this physical model help to explain from the single position a lot of key questions of cognition, in particular why

- it is impossible to obtain infinity high accuracy and to create absolutely adequate mathematical model for any real physical object;
- all real estimates and created on them estimate and object models are not consistent ones (when the data volume is rising their errors are not tend to zero);
- it is impossible to prove by any experiments that even the best perfect physical theory absolutely exactly describes nature lows;
- it is impossible by any real measurement means to confirm or refute the existence of world constants;
- it is impossible to make absolutely exact prediction on the basis of the current and former data.

This list can be continued.

From the hypothesis of statistic unpredictability and the hypothesis of hyper-random world organization follow that cognition horizon is defined by the range of unpredictable changing of physical phenomena and conditions of their observation.

* * *

All hypotheses and theories have limited age. It is impossible to foresee exactly if they will be claimed in a future and in what degree will be claimed. True importance of scientific results is covered in a process of time tests. It would be wanted to hope that the hypothesis of statistic unpredictability, the hyper-random hypothesis, and the theory of hyper-random phenomena will be useful for cognition of our world.

Bibliography

[Batyrshin, Kacprzyk, Sheremetov, Zadeh, 2007] I.Batyrshin, J.Kacprzyk, L.Sheremetov, L.A.Zadeh (Eds.). Perception-based Data Mining and Decision Making in Economics and Finance. Series: Studies in Computational Intelligence, Vol. 36. 2007.

[Crownover, 1999] R.M.Crownover. Introduction to Fractals and Chaos. – Boston-London: Jones and Bartlett Publishers, 1999. – 348 p.

[Ferson, Kreinovichy, Ginzburg, Myers, 2003] S.Ferson, V.Kreinovichy, L.Ginzburg, D.S.Myers. Constructing probability boxes and Dempster-Shafer structures. SAND report SAND2002-4015, 2003. – 143 p.

[Guide, 1993] Guide to the expression of uncertainty in measurement. First edition. International organization for standardization. – Switzerland. – ISBN 92-67-10188-9, 1993. – 101 p.

[Gorban, 2006] I.I.Gorban. The hyper-random functions and their description // Radioelectronics and Communications Systems. – 2006. – V. 49, No 1. – P. 1 – 10.

[Gorban, 2006] I.I.Gorban. Stationary and ergodic hyper-random functions // Radioelectronics and Communications Systems. – 2006. – V. 49, No 6. – P. 39 – 49.

[Gorban, 2008] I.I.Gorban. Hyper-random phenomena: definition and description // Information Theories and Applications. – 2008. – V. 15, No 3. – P. 203 – 211.

[Gorban, 2008] I.I.Gorban. Value measurement in statistically uncertain conditions // Radioelectronics and Communications Systems. – 2008. – V. 51, No 7. – P. 349– 363.

[Hagan, Demuth, Beale, 1996] M.T.Hagan, H.B.Demuth, and M.H.Beale. Neural network design. – Boston, MA: PWS Publishing, 1996.

[International standard, 2006] International standard ISO 3534-1:2006(E/F). Statistics. Vocabulary and symbols. Part I: General statistical terms and terms used in probability, 2006.

[Kreinovich, Berleant, Ferson, Lodwick, 2005] V.Kreinovich, D.J.Berleant, S.Ferson, W.A.Lodwick. Combining interval and probabilistic uncertainty: foundations, algorithms, challenges. – An Overview "Proceedings of the International Conference on Fuzzy Systems, Neural Networks, and Genetic Algorithms FNG'05". – Tijuana, Mexico. – 2005. – P. 1-10.

[Mises, 1964] R. von Mises. Mathematical theory of probability and statistics / Edited and complemented by H. Geiringer. – N.Y. and London: Acad. Press, 1964. – 232 p.

[Sharkovsky, Romanenko, 2005] , A.N.Sharkovsky and E.Yu. Romanenko. Turbulence, ideal. – Encyclopedia of Nonlinear Science, New York and London. – 2005. – P. 955-957.

[Shary, 2002] S.P.Shary. A new technique in systems analysis under interval uncertainty and ambiguity. Reliable computing. – 2002. – No 8. – P. 321 – 418.

[Zadeh, Kacprzyk, 1992] L.A.Zadeh and J.Kacprzyk. Fuzzy logic for the management of uncertainty. – New York: John Wiley & Sons, 1992. – 256 p.

[Алефельд, Херцбергер, 1987] Г.Алефельд, Ю.Херцбергер. Введение в интервальные вычисления. – М.: Мир, 1987. – 356 с.

[Алимов, Кравцов, 1981] Ю.И.Алимов, Ю.А.Кравцов. Является ли вероятность «нормальной» физической величиной // Успехи физических наук. – 1992. – Т. 162, № 7. – С. 149 – 182.

[Анищенко, Вадивасова, Окрокверцхов, Стрелкова, 2005] В.С.Анищенко, Т.Е.Вадивасова, Г.А.Окрокверцхов, Г.И.Стрелкова. Статистические свойства динамического хаоса // Успехи физических наук. – 2005. – Т. 175, № 2. – С. 163 – 179.

[Боровков, 1986] А.А.Боровков. Теория вероятностей. М.: Наука, 1986. – 432 с.

[Бочарников, 2001] В.П.Бочарников. Fuzzy-технология: Математические основы. Практика моделирования в экономике. – Санкт-Петербург: Наука. 2001. – 328 с.

[Гнеденко, 1961] Б.В.Гнеденко. Курс теории вероятностей. – М.: Изд-во физ.-мат. литературы, 1961. – 406 с.

[Горбань, 2003] І.І.Горбань Теорія ймовірностей і математична статистика для наукових працівників та інженерів. К.: Інститут проблем математичних машин і систем НАН України, 2003. – 245 с.

[Горбань, 2005] И.И.Горбань Гиперслучайные явления и их описание // Акустичний вісник. – 2005. – Т. 8, № 1 – 2. – С. 16 – 27.

[Горбань, 2005] И.И.Горбань Методы описания гиперслучайных величин и функций // Акустичний вісник. – 2005. – Т. 8, № 3. – С. 24 – 33.

[Горбань, 2006] И.И.Горбань. Оценки характеристик гиперслучайных величин // Математичні машини і системи. – 2006. – № 1. – С. 40 –48.

[Горбань, 2007] И.И.Горбань. Теория гиперслучайных явлений. К.: ИПММС НАН Украины, 2007. – 184 с. (http://ifsc.ualr.edu/jdberleant/intprob/).

[Горбань, 2008] И.И.Горбань. Гиперслучайные марковские модели // Proceedings of XIII-th International conference KDS–2. Sofia –Uzhgorod, Bulgaria – Ukraine, 2008.

[Гринченко, Мацыпура, Снарский, 2005] В.Т.Гринченко, В.Т.Мацыпура, А.А.Снарский. Введение в нелинейную динамику. – К.: Наукова думка, 2005. – 263 с.

[Дыхне, Снарский, Женировский, 2004] А.М.Дыхне, А.А.Снарский, М.И.Женировский. Устойчивость и хаос в двумерных случайно-неоднородных середах и LC-цепочках. // Успехи физических наук. – 2004. – Т. 1174, № 8. – С. 887 – 894.

[Колмогоров, 1936] А.Н.Колмогоров. Основные понятия теории вероятностей. – М.: ОНТИ, 1936. – 175 с.

[Колмогоров, 1987] А.Н.Колмогоров. Теория информации и теория алгоритмов. – М.: Наука. – 1987. – 232 с.

[Королюк, 1985] В.С.Королюк и др. Справочник по теории вероятностей и математической статистике. – М.: Наука, 1985. – 637 с.

[Кравцов, 1989] Ю.А.Кравцов. Случайность, детерминированность, предсказуемость // Успехи физических наук. – 1989. – Т. 158, Вып. 1. – С. 93–122.

[Кузнецов, 1991] В.П.Кузнецов. Интервальные статистические модели. – М.: Радио и связь, 1991. – 348 с.

[Леман, 1979] Э.Леман. Проверка статистических гипотез. – М.: Наука, 1979. – 407 с.

[Степин, 1999] В.С.Степин. Теоретическое знание. М.: Наука, 1999. – 472 с.

[Тутубалин, 1972] В.Н.Тутубалин. Теория вероятности в естествознании. – М.: Знание, 1972. – 48 с.

[Тутубалин, 1972] В.Н.Тутубалин. Теория вероятностей. – М.: Из-во Московского университета, 1972. – 230 с.

[Тутубалин, 1993] В.Н.Тутубалин. Вероятность, компьютеры и обработка результатов эксперимента // Успехи физических наук. – 1993. – Т. 163, № 7. – С. 93 – 109.

[Шокин, 1981] Ю.И.Шокин. Интервальный анализ. – Новосибирск: Наука, 1981. – 112 с.

Authors' Information

Igor Gorban *– Principal scientist of the Institute of Mathematical Machines and Systems Problem, National Academy of Sciences of Ukraine, Glushkov ave., 42, Kiev, 03187, Ukraine; e-mail: igor.gorban@yahoo.com*

Major Fields of Scientific Research: Probability theory and mathematical statistics, Theory of hyper-random phenomena, Space-time signal processing in complicated dynamic conditions, Fast multichannel space-time signal processing

SOLVING LARGE SYSTEMS OF BOOLEAN EQUATIONS

Arkadij Zakrevskij

Abstract. *Systems of many Boolean equations with many variables are regarded, which have a lot of practical applications in logic design and diagnostics, pattern recognition, artificial intelligence, etc. A special attention is paid to systems of linear equations playing an important role in information security problems. A compact matrix representation is suggested for such systems. A series of original methods and algorithms for their solution is surveyed in this paper, as well as the information concerning their program implementation and experimental estimation of their efficiency.*

Keywords: *Solving Boolean equations, large systems, combinatorial search*

ACM Classification Keywords: *G.2.1 Combinatorics. I.2.8. Problem solving, G.3 Probability and Statistics*

Introduction

A special type of systems of logical equations is regarded here, which seems to be very important for applications in logic design, pattern recognition and diagnostics, artificial intelligence, information security, etc. Such systems consist of many equations and Boolean variables (up to thousand and more), but with restricted number of variables in each equation (for example, not exceeding 10). That allows one to represent every equation by a rather short Boolean vector of its roots, providing a compact description of the system as a whole and efficient use of vector logical operations.

In that case each function $\varphi_i(\boldsymbol{x})$ with k arguments from some system F can be represented by a pair of Boolean vectors: 2^k-component *vector* $\boldsymbol{v}_i$ *of function values* (using the conventional component ordering) and n-component *vector* $\boldsymbol{w}_i$ *of function arguments*.

For instance, if $\boldsymbol{x} = (a, b, c, d, e, f, g, h)$, then the pair of vectors $\boldsymbol{v}_i = 01101010$ and $\boldsymbol{w}_i = 00101001$ represents the function $\varphi_i(c, e, h)$ which takes value 1 on four combinations 001, 010, 100 and 110 of argument values and takes value 0 on all others.

The whole system F can be represented by a pair of corresponding Boolean matrices: $(m \times 2^k)$ *matrix* $\boldsymbol{V}$ *of functions* and $(m \times n)$ *matrix* $\boldsymbol{W}$ *of arguments*.

Example 1. The system of Boolean equations

$$\varphi_1 = a'b'cd' \vee a'bc'd \vee ab'c'd$$

$$\varphi_2 = c'd'e'f' \vee c'd'e'f \vee cd'e'f' \vee cd'ef \vee cde'f \vee cdef'$$

$$\varphi_3 = e'fgh' \vee ef'g'h' \vee ef'gh \vee efgh'$$

is represented in matrix form as follows:

$$\boldsymbol{V} = \begin{array}{cl} 0010\ 0100\ 0100\ 0000 & v_1 \\ 1100\ 0000\ 1001\ 0110 & v_2 \\ 0000\ 0010\ 1001\ 0010 & v_3 \end{array} \qquad \boldsymbol{W} = \begin{array}{cccccccc|l} a & b & c & d & e & f & g & h & \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & w_1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & w_2 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & w_3 \end{array}$$

Let us designate these systems as *large SLE*. It is supposed that in many applications these systems usually have few roots or none at all.

A series of original methods and algorithms for solving large SLE is presented in this survey, together with the results of their program implementation. They were published in various papers (see *References*).

Search Tree Minimization

Two combinatorial methods using tree searching technique could be applied to solve large SLE: the *equation scanning method* and the *argument scanning method*. The first method is implementing consecutive multiplication of orthogonal DNFs of the equations from a considered system and uses the search tree T_e which levels correspond to equations. The second method realizes a scanning procedure over arguments corresponding to levels of the search tree T_a. In both cases the run-time is roughly proportional to the size of the tree, i.e. to the number of its nodes. Two original algorithms were worked out that considerably reduce that number in trees T_e and T_a.

Solving large SLE can be considerably accelerating by the described below methods taking into account only the matrix of arguments $\boldsymbol{W}$ [1, 2].

Raising efficiency of the equation scanning method. In that method the nodes of i-th level of the search tree T_e represent the roots of the subsystem, constructed from the first i equations. Let us consider the set of variables, on which this subsystem depends, as $\boldsymbol{U}_i = \boldsymbol{u}_1 \cup \boldsymbol{u}_2 \cup \ldots \cup \boldsymbol{u}_i$ and denote the number of elements in $\boldsymbol{U}_i$ (in other words, the variables included in the first i equations) as $r(i)$. Then roots of the subsystem under review are the elements of the $r(i)$-dimensional Boolean space. Suppose, the functions are random, taking value 1 with probability p on every combination of argument values, independently of each other.

Affirmation 1. The expected value $M_e(i)$ of the number of nodes on the i-th level of tree T_e can be calculated as $M_e(i) = p^i 2^{r(i)}$.

In particular, the number of nodes on the last level is estimated as $M_e(m) = p^m 2^n$. (These nodes represent the solutions of the whole system)

When we include the next equation (given by function $f_{i+1}(\boldsymbol{u}_{i+1})$) into the subsystem, the set of considered variables will expand by the arguments, which are included in $f_{i+1}(\boldsymbol{u}_{i+1})$ but were not presented in any previous function. Thus, the number of possible solutions $M_e(i+1)$ can increase compared to $M_e(i)$. On the other hand, since each new equation represents a new restriction on the set of solutions, $M_e(i+1)$ may be also smaller than $M_e(i)$. The total effect of both tendencies can be represented by the following formula:

Affirmation 2. $M_e(i+1) = M_e(i)\, p\, 2^{r(i+1) - r(i)}$.

This formula shows that the increase in the number of nodes by the transition to the next level depends on the number of new arguments in $f_{i+1}(\boldsymbol{u}_{i+1})$. This number is usually much smaller than the total number of variables in $f_{i+1}(\boldsymbol{u}_{i+1})$.

The algorithm complexity for finding all solutions of the considered system is proportional to the total number of nodes in the tree T_e: $M_e = M_e(1) + M_e(2) + \ldots + M_e(m)$. The number of nodes at the last level can be determined unambiguously: $M_e(m) = p^m 2^n$. However, numbers of nodes on other levels and the total number of nodes M_e depend on the order, in which equations are considered.

We suggest the following method to decrease the algorithmical complexity. All equations are ordered by the following rule: the next equation must contain the minimum number of new variables. At the first step, the equation depending on the minimum number of arguments is selected.

Example 2. Suppose, that $p = 0.5$ and the distribution of the variables by the equations is given by the matrix ***W*** shown below. Note, that the case $p = 0.5$ corresponds to the often encountered in practice situation when characteristic Boolean functions are completely random. Considering the equations in the natural order (according to the rows of matrix ***W***), we get: $M_e(1) = 4$, $M_e(2) = 16$, $M_e(3) = 16$, etc., with the total estimated number of the nodes in the tree $M_e = 67$.

i	***W***	$M_e(i)$
1	0 1 0 0 0 1 0 1	4
2	1 0 0 0 1 1 1 1	16
3	0 1 1 0 1 0 0 1	16
4	1 0 0 1 0 0 1 0	16
5	0 1 1 0 1 0 0 0	8
6	0 1 0 0 0 1 1 0	4
7	1 0 0 1 0 1 1 0	2
8	0 0 1 0 1 0 0 0	1
	M_e =	67

But if we will reorder equations according to the proposed method (using the substitution (8, 5, 3, 1, 6, 2, 4, 7) on the set of rows), we will considerably decrease the computational complexity: $M_e(1) = 2$, $M_e(2) = 2$, $M_e(3) = 2$, etc., with the total estimated number of the nodes in the tree $M_e = 15$.

i	***W***	$M_e(i)$
1	0 0 1 0 1 0 0 0	2
2	0 1 1 0 1 0 0 0	2
3	0 1 1 0 1 0 0 0	2
4	0 1 0 0 0 1 0 1	2
5	0 1 0 0 0 1 1 0	2
6	1 0 0 0 1 1 1 1	2
7	1 0 0 1 0 0 1 0	2
8	1 0 0 1 0 1 1 0	1
	M_e =	15

Affirmation 3. Suppose that $p = 0.5$; $m = n$; $w_i^j = 1$ if $i \le j$, and $w_i^j = 0$ otherwise. In this case the search tree will contain 2^n nodes for the initial order of the equations, and only n nodes for the optimal order.

Raising efficiency of the argument scanning method. In this method we construct the search tree T_a, which shows the bifurcation hierarchy by the values of the Boolean arguments $x_1, x_2, \ldots, x_n$. Each x_j corresponds to one (and only to one) level of the tree, there are n levels in the tree T_a. The nodes on j–th level represent all input vectors of the variables x_1, x_2, …, x_i , for which no function of the initial system will have a zero value. Let us denote by $M_a(j)$ the expected number of nodes on level j.

Affirmation 4. $M_a(j) = 2^j \prod_{i=1}^{m} S(p, q(i, j))$,

where $S(p, r) = 1 – (1–p)^{2^r}$ is the probability that the random function with r arguments (having parameter p) is not equal to 0, and $q(i, j)$ is the number of ones in the i–th row of the matrix $\boldsymbol{W}$, located to the right from the component j .

In particular, the number of solutions of the system equals the number of nodes on the last level, which can be estimated as $M_a(n) = 2^n p^m$. The total number of nodes in tree T_a is given by the formula

$$M_a = M_a(1) + M_a(2) + … + M_a(n).$$

When the number r of the arguments of a random Boolean function is increasing, the probability $S(p, r)$ that this function is not constant zero is swiftly going to 1. For example, if $p = 0.5$, then $S(p, r) = 1 – 2^{-2^r}$ ($S(0) = 1/2$, $S(1) = 3/4$, $S(2) = 15/16$, $S(3) = 255/256$, $S(4) = 65535/65536$, etc.) In practice, we can take $S(r) = 1$ if $r > 3$.

In the proposed algorithm we are optimizing the order, in which variables are selected. As the criteria of minimization, the expected number of nodes in the sequentially considered tree levels is used.

When the next level j is considered and the corresponding argument is selected, the effect of this choice is estimated in advance. Whenever some specific value of some argument is selected and substituted into the equation depending on this variable, the number u of the free variables in this equation decreases by one. As a result, the probability $S(u)$ that the equation can be satisfied, is changed for $S(u–1)$, i.e. decreases in $S(u)/S(u–1)$ times. We will use the notation $R(u)=S(u)/S(u–1)$, for example, $R(1) = 3/2$, $R(2) = 15/12$, $R(3) = 255/240$, $R(4) = 65535/65280$. In practice, we can assume that $R(u) = 1$ if $u > 4$.

Affirmation 5. During the transition from level $j – 1$ to level j the mathematical expectation of the number of nodes in the level is increasing in $2 / \prod R(q(i, j))$ times, where the product is taken by all i, for which $w_i^j = 1$.

In the proposed algorithm at each step an argument is selected such that the number of nodes in corresponding tree level is minimized. The procedure works differently, depending on whether there exists a row in the argument matrix $\boldsymbol{W}$ containing not more than 4 ones. If all rows in this matrix contain more than 4 ones, we choose rows with the minimum number of ones, and select the column j having the maximal number of ones in the chosen rows. The argument x_j is taken as the next one, and the j-th column is deleted from the further consideration. The procedure is repeated until a row will appear which contains not more than 4 ones.

To choose the next argument, we calculate the value $\prod R(q(i, j))$, using the already known values of $R(1)$, $R(2)$, $R(3)$, $R(4)$. The variable j with the maximal value of $\prod R(q(i, j))$ is selected.

Example 3. Let us consider the system from the example 1, considering the arguments in the order (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8). Taking into account the number of ones to the right from the position i, we obtain:

$$M_a(1) = 2 \cdot S(3) \cdot S(4) \cdot S(4) \cdot S(2) \cdot S(3) \cdot S(3) \cdot S(3) \cdot S(2) = 1.731,$$

$$M_a(2) = 4 \cdot S(2) \cdot S(4) \cdot S(3) \cdot S(2) \cdot S(2) \cdot S(2) \cdot S(3) \cdot S(2) = 2.874.$$

For other j we calculate the following values of $M_a(j)$:

$$M_a(3) = 3.463;\ M_a(4)= 5.214;\ M_a(5)= 3.693;$$
$$M_a(6) = 3.560;\ M_a(7) = 1.687;\ M_a(8)= 1.000.$$

The total number of nodes in the tree T_a is calculated as

$$M_a = M_a(1) + M_a(2) + \ldots + M_a(8) = 23.223.$$

Now, we will use another ordering of the arguments, applying the proposed algorithm. First, we will select the input variable x_5 , included into equations 2, 3, 5 and 8, since for $j = 5$ the value of $\prod R(q(i, j))$ for $w_i^j = 1$ is maximal (equal to 1.333). We will delete the corresponding (x_5) column from the further consideration. At the second step, variable x_3 will be selected. The final optimized order (x_5, x_3, x_2, x_8, x_6, x_7, x_1, x_4) is represented by the following column transfer in the matrix ***W***:

$$\begin{array}{c|cccccccc|cccccccc}
 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_5 & x_3 & x_2 & x_8 & x_6 & x_7 & x_1 & x_4 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
2 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\
3 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
4 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
5 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
6 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\
7 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
8 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}$$

Let us estimate the complexity of the search tree for the new argument order:

$$M_a(1) = 2 \cdot S(3) \cdot S(4) \cdot S(3) \cdot S(3) \cdot S(2) \cdot S(3) \cdot S(4) \cdot S(1) = 1.384,$$
$$M_a(2) = 4 \cdot S(3) \cdot S(4) \cdot S(2) \cdot S(3) \cdot S(1) \cdot S(3) \cdot S(4) \cdot S(0) = 1.390.$$

Similarly we calculate the next values $M_a(j)$: $M_a(3) = 1.313$; $M_a(4) = 1.395$; $M_a(5) = = 1.395$; $M_a(6) = 1.318$; $M_a(7) = 1.125$; $M_a(8) = 1.000$. Thus, we see that the expected value of the number of nodes in the tree T_a equals 10.620, which is more than twice less than it was for the initial order of arguments.

The program implementation and computer experiments confirm the high efficiency of the both methods. They show also that the argument scanning method greatly surpasses in efficiency the other one.

***

The search for solutions can be greatly facilitated by preliminary reducing the number of roots in separate equations, which, in its turn, could lead to decreasing the number of variables in a considered system and the number of equations. Three reduction methods are suggested for that, called *local reduction*, *spreading of constants* and *technique of syllogisms* [3].

The main idea of these methods consists in analyzing one by one equations of the system F, revealing there so called *k-bans* (affirmations about existence of some empty interval of the rank k in the Boolean space over the equation variables – were the equation has no root), and using them for reducing the sets of roots in other equations which, in its turn, contributes to finding new bans. That process has the chain character and can result in reducing the number of equations and variables in the system F. The method of constants spreading deals with 1-bans, the technique of syllogisms operates with 2-bans (using original deduction procedures for solving polysyllogisms), and the method of local reduction is using bans of arbitrary rank. Each of them has its own area of preferable application.

Local Reduction

This method suggested in [4] has the local nature. That means that the possibility of reduction is looked for when examining various pairs of functions $\varphi_i(\boldsymbol{u}_i)$ and $\varphi_j(\boldsymbol{u}_j)$ from the system F with intersecting sets of arguments: $\boldsymbol{u}_{i,j} = \boldsymbol{u}_i \cap \boldsymbol{u}_j \neq \varnothing$.

Let us introduce some denotations. Consider the characteristic set M_i of function $\varphi_i(\boldsymbol{u}_i)$ in the space of arguments from the set $\boldsymbol{u}_i$, and let $\boldsymbol{a}$ be its arbitrary element: $\boldsymbol{a} \in M_i$. The latter is a k-component Boolean vector, where k is the number of arguments of function $\varphi_i(\boldsymbol{u}_i)$: $k = |\boldsymbol{u}_i|$. Let $\boldsymbol{v}$ be an arbitrary subset from $\boldsymbol{u}_i$ ($\boldsymbol{v} \subseteq \boldsymbol{u}_i$) and $\boldsymbol{a} / \boldsymbol{v}$ – the *projection of element* ***a*** *onto* ***v***, i.e. the vector composed of those components of vector $\boldsymbol{a}$ which correspond to variables included in set $\boldsymbol{v}$.

The set of all different projections of elements from M_i on $\boldsymbol{v}$ is named the *projection of set M_i on* ***v*** and designated as $M_i / \boldsymbol{v}$. Let $M_{i,j}$ be the intersection of sets $M_i / \boldsymbol{u}_{i,j}$ and $M_j / \boldsymbol{u}_{i,j}$, and $M_{i/j}$ – the set of all such elements from M_i which projections on $\boldsymbol{u}_{i,j}$ belong to the set $M_{i,j}$.

For example, if $\boldsymbol{u}_i = (a, b, c, d, e)$, $\boldsymbol{u}_j = (c, d, e, f, g, h)$, $M_i = (01101, 11010, 10011)$ and $M_j = (101110, 001101, 010010)$, then $\boldsymbol{u}_{i,j} = \boldsymbol{u}_{j,i} = (c, d, e)$, $M_{i,j} = M_{j,i} = (101, 010)$, $M_{i/j} = (01101, 11010)$ and $M_{j/i} = (101110, 010010)$.

Let us introduce the operation $M_i := M_{i/j}$ of changing M_i for $M_{i/j}$.

Affirmation 6. For any $i, j = 1, 2, ..., m$ the operation $M_i := M_{i/j}$ is an equivalence transformation of system F, preserving the set of its roots.

Note that the application of this operation to the shown above example reduces each set M_i and M_j by one element.

Let us say that operation $M_i := M_{i/j}$ is *applicable* to an ordered pair of functions (φ_i, φ_j) if $M_i \neq M_{i/j}$. The probability of its applicability rises with increasing of the cardinality $|\boldsymbol{u}_{i,j}|$ of set $\boldsymbol{u}_{i,j}$ and goes down when $|\boldsymbol{u}_{i,j}|$ decreases. For instance, it is rather high when $|M_j| < 2^s$, where $s = |\boldsymbol{u}_{i,j}|$.

Consider now the procedure of sequential execution of this operation on pairs where it can be applied. It could terminate with reducing some of the sets M_i down to the empty set, which will mean that system F is inconsistent, or some set of reduced functions will be found where the given operation cannot be applied to any pair. This procedure is called the *local reduction* of system F.

Let us demonstrate the described algorithm of local reduction using the following example of system F.

$$\boldsymbol{V} = \begin{array}{cl} 0010\ 0100\ 0100\ 0000 & \boldsymbol{v}_1 \\ 1100\ 0000\ 1001\ 0110 & \boldsymbol{v}_2 \\ 0000\ 0010\ 1001\ 0010 & \boldsymbol{v}_3 \end{array} \qquad \boldsymbol{W} = \begin{array}{ccccccccl} a & b & c & d & e & f & g & h & \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & \boldsymbol{w}_1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & \boldsymbol{w}_2 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & \boldsymbol{w}_3 \end{array}$$

Regard in succession pairs of functions, beginning with the first one: (φ_1, φ_2). Using the operation of component-wise conjunction of corresponding rows of matrix ***W***, we find for this pair common arguments c and d. Going through all combinations of values of these variables, we examine defined by them intervals in the space of arguments of function φ_1 (this space is presented by vector $\boldsymbol{v}_1$) and find between them intervals free of values 1 of this function. Then we delete all 1s in corresponding intervals of vector $\boldsymbol{v}_2$.

Vector representation of intervals and component-wise logical operations are used during this procedure. For example, considering combination 00 of values of variables c and d, we construct vector 1000 1000 1000 1000 which marks with 1s the corresponding interval in the space of variables a, b, c, d. Its conjunction with vector $\boldsymbol{v}_1$ does not contain ones, therefore equation $\varphi_1 = 1$ has no roots in this interval. The respective interval in the space of arguments of function φ_2 is represented by vector 1111 0000 0000 0000, inasmuch as variables c and d take now left positions. All ones contained in this interval are deleted from vector $\boldsymbol{v}_2$, so the latter receives the value 0000 0000 1001 0110.

These operations could be presented in a more compact form, by the formula

$$C' d' \varphi_1 = 0 \ \rightarrow \ \boldsymbol{v}_2 := 0000\ 0000\ 1001\ 0110\ .$$

Continuing the reduction algorithm, we perform one by one the following operations presented similarly:

$$c\, d\, \varphi_1 = 0 \ \rightarrow \ \boldsymbol{v}_2 := \ 0000\ 0000\ 1001\ 0000\ ,$$

$$c' d\, \varphi_2 = 0 \ \rightarrow \ \boldsymbol{v}_1 := \ 0010\ 0000\ 0000\ 0000$$

$$e' f\, \varphi_2 = 0 \ \rightarrow \ \boldsymbol{v}_3 := \ 0000\ 0000\ 1001\ 0010$$

$$e\, f'\, \varphi_2 = 0 \ \rightarrow \ \boldsymbol{v}_3 := \ 0000\ 0000\ 0000\ 0010$$

$$e' f'\, \varphi_3 = 0 \ \rightarrow \ \boldsymbol{v}_2 := \ 0000\ 0000\ 0001\ 0000$$

As a result, the initial system of Boolean functions is reduced to the following one:

$$\boldsymbol{V} = \begin{array}{cc} 0\,0\,1\,0\ \ 0\,0\,0\,0\ \ 0\,0\,0\,0\ \ 0\,0\,0\,0 & \boldsymbol{v}_1 \\ 0\,0\,0\,0\ \ 0\,0\,0\,0\ \ 0\,0\,0\,1\ \ 0\,0\,0\,0 & \boldsymbol{v}_2 \\ 0\,0\,0\,0\ \ 0\,0\,0\,0\ \ 0\,0\,0\,0\ \ 0\,0\,1\,0 & \boldsymbol{v}_3 \end{array}$$

from where the unique root of the system is easily obtained: 00101110.

Spreading of Constants

This method can be regarded as a simplified version of the local reduction. It can be efficiently used when the number of roots in some equations $\varphi_k = 1$ is very small. In that case it is enough to look only for 1-bans regarding separate literals x_i and x_i' and checking them consecutively for satisfying equations.

When $x_i \wedge \varphi_k = 0$ for function φ_k in some equation of the system, 1-ban $x_i = 0$ is found. In that case value 0 is assigned to variable x_i (value 1 in the case $x_i' \wedge \varphi_i = 0$), and the latter is changed for constant 0 (or 1) in all other equations. In such a way finding constants is followed by their spreading over the whole system F. Replacing some variables by constants usually decreases the number of roots in regarded equations which, in its turn, helps to discover new constants. So, the process of constants spreading has the cyclic chain nature. As a result, the dimension of processed equations is decreased, sometimes down to zero – when all variables of the regarded equation receive definite values. If function φ_k turns into 1, the corresponding equation is deleted from the system; if φ_k turns into 0, it becomes evident that the system is inconsistent.

Simple enough, this method turned out to be very efficient, being applied to some problems of cryptology. A special problem of cryptanalysis of the mechanical rotor encryption machine Hagelin M-209-B, which was used in several forms by Germans during the second world war, was investigated in [5]. It was shown that its cryptanalysis can be reduced to solving a definite system of many Boolean equations (about five hundred) each of which contains six Boolean variables, meanwhile the general number of variables equals 131 – the set of their values constitutes the sought-for key. To solve this system a method was proposed in [5] based on using Reduced Ordered Binary Decision Diagrams (ROBDDs) for representation of the regarded functions. Its computer implementation on Pentium Pro 200 showed that under some suppositions it enables to find the key in several minutes.

Application of the method of spreading of constants using vector representation of the considered Boolean functions and taking into account the specific of the regarded system of logical equations turned out to be considerably more efficient. It accelerates the search for the key more than in thousand times [6].

Technique of Syllogisms

Here 2-bans are looked-for and used in the reduction procedure. Besides, the latter takes into account all logical consequences deduced from the set of found 2-bans by syllogisms [7]. An improved technique of polisyllogisms is applied for that [8].

Let us regard equation $\varphi(z_1, z_2, \ldots, z_k) = 1$ with function φ taking value 1 on s randomly selected inputs. When s is small, it is possible to find some constant, which prohibits the value of some variable (1-ban). But it is more probable to reveal a prohibition on some combination of values of two variables (2-ban), which determines the corresponding *implicative regularity*, or connection between these variables. For example, connection "if a, then not b" prohibits combination of values $a = 1$, $b = 1$. It could be revealed in φ if $ab\varphi = 0$. For convenience, represent this ban by product ab (having in mind equation $ab = 0$).

In a similar way, 2-bans ab', $a'b$, $a'b'$ are defined. They are interpreted easily as general affirmation and negation category statements. By that besides three such statements of Aristotle syllogistic (ab' – all A are B, $a'b$ – all B are A, ab – none of A is B) the fourth is also used: $a'b'$ – none of objects is A and is not B. Such a statement was not considered by Aristotle, inasmuch as he did not regard empty classes [9].

Suppose, that by examining equations of the system F one by one, we have found a set P of 2-bans. Let us consider the task of *closing* it, i. e. adding to it all other 2-bans which logically follow from P (so called *resolvents* of P). This task is equivalent to the polysyllogistic problem. Denote the resulting *closed set of 2-bans* as $Cl(P)$. A method to find it is suggested below. It differs from the well-known method of resolution and its graphical version by application of vector-matrix operations which speed up the logical inference.

Let X_t^1 and X_t^0 be the sets of all literals that enter 2-bans contained in F together with literal x_t or x_t', correspondingly. We introduce operator Cl_t of *partial closing* of set P in regard to variable x_t, extending this set by uniting it with direct product $X_t^1 \times X_t^0$ containing results of all possible resolutions by this variable.

Affirmation 7. $\mathrm{Cl}_t(\mathrm{P}) = \mathrm{P} \cup \mathrm{X}_t^1 \times \mathrm{X}_t^0 \subseteq \mathrm{Cl}(\mathrm{P})$.

Affirmation 8. $\mathrm{Cl}(\mathrm{P}) = \mathrm{Cl}_1\,\mathrm{Cl}_2 \ldots \mathrm{Cl}_n\,(\mathrm{P})$.

In such a way, the set P can be closed by separate variables, one by one.

The set P can be represented by a square Boolean matrix $\boldsymbol{P}$ of the size $2n$ by $2n$, with rows $\boldsymbol{p}_{t1}$, $\boldsymbol{p}_{t0}$ and columns $\boldsymbol{p}^{t1}$, $\boldsymbol{p}^{t0}$ corresponding to literals x_t, x_t', $t = 1, 2, \ldots, n$. Elements of matrix $\boldsymbol{P}$ correspond to pairs of literals, and non-diagonal elements having value 1 represent discovered 2-bans. So, the totality of 1s in row $\boldsymbol{p}_{t1}$ (as well as in column $\boldsymbol{p}^{t1}$) indicates set X_t^1, and the totality of 1s in row $\boldsymbol{p}_{t0}$ (column $\boldsymbol{p}^{t0}$) indicates set X_t^0. Using vector operations, we can construct the matrix $\boldsymbol{P}^+$, presenting the result of closing operation: $\boldsymbol{P}^+ = Cl(P)$.

For example, if $\boldsymbol{x} = (a, b, c, d)$ and 2-bans ab', ac, $a'd'$, bc' are found forming set P, then

$$\boldsymbol{P} = \begin{array}{cccccccc|c}
a & a' & b & b' & c & c' & d & d' & \\
0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & a \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & a' \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & b \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & b' \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & c' \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & d \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & d'
\end{array}
\qquad
\boldsymbol{P}+ = \begin{array}{cccccccc|c}
a & a' & b & b' & c & c' & d & d' & \\
\boldsymbol{c} & 0 & \boldsymbol{c} & 1 & 1 & \boldsymbol{b} & 0 & \boldsymbol{c} & a \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & a' \\
\boldsymbol{c} & 0 & 0 & 0 & 0 & 1 & 0 & \boldsymbol{c} & b \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{a} & b' \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{a} & c \\
\boldsymbol{b} & 0 & 1 & 0 & 0 & 0 & 0 & \boldsymbol{b} & c' \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & d \\
\boldsymbol{c} & 1 & \boldsymbol{c} & \boldsymbol{a} & \boldsymbol{a} & \boldsymbol{b} & 0 & \boldsymbol{c} & d'
\end{array}$$

– the bans-consequences are marked in matrix ***P***+ by symbols of variables by which the corresponding resolutions were executed.

The closed set *Cl*(*P*) could be found also by the *increment algorithm of expansion* of *P* : every time when a new 2-ban *p* is added by a special operation *ins*(*p, P*) all resolvents are included in *P,* too. In that case after each step the set *P* will remain closed: *P* = *Cl*(*P*).

Operation *ins*(*p, P*) is defined as follows.

Affirmation 9. If $P = Cl(P)$, than $Cl(P \cup \{p\}) = P \cup D$, where

$D = (\{x\} \cup X0\,) \times (\{y\} \cup Y0\,)$, if $p = xy$,

$D = (\{x\} \cup X0\,) \times (\{y'\} \cup Y1\,)$, if $p = xy'$,

$D = (\{x'\} \cup X1\,) \times (\{y\} \cup Y0\,)$, if $p = x'y$,

$D = (\{x'\} \cup X1\,) \times (\{y'\} \cup Y1\,)$, if $p = x'y'$.

Consider now the problem of finding all *prime bans* (which do not follow from one another) deduced from system *P*. It is known that no set of 2-bans can produce any bans of higher rank. But it can produce some 1-bans, prohibiting definite values of separate variables.

Affirmation 10. All 1-bans deduced from set *P* are represented by 1-elements of the main diagonal of matrix $\boldsymbol{P}^{+}$.

In the regarded example 1-bans *a* and *d*' are presented in such a way.

Affirmation 11. If the pair of 1-bans *x* and *x′* is found for some variable *x*, the system *F* is inconsistent.

Note that inconsistency of *F* follows from inconsistency of *P*, but not vice versa.

Based on the technique of syllogisms an efficient reduction method was developed, dealing with a set of logical equations *F*, empty at the beginning. It examines the equations in cyclic order, reduces the set of roots of the current equation $f_j = 1$ by considering bans enumerated in *P* (prohibited roots are deleted) and looks there for new 2-bans not existing in *P*. These bans are added to *P*, at the same time operation of closing *P* is performed. By that some variables can receive unique values – when 1s appear on the main diagonal of matrix ***P*** (1-bans are found). The procedure comes to the end when inconsistency is revealed (0-ban is found represented by a pair of 1s on the main diagonal of ***P***) or when processing *m* equations one by one turns out to be unsuccessful. In that case we have as a result a reduced system of equations equivalent to the initial one.

Computer Experiments

Extensive computer experiments were conducted on PC Pentium 100 to evaluate the efficiency and applicability of the suggested reduction methods [10, 11]. A series of pseudo-random consistent (having at least one root) systems of Boolean equations with given parameters (m – the number of equations, n – the total number of variables, k – the number of variables in each equation and p – the relative number of roots in equations) was generated [12] and subjected to the reduction procedures programmed in C++. Two important results were obtained by that.

First, the *avalanche* effect of reduction was revealed experimentally, both for the local reduction and the technique of syllogisms. When we conduct experiments for fixed values of p, n and k, gradually increasing m, it turns out that for some crucial value of m an avalanche occurs. It means that the number of roots in the equations dramatically decreases in such a high degree that it could be easy to find the complete solution of the regarded system. This effect is well shown on Table 1, where partial results of some experiments are presented. Note that q s the average number of remaining roots in one equation after the reduction. Evidently, if $q = 1$, the system has only one root, and it is found.

Table 1. Examples illustrating avalanche effect of the reduction procedures

Local reduction

Experiment 1. $p = 1/2$, $n = 50$, $k = 5$. $(m: q) = 113: 8.19,\ 114: 8.21,\ 115: 1.$

Experiment 2. $p = 1/2$, $n = 50$, $k = 6$. $(m: q) = 298: 30.02,\ 299: 1.$

Experiment 3. $p = 1/4$, $n = 100$, $k = 6$. $(m: q) = 167: 13.88,\ 168: 1.$

Experiment 4. $p = 1/4$, $n = 100$, $k = 7$. $(m: q) = 390: 28.53,\ 391: 1.$

Experiment 5. $p = 1/4$, $n = 200$, $k = 6$. $(m: q) = 384: 14.29,\ 385: 1.$

Experiment 6. $p = 1/8$, $n = 200$, $k = 6$. $(m: q) = 72: 7.93,\ 73: 1.09.$

Experiment 7. $p = 1/8$, $n = 200$, $k = 7$. $(m: q) = 196: 13.25,\ 197: 1.$

Technique of syllogisms

Experiment 8. $p = 1/2$, $n = 50$, $k = 5$. $(m: q) = 74: 13.32,\ 75: 1.07.$

Experiment 9. $p = 1/4$, $n = 100$, $k = 6$. $(m: q) = 85: 14.25,\ 86: 1.05.$

Experiment 10. $p = 1/8$, $n = 200$, $k = 7$. $(m: q) = 128: 15.76,\ 129: 1.02.$

Second, the existence of avalanche effect enables practically for every combination of values of parameters p, n, k to find the crucial value m_c indicating the number of equations m at which the system collapses under the influence of the reduction procedure. Assume that such a collapse occurs when q becomes less than 1.1.

That crucial value m_c is shown below (the first number in a pair playing the role of the table element) both for local reduction and technique of syllogisms, for $p = 1/2$, 1/4 and 1/8, as the function of n and k. The run-time t in seconds is presented by the second number in the pair. For instance, if $n = 80$ and $k = 7$, then for local reduction and $p = 1/4$ it follows that $m_c = 245$ and $t = 54\ s$. These results show that the area of applicability of the suggested method is rather broad, up to thousand variables under certain conditions.

Table 2. Dependence of the crucial value m_c of the number of equations m and the run-time t on the total number of variables n, the number of variables k and the density of roots p in separate equations

Local reduction

	n	*k* = 4	*k* = 5	*k* = 6	*k* = 7	*k* = 8	*k* = 9	*k* = 10
	20	38-1	30-2	34-6	41-10	37-39	64-173	60-431
	40	67-3	74-8	195-57	426-404			
	80	144-10	273-47	1355-989				
p = 1/2	160	372-33	1094-973					
	320	655-96						
	640	1245-358						
	1280	1500-647						
	20	14-0	13-0	10-0	13-0	13-2	13-5	17-5
	40	20-0	26-1	26-1	45-4	113-41	204-230	431-2493
	80	40-1	46-1	78-5	245-54	1031-1368		
p = 1/4	160	85-2	123-4	314-36	1226-648			
	320	250-10	250-17	821-174				
	640	550-42	475-60	2265-1146				
	1280	1100-179	1000-251					
	2560	2200-995						
	20	10-0	10-0	10-0	10-0	10-0	10-1	10-3
	40	25-0	21-0	15-1	17-1	22-2	36-8	52-27
	80	35-1	30-1	25-1	43-2	123-18	256-130	1008-2598
p = 1/8	160	82-1	67-1	74-2	134-9	662-267	2287-4000	
	320	199-4	166-5	158-6	346-42			
	640	415-19	329-15	352-29	705-129			
	1280	749-74	736-96	742-137	1568-673			
	2560	1965-633	1802-626	1423-565				

Technique of syllogisms

	n	*k* = 4	*k* = 5	*k* = 6	*k* = 7	*k* = 8
	20	30-0	39-1			
	40	42-0	73-2			
	80	124-2	217-7			
p = 1/2	160	217-5	615-23			
	320	573-25	1446-61			
	640	1191-166				
	1280	2438-1371				
	20	14-0	13-0	9-1	471-33	
	40	28-0	20-2	24-2	321-25	
	80	50-1	47-1	39-4	620-53	
p = 1/4	160	129-3	109-5	166-12	1825-142	
	320	296-24	333-26	452-44		
	640	797-201	663-186	894-220		
	1280	1371-1585				
	20	10-0	10-0	10-0	10-1	86-11
	40	18-0	21-1	14-1	27-3	177-23
	80	40-0	31-1	32-2	46-5	357-52
p = 1/8	160	111-4	78-4	79-7	101-15	917-125
	320	271-26	267-29	212-32	227-46	
	640	596-226	537-223	409-166	413-187	
	1280	1027-1701				

Systems of Linear Logical Equations – Finding Shortest Solutions

In general case any system of linear logical equations (SLLE) can be presented as

$$a_1^1x_1 \oplus a_1^2x_2 \oplus \ldots \oplus a_1^nx_n = y_1 ,$$
$$a_2^1x_1 \oplus a_2^2x_2 \oplus \ldots \oplus a_2^nx_n = y_2 ,$$
$$a_m^1x_1 \oplus a_m^2x_2 \oplus \ldots \oplus a_m^nx_n = y_m ,$$

or in a more compact form of one matrix equation

$$\boldsymbol{A}\ \boldsymbol{x} = \boldsymbol{y}.$$

Here ***A*** is a Boolean ($m \times n$)-matrix of coefficients, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ – a Boolean vector of unknowns and $\boldsymbol{y} = (y_1, y_2, \ldots, y_m)$ – a Boolean vector of free terms. The operation of multiplying matrix ***A*** by vector ***x*** is defined as follows:

$$\bigoplus_{j=1}^{n} a_i^jx_j = y_i , \quad i = 1, 2, \ldots, m.$$

Suppose ***A*** and ***y*** are known and vector ***x*** is to be found. It is accepted usually, that the problem consists in finding a *root* of the system – a value of vector ***x***, satisfying all equations, i.e. turning them into identities. However, when several roots exist, a problem arises to choose one of them, optimal in some sense.

Alongside with two parameters m and n the third parameter of an SLLE is important – the *rank* r, i.e. the maximal number of linearly independent columns in matrix ***A***. Remind, that a set of Boolean vectors is called *linearly independent* if the component-wise sum (modulo 2) of any of its elements differs from zero. It is known that the rank equals as well the maximal number of linearly independent rows in the same matrix. The relations between parameters m, n and r determine if the system has some roots and how many of them.

In case $n = m = r$ the system has exactly one root and is called *defined*, or *deterministic*. When $n < m$, the system could have no roots and is called in this case *over-defined*, or *inconsistent*, or *contradictory*. When $n > m$, the system has 2^{n-r} roots and is called *undefined*, or *non-deterministic*.

In this section the last case is considered, and the optimization task of finding a *shortest* solution (with minimum number of ones in vector ***x***) is to be solved. That task has important application at design of linear finite automata [13] and logic circuits synthesized in Zhegalkin basis and possessing such attractive properties, as good testability and compactness at implementation of arithmetic operations [14]. It is useful also when solving information security problems [15, 16].

A simplest algorithm. Let us suppose that a regarded system is undefined and $m = r$, i.e. all rows of matrix of coefficients ***A*** are linearly independent. In that case a shortest solution could be found by means of selecting from matrix ***A*** one by one all different combinations of columns, consisting first of 1 column, then of 2 columns, etc. and examining them to see if their sum equals vector ***y***. As soon as it happens, the current combination is accepted as the sought-for solution.

That moment could be forecasted. If the *weight* of the shortest solution (the number of 1s in vector ***x***) is *w*, the number *N* of checked combinations is defined approximately by the formula

$$N = \sum_{I=0}^{w} C_n^i$$

and could be very large, as is demonstrated below.

It was shown in [17] that the expected weight γ of the shortest solution of an SLLE with parameters *m* and *n* can be estimated before finding the solution itself. We represent this weight as the function $\gamma(m, n)$. First we find the mathematical expectation $\alpha(m, n, k)$ of the number of solutions with weight *k*. We assume that the considered system was randomly generated, which means that each element of ***A*** takes value 1 with the probability 0.5 and any two elements are independent of each other. Then the probability that a randomly selected column subset in matrix ***A*** is a solution equals 2^{-m} (probability that two randomly generated Boolean vectors of size *m* are equal). Since the number of all such subsets having *k* elements equals C_n^k, we get:

$$\alpha(m, n, k) = C_n^k 2^{-m}, \text{ where } C_n^k = n! / ((n-k)!\, k!).$$

Similarly, we denote as $\beta(m, n, k)$ the expected number of the solutions with weight not greater than *k*:

$$\beta(m, n, k) = \sum_{i=0}^{k} C_n^i 2^{-m}.$$

Now, the expected weight γ of the shortest solution can be estimated well enough by the maximal value of *k*, for which $\beta(m, n, k) < 1$:

$$\gamma(m, n) = k,$$

where $\beta(m, n, k) < 1 \le \beta(m, n, k+1)$

For example, the values of α and β for the system of 40 equations with 70 variables and the values of *k* from 7 to 13 are shown in Table 3.

k	α	β
7	0.001	0.001
8	0.009	0.010
9	0.059	0.069
10	0.361	0.430
11	1.968	2.398
12	9.676	12.074
13	43.170	55.244

It is clear enough that the weight of the shortest solution for this system will be probably equal to 10.

Unfortunately, the described above simple algorithm could appear too difficult to implement. Regarding another example with $m = 100$ and $n = 130$ we find that $\gamma = 31$, and the number N of checked combinations is about 10^{30}. Examining them with the speed of one million combinations per second we need about **30 000 000 000 000 000 years** to find the solution. Too much!

Gaussian method.The well-known Gaussian method of variables exclusion [18] was developed for solving systems of linear equations with real variables, and is adjusted here for Boolean variables. It enables to avoid checking all 2^n subsets of columns from Boolean matrix ***A*** which have up to w columns, when only one of 2^{n-m} regarded combinations presents some root of the system.

Its main idea consists in transforming the extended matrix of the system (matrix ***A*** with the added column ***y***) to the *canonical form*. A maximal subset of m linear independent columns (does not matter which one) is selected from ***A*** and by means of equivalent matrix transformations (adding one row to another) is transformed to **I**-matrix, with 1s only on the main diagonal. That part is called a *basic*, the rest $n - m$ columns of the transformed matrix ***A*** constitute a *remainder*. The column ***y*** is changed by that, too.

According to this method the subsets of the remainder are regarded, i.e. combinations selected from the set which has only $n - m$ columns (not all n columns!). It is easy to show that every of these combinations enables to get a solution of the considered system. Indeed, any sum (modulo 2) of its elements can be supplemented with some columns from **I** to make it equal to ***y***.

When we are looking for a shortest solution (solution with the minimum weight) using this method, described in detail in [19], we have to consider different subsets of columns from the remainder, find the solution for each such subset and select a subset, which generates the shortest solution. If it is known that the weight of the shortest solution is not greater than w, then the level of search (the cardinality of inspected subsets) is restricted by w. Note that if $w \geq n - m$, then all 2^{n-m} subsets must be searched through.

For the same example ($m = 100$ and $n = 130$) $N \cong 10^9$, which means that the run-time of Gaussian method is about **17 minutes**.

Decomposition method. An additional gain can be received by decomposition of the process of solution, at which instead of one canonical form of matrix ***A*** several canonical forms are considered. That idea was realized before by the author who suggested a decomposition method for finding shortest solutions [17]. That method is based on constructing a set of different but equivalent canonical forms of the regarded SLLE and solving them in parallel until a shortest solution is found. The run-time of the implementation program depends much on the level of combinatorial search, and the lowering of this level can greatly accelerate the search process.

Let us assume that we can find q maximal subsets of linear independent columns in matrix ***A***, such that the corresponding remainders do not intersect. In this case $n \geq q(n - m)$. The following method can be used, which was called the *method of non-intersecting remainders*.

Let us construct the set Q, consisting of q canonical forms, such that the basics of these forms are obtained using the considered subsets. We will search for the optimal solution within the set Q, with the subsequent increase in the level of search up to some value.

Affirmation 12. A canonical form always exists in Q, such that a shortest solution can be found on the search level not greater than $\rfloor \gamma/q \lfloor$.

Affirmation 13. A shortest solution for the given system can be found by the subsequent consideration of the remainders of the canonical forms from the set Q, restricting the search level by the value $\rfloor (w-1)/q \lfloor$, where w is the weight of the shortest already found solution.

Based on these statements the decomposition method was proposed to find a shortest solution of a system of linear logical equations. Using this method we search through the subsets in all remainders first on level 0, then on level 1, etc. until the level of search reaches $\rfloor (w-1)/q \lfloor$(the nearest integer from below).

Affirmation 14. The number $N_r(m, n)$ of the subsets of the columns, which are considered using the not-intersecting remainders method, is defined by the formula:

$$N_r(m, n) = q \sum_{i=0}^{p} C_{n-m}^{i} \,,$$

where $p = \rfloor \gamma\, (m, n) / q \lfloor$.

For the same example ($m = 100$ and $n = 130$), $q = 4$, $\gamma = 31$ and $p = \rfloor 31/4 \lfloor = 7$. In this case $N_r \cong 10^7$, that means that the run-time of this method is about **ten seconds**.

Recognizing short solutions. A much bigger progress in the run-time saving can be achieved in the case when some short solution exists, with weight w perceptibly smaller than γ [19].

Such a solution which satisfies the relation $w < \gamma$ or even $w < \gamma - 1$ could be immediately recognized and accepted without any additional proof. That enables to increase considerably the size of regarded and solved systems, which is measured in number of equations and variables (m and n).

Consider, for example, a random SLLE with $m = 300$ and $n = 350$ with expected weight of a shortest solution $\gamma = 101$. In general case such solution could be found on the level of search 21, and we should spend about **61 years** to find it by the decomposition method (examining one million combinations per second). However, when a solution with weight 70 exists, it can be found and recognized on the level of search 7 in 7 minutes, and a solution with weight 35 can be found on level 2 in **only 0.5 seconds**.

Randomized parallelization. A new version of the decomposition method was suggested in [20, 21], in which a set of canonical forms is prepared beforehand, all different but equivalent to the given one. They have various basics specified by some maximal linearly independent subsets of columns of matrix A, selected *at random*, independently of each other. In such a way the process of looking for a shortest solution is randomized. The number q of used canonical forms could be arbitrary, being chosen by some additional considerations.

A solution is searched in parallel over all these forms, first at the level 0 of exhaustive search, then at the level 1, etc., until at the current level k a solution with weight w, satisfying condition $w < \gamma - 1$ will be found. With raising q this level k can be reduced, which reduces the run-time as well.

Suppose there exists a solution with weight w. The chances to detect it at level k of exhaustive search can be estimated as follows. Consider an n-component Boolean vector $\boldsymbol{a}$, with a randomly selected $(n - m)$-component sub-vector $\boldsymbol{a}'$. There exist C_n^w (the number of different combinations from n by w) values of vector $\boldsymbol{a}$ each of which has exactly w ones. Let us assume that all of them are equiprobable. The number of those of them, which have exactly k ones in vector $\boldsymbol{a'}$ ($k \leq n - m$ by that), is evaluated by the formula $C_{n-m}^{k}C_{m}^{w-k}$, and the number N_k of those which have no more than k ones in vector $\boldsymbol{a'}$ is evaluated by the formula

$$N_k = \sum_{j=0}^{k} C_{n-m}^{j}\, C_{m}^{w-j}$$

Evidently,

$$C_n^w = \sum_{i=0}^{\min(w,\, n)} C_{n-m}^{i} C_{m}^{w-i}$$

and the formula

$$P = 100\, N_k / C_n^w$$

shows the percentage of situations wgere a short solution with weight w can be found at the level of search k.

For example, in Table 4 are shown the calculated values received by P for different levels of search k at $m = 420$, $n = 500$ and $w = 75$.

The following conclusion could be deduced from that table. Preparing beforehand $q = 100$ random canonical forms of the considered SLLE with given parameters, we could hope to find the solution on level 5 or 6. In that case about 25 or 300 million combinations should be checked for every of 100 canonical forms.

Table 4. Evaluation of chances to detect a shortest solution at given level of search k

($m = 420$, $n = 500$ and $w = 75$).

k	1	2	3	4	5	6	7	8	9	10	11	12	13
P	0.00	0.01	0.06	0.23	0.85	2.41	5.62	11.27	19.84	23.12	36.36	50.00	62.34

Programming and experiments. The suggested randomized parallel algorithm was programmed and verified (C++, PC COMPAC Presario – processor Intel Pentium III, 1000 MH). The dependence of the run-time T on the number q of randomly selected canonical forms was investigated for different systems of linear equations. Some results are presented below.

Thirty different random SLLEs with $m = 900$ and $n = 1000$ were generated (each having a solution with weight $w = 100$) and solved, using q randomly chosen canonical forms for every system, for different q: 1, 10, 30 and 300. The following parameters of the solution process were measured and shown in Table 5:

N – the ordinal number of the solved SLLE, L – the level of search at which the solution was found, F – the number of canonical form where the solution was found, T – the time spent for finding the solution (measured in seconds (s), minutes (m), hours (h), days (d) and years (y).

For instance, the short solution with $w = 100$ was found for SLLE number 22 in 6 minutes, at the level of search 3 while solving the canonical form number 190.

Note that the last parameter T was found not immediately but forecasted according to the method described in [22], which changes the real solution process for a virtual one. That saves much time spent for the experiment.

The positive result of increasing the number q of canonical forms is evident: at $q = 300$ every of the considered 30 examples is solved in several minutes – instead of many (thousands sometimes) years at $q = 1$.

Solving Over-defined Systems

It can appear, that the regarded SLLE has no root – when it is over-defined (inconsistent, or contradictory, when usually *m* > *n*). In that case it is possible to put the task of finding a value of vector x, fitting to maximum number of the equations and accepted therefore as a solution of the system. Such task arises at development of information security systems and can be interpreted as follows. Suppose, the appropriate value of vector y is received for given A and x, and then distorted (in components marked by ones in the vector of distortions e). As a result a vector $z = y \oplus e$ appears, whereas x and y are "forgotten". It is required to restore the initial situation on known now values A and z.

This task was solved in [23], where it was reduced to finding a shortest root of an undefined SLLE obtained from the initial over-defined SLLE by appropriate transformation of matrix A and vector y. The boundaries of correct setting of the task were defined in [16], within which the values of x and y can be restored practically uniquely. A new method of solution of the given task was offered in [24], based on compact representation of the processed information and usage of the procedure of random sampling. The given over-defined SLLE is converted by that to other over-defined SLLE equivalent to initial one but solved more easily.

Table 5. The results of solving undefined SLLEs with parameters *n* = 1000, *m* = 900, *w* = 100.

	q=1			*q*=10			*q*=30			*q*=300		
N	L	F	T	L	F	T	L	F	T	L	F	T
1	1	8	7d	9	6	16h	9	6	18h	33	4	19m
2	1	9	69d	8	6	14h	16	5	2h	254	2	4m
3	1	10	2y	7	7	7d	28	6	2d	234	2	4m
4	1	13	776y	3	6	5h	3	6	8h	117	3	5m
5	1	3	4s	1	3	4s	1	3	5s	1	3	4m
6	1	10	2y	5	7	5d	26	5	3h	56	2	4m
7	1	12	112y	9	4	3m	9	4	3m	57	3	5m
8	1	8	7d	10	5	1h	10	5	1h	106	3	5m
9	1	12	112y	10	5	1h	28	4	10m	141	3	6m
10	1	9	69d	2	6	4h	2	6	6h	95	2	4m
11	1	13	776y	7	8	82d	15	4	5m	50	2	4m
12	1	13	776y	9	8	104d	14	5	2h	117	3	5m
13	1	10	2y	8	6	14h	8	6	16h	35	3	4m
14	1	6	58m	1	6	2h	1	6	4h	39	3	4m
15	1	6	58m	1	6	2h	11	5	1h	134	3	6m
16	1	14	4954y	2	8	27d	28	4	10m	205	3	7m
17	1	6	58m	1	6	2h	14	5	2h	49	3	4m
18	1	10	2y	2	7	2d	27	5	3h	285	2	4m
19	1	13	776y	8	7	8d	8	7	9d	84	3	5m
20	1	10	2y	2	6	4h	2	6	6h	203	4	1,3h
21	1	8	7d	7	5	45m	7	5	52m	93	4	39m
22	1	7	13h	10	6	17h	27	4	9m	190	3	6m
23	1	12	112y	2	5	13m	16	2	4s	16	2	4m
24	1	15	29185y	4	7	4d	24	6	2d	226	2	4m
25	1	10	2y	2	6	4h	2	6	6h	46	3	4m
26	1	13	776y	9	7	9d	16	5	2h	112	4	45m
27	1	10	2y	6	8	71d	16	4	6m	281	3	8m
28	1	12	112y	7	8	82d	18	6	1d	87	3	5m
29	1	12	112y	6	4	2m	6	4	2m	254	2	4m
30	1	12	112y	7	8	82d	22	6	2d	31	4	18m
Sum:			38704y			1,3y			20d			5,3h

Bibliography

1. Zakrevskij A., Zakrevski L. Solving systems of logical equations using search tree minimization technique. – Proceedings of the PDPTA'02 International Conference, June 24-27, 2002, Las Vegas, USA. – pp. 1145-1150.
2. Zakrevskij A., Vasilkova I. Reducing search trees to accelerate solving large systems of Boolean equations. – Boolean Problems // 5-th International Workshop, Sept. 19-20, 2002, Freiberg (Sachsen). – pp. 71-76.
3. Zakrevskij A. Reduction algorithms for solving large systems of logical equations. – Computer Science Journal of Moldova, 2000, v. 8, No 1. – pp. 3-15.
4. Zakrevskij A.D. Solving systems of logical equations by the method of local reduction. – Doklady NAN B, 1999, v. 43, No 5, pp. 5-8. (in Russian).
5. Baumann M., Rohde R., Barthel R. Cryptanalysis of the Hagelin M-209 Machine. – 3rd International Workshop on Boolean Problems, Sept. 17-18, 1998, Freiberg (Sachsen), pp. 109-116.
6. Zakrevskij A.D., Vasilkova I.V. Cryptanalysis of the Hagelin machine by the method of spreading of constants. – Proceedings of the Third International Conference on Computer-Aided Design of Discrete Devices (CAD DD'99), Minsk, November 10-12, 1999, vol. 1. – pp. 140-147.
7. Zakrevskij A.D. Solving large systems of logical equations by syllogisms. – Doklady NAN B, 2000, v. 44, No 3, pp. 40-42 (in Russian).
8. Zakrevskij A.D. To formalization of polysyllogistic. – Logical Inference, Moscow: Nauka, 1979, pp.300-309 (in Russian).
9. Lukasiewich J. Aristotle syllogistic from the point of view of modern formal logic. – Moscow, 1959 (in Russian).
10. Zakrevskij A., Vasilkova I. Reducing large systems of Boolean equations. – 4-th International Workshop on Boolean Problems, September 21-22, 2000, Freiberg, Germany. – pp. 21-28.
11. Zakrevskij A. D. Solving large systems of logical equations. – Sixth ISTC Scientific Advisory Committee Seminar "Science and Computing", Moscow, Russia, 15-17 September 2003. – Proceedings, volume 2, pp. 528-533.
12. Zakrevskij A.D, Toropov N.R. Generators of pseudo-random logical-combinatorial objects in C++. - Logical Design, No 4, 1999, Minsk, Institute of Engineering Cybernetics, pp. 49-63 (in Russian).
13. Gill A. Linear sequential circuits. McGraw-Hill Book Co., New York, 1966.
14. Zakrevskij A.D., Toropov N.R. Polynomial implementation of partial Boolean functions and systems. – Moscow, URSS, 2003 (in Russian).
15. Balakin G.V. Introduction into the theory of random systems of equations. – Proceedings on discrete mathematics, Moscow, TVP, 1997, vol. 1, pp. 1-18 (in Russian).
16. Zakrevskij A. Solution of a system of linear logical equations with distorted right members – when it could be found. – New Information Technologies. Proceedings of the Fifth International Conference NITe'2002, Minsk, BSEU. – Vol. 1, pp. 54-58.
17. Zakrevskij A. D., Zakrevski L. Optimizing solutions in a linear Boolean space – a decomposition method // Proc. of STI '2003, Orlando, Florida, USA, July 2003, pp. 276-280.
18. Gauss C.F. Beitrage zur Theorie der algebraischen Gleichungen. – Gött., 1849
19. Zakrevskij A.D. Looking for shortest solutions of systems of linear logical equations: theory and applications in logic design. – 2. Workshop "Boolesche Probleme", 19./20. September 1996, Freiberg/Sachsen, pp. 63-69.
20. Zakrevskij A.D. Randomization of a parallel algorithm for solving undefined systems of linear logical equations. – Proceedings of the International Workshop on Discrete-Event System Design – DESDes'04. – University of Zielona Gora Press, Poland, 2004, pp. 97-102.
21. Zakrevskij A.D. Raising efficiency of combinatorial algorithms by randomized parallelization. – XI-th International Conference "Knowledge-Dialogue-Solution" KDS – 2005, June 20-30, 2005, Varna, Bulgaria, pp. 491-496.
22. Zakrevskij A.D., Vasilkova I.V. Forecasting the run-time of combinatorial algorithms implementation. – Methods of logical design, 2003, issue 2. Minsk: UIIP of NAS of Belarus, pp. 26-32 (in Russian).
23. Zakrevskij A.D. Solving inconsistent systems of linear logical equations. – 6-th International Workshop on Boolean Problems, September 23-24, 2004, Freiberg (Sachsen), pp. 183-190.
24. Zakrevskij A.D. A new algorithm to solve overdefined systems of linear logical equations. – Computer-Aided Design of Discrete Devices (CAD DD'04). Proceedings of the Fifth International Conference, 16-17 November 2004, Minsk, vol. 1, pp. 154-161.

Authors' Information

Arkadij Zakrevskij - *United Institute of Informatics Problems of the NAS of Belarus, Surganov Str. 6, 220012 Minsk, Belarus; e-mail: zakrevskij@tut.by*

ADAPTIVE APPROACH TO STATIC CORRECTION WITH RESPECT TO APRIORISTIC INFORMATION IN PROBLEM OF SEISMIC DATA PROCESSING

Tatyana Stupina

Abstract: *In that paper problems that arising in seismic data processing for real time are considered. The main issue is defects in investigation of mathematical and physical model of object. The static and kinematic residual correction problem is relevant for real time in spite of the modern graph of seismic data processing contains such procedures. The quality of final results that presented by seismic and depth-distance plot doesn't answer to modern requirements. Three adaptive approaches to realization of numerical method for regression model of static and kinematic residual correction problem are considered. Its use a priory information about number of supporting points of seismic profile at the most.*

Keywords: *the model of residual time of reflected waves, kinematic and static corrections, number of supporting points of seismic profile, line regression model, measure of a priori information, functional of quality, adaptive algorithm, the condition of stability.*

ACM Classification Keywords: *G1.10 Numerical Analysis - Applications, G3 Probability and statistics - Correlation and regression analysis.*

Introduction

The results of the seismic data processing depend on the correspondence of experimental (survey notes) data to theoretical model of environment under investigation. That variance usually concerns with variance of time of reflected waves and it is conditioned by two main reasons. The first is the heterogeneities of upper part of formation (static corrections). The second is difference of given unequal distance between source point and receiving point (kinematic corrections). The modern geophysical software contains enough effective algorithms for static and kinematic corrections, but usually it is founded on supposition of vertical waves transmission from datum to original ground [V.S. Kozirev]. Such supposition doesn't give us expected or satisfactory result [A.P. Sisoev], because it doesn't consider full information about heterogeneities of upper part of formation, that we can be described as:

1. The weathering zone,
2. The relief of original ground,
3. The deep heterogeneities.

The using of the most total information about environment also gives difficult. It is complexity of mathematical model. In addition we have to construct a stability algorithm. Thus that is fundamental problem of numerical mathematics. This paper considers and analyses a line regression model of the static and kinematic corrections for fixed observing system. The measure of a priory information is defined as number of supporting points of seismic profile. The adaptive approach is applied to realization numerical algorithm for increase of stability.

The Problem of Static Correction

The complication of mathematical model, as a rule, leads to additional expenses not only in respect of algorithmic and computing realization of a method, but also to a problem of presence of the additional information on studied object. Therefore now at the first stage as the research tool the linear tetra-factorial model of residual times of reflexions for the single-layered environment is considered, which with reference to CDP arrival-time curve looks like [V.S. Kozirev]:

$$t_{ij} = S_i + R_j + G_k + M_k x_{ij}^2 + N_{ij} \tag{1}$$

where t - double time of reflexion after introduction of primary aprioristic static and kinematic amendments, *i*, *j*, *k* - indexes of position of a source, the receiver and the common midpoints (CMP) accordingly, *G* - structural a component (a double transit time of a wave from level of reduction to reflecting border), *S* - superficially co-ordinated static amendment for a source, *R* - the same for the receiver, *M* - factor of the residual kinematic amendment, x - source-receiver offset, *N* - a noise component.

Model (1) turns out from the equation concerning times T_{ij} of arrival of the reflected wave:

$$T(x,l) = a(x-l) + b(x+l) + \sqrt{t_0^2(x) + 4l^2 / v^2(x)} \tag{2}$$

describing CMP arrival-time curve for single-layered model of the environment, counted for a reduction line in system of co-ordinates (x,l): x - co-ordinate of CMP, l – half of the source-receiver offset. Thanks to possibility of reception and the account of the aprioristic information which are carried out by input of aprioristic model of arrival-time curve $T_a(x,l)$ with parameters $a_a(x-l)$, $b_a(x+l)$, t_{0a}, v_a, and having made the linearization procedure of models (2) (expansion in Taylor's series of function $\sqrt{(\bullet)}$ about a point 0), we will receive:

$$\tau(x,l) = T(x,l) - a_a(x-l) - b_a(x+l) - \sqrt{t_{0a}^2(x) + 4l^2 / v_a^2(x)} \approx$$

$$\approx \ s(x-l) + r(x+l) + g(x) + m(x)l^2 \tag{3}$$

the model defining residual time shifts $\tau(x,l) = \tau_{ij}$ in the form of (1).

Results of the spent correction at the made aprioristic assumptions of the arrival-time curve form are times of reflexions

$$T(x,l) = T_a(x,l) + \tau(x,l) \tag{4}$$

In the operational form the equation (1) registers as

$$Ap = \tau^* + \delta\tau \tag{5}$$

Matrix A - strongly rarefied, consisting from 0, 1 and x_{ij}^2; $p = (S_i, R_j, G_k, M_k)^T$ - a vector of estimated factors, $p \in D_p$ (space of admissible decisions); τ^* - a vector of true supervision; $\delta\tau$ - a vector of hindrances.

Taking in account the big dimension of entrance parameters and presence between them certain communications the system of the equations is badly caused, and decisions - unstable [V.S. Kozirev, A.P. Sisoev]]. Solving out of such systems the various methods of regularization are applied, involve the additional information, simplify models to smaller number of parameters, narrowing a class of decision functions. However methods of regularization are not always effective in reception of satisfactory result in a case when it is necessary to consider aprioristic, presented by quantity of reference points, information in the course of reception of the decision [A.P. Sisoev].

In a considered problem the strong rarity of an operational matrix allows to write out iterative formulas of reception Least-Squares Method - estimations concerning criterion F_1 (item 4) for model (1) not resorting to direct transposing and the reference of matrixes of the big sizes:

$$G_k = \tfrac{1}{n_k}\sum_{(k)}^{n_k}(\tau_{ij} - S_i - R_j - M_k x_{ij}^2)\,,$$

$$S_i = \tfrac{1}{n_i}\sum_{(i)}^{n_i}(\tau_{ij} - G_k - R_j - M_k x_{ij}^2)\,,$$

$$R_j = \tfrac{1}{n_j}\sum_{(j)}^{n_j}(\tau_{ij} - S_i - G_k - M_k x_{ij}^2)\,,$$

$$M_k = \frac{1}{\sum\limits_{(k)}^{n_k} x_{ij}^2}\sum_{(k)}^{n_k}(\tau_{ij} - S_i - R_j - G_k) \tag{6}$$

where n_k , n_i , n_j means multiplicity with which k'th CMP, i'th point of excitation and j'th point of reception are present at supervision, summation is conducted on traces in which they participate. From the presented dependences (3) it is visible, that initial approach is possible on any of factors S_i , R_j , G_k , M_k , hence the result of the solution will depend on a choice of initial approach. Moreover, minimization of criteria in general case of errors correlation with the unknown law of distribution and in case of stochastic of results of measurements (not planned experiment or experiment with errors in measurements) does not guarantee the stable solution.

Let's show on an example, that minimization of only functional F_1 not always can lead to desirable result. We will designate through g^* , f^* (we will drop parameters x and l) true, in essence unknown functions of arrival-time curve and "relief", through g_a , f_a - aprioristic functions of arrival-time curve and "relief", through $\tilde{g}$, $\tilde{f}$ - kinematic and static amendments. Then if we will present $A = (A_1, A_2)$, $C_1 = \|A_1\|$, $C_2 = \|A_2\|$, a rough estimate from above

$$\begin{aligned} &\|\tau - \tilde{\tau}\| = \|Ap - A\tilde{p}\| = \left\|A(f^* - f_a, g^* - g_a) - A(\tilde{f}, \tilde{g})\right\| \\ &\le C_1(\left\|f^* - f_a\right\| + \|\tilde{f}\|) + C_2(\left\|g^* - g_a\right\| + \|\tilde{g}\|) = \varepsilon \end{aligned} \tag{7}$$

shows, that to one value of criterion there can correspond various values of the items entering into the sum. Even under condition of a priori known model of environment (the first items in brackets it is possible to name ineradicable errors of model) kinematic and static amendments can be mutually replaced. From here there was a research problem of algorithm of correction of a static at an iterative stage of minimization of discrepancy F_1 with the account of the aprioristic information.

The Aprioristic Information

At work with real (field) data the aprioristic information are:

1. Assumptions about agent and laws of distribution of waves (for the decision of a direct problem). For example, in a considered case parameters t_{0a}, v_a for the equation of CDP arrival-time curve for the reflected single-layered environment.
2. Quantity of the reference points necessary for interpolation of function of "relief" or heterogeneity of agent, we will designate through μ. We will notice, that generally algorithms of interpolation are unstable, relief function actually is discretely multiextremal, and assumptions of a "good" class of functions at times are far from practice.

Therefore in the researches we will stop on studying three interdependent characteristics: a frequency component of "relief", number of reference points, length of arrangement of a seismic profile.

Construction of the Multiobjective Decision-making

Pursuing an ultimate goal - reception of the stable solution, on a basis multiobjective decision-making we will construct an optimum subset of decisions $optD_p \in D_p$.

Let's consider the criteria to which optimum values should satisfy required decision $\tilde{p}$:

1. $F_1(\tilde{p}) = \|\tau - \tilde{\tau}\|_{L^2} \to \min\limits_{p \in D_p}$ - minimization of discrepancy by, for example, an iterative method, where $A\tilde{p} = \tilde{\tau}$;

2. $F_2(\tilde{p}) = \|p - \tilde{p}\|_{L^2} \to \min\limits_{p \in D_a}$ - minimization of an error of the decision on reference points, D_a - set of reference points on which heterogeneity of upper part of formation is a priori restored, $|D_a| = \mu$;

3. $F_3(\tilde{p}) = \|p^* - \tilde{p}\|_{L^2} \to \min\limits_{p \in D_p}$ - minimization of an error of the decision on synthetic data, where $Ap^* = \tau^*$;

4. $F_4(\tilde{p}) = |D_a| \to \min$ - minimization of number of reference points, use a big number of which is connected with essential economic expenses;

Listed above functions forms vector criterion $F(p) = (F_1(p), F_2(p), F_3(p), F_4(p)) \in \Re^4$. There are some variants of solving of such problems, of which it is possible to allocate with most widespread:

a) The Pareto construction of optimum set of decisions [V.V. Podinovskij];

b) Ranging of criteria or introduction of relative importance of criteria;

c) Construction of global criterion or secularization of vector criterion.

First two variants mean presence of the aprioristic information, possibly even expert knowledge of admissible value of norm of discrepancy, whereas the third is based on more automated decision-making. For investigated model we will in detail analyze synthesis of the second and third variant.

Let on n -th iteration of the equations (6), $n = 1, 2, 3...$, solution $\tilde{p}^{(n)} = (\tilde{f}^{(n)}, \tilde{g}^{(n)})^T$ of system $\tilde{\tau}^{(n)} = A\tilde{p}^{(n)} = (A_1\tilde{f}^{(n)}, A_2\tilde{g}^{(n)})^T$ is received, with value of discrepancy equal $\varepsilon^{(n)} = \left\|\tau - \tilde{\tau}^{(n)}\right\|$. We will designate through $\tilde{D}_p = \left\{\tilde{p}^{(n)} : \delta_{\min} \le \left\|\tau - A\tilde{p}^{(n)}\right\| \le \delta_{\max}, n_{\min} \le n \le n_{\max}\right\}$ set of "trial" solutions, i.e. set of decisions which within several iterations satisfy to admissible quality on norm of discrepancy F_1 . We will formulate three approaches to decision construction:

1. The first approach.

We build solution $\tilde{p} = \sum_n \lambda_n \tilde{p}^{(n)}$ of elements of set $\tilde{D}_p$, where $\sum_n \lambda_n = 1$ and weight number $\{\lambda_n\}$ ranked with the maximum value for minimum norm $\left\|\tilde{f}^{(n)}\right\| = \min\limits_{\tilde{p} \in \tilde{D}_p \cap D_a} F_2(\tilde{p})$.

2. The second approach.

Consistently we correct solution $\tilde{p}^{(n)}$ to $p^{(n)}$, $p^{(n)} = \tilde{p}^{(n)} + p_1$, on reference points remaining in "trial" set, i.e. $\left\|\tau - \tau^{(n)}\right\| = \varepsilon^{(n)} \le \delta_{\max}$ and $\left\|f^{(n)}\right\| \le \left\|\tilde{f}^{(n)}\right\|$, where $p_1 = (f^{(n)} - \tilde{f}^{(n)}, 0)$.

3. The third approach.

$F = \lambda_1 \left\|\tau - \tilde{\tau}^{(n)}\right\| + \lambda_2 \left\|\tilde{f}^{(n)}\right\| \to \min\limits_{\tilde{p} \in \tilde{D}_p \cap D_a}$, where $\lambda_1 + \lambda_2 = 1$.

Condition $\lambda_2 > \lambda_1$ means stronger requirements to performance of aprioristic conditions.

The Description of a Modeling Example

The synthetic data simulated for the single-layered formation and flank system of supervision, is schematically shown on fig. 1. Model parameters:

Velocity in weathering zone, V_1 = 600 m/s;

1. Velocity in the agent to the first reflecting horizon, V_0 = 1950 m/s;
2. The equation of "relief" $y(x) = \frac{1}{V_1}(c_1 + c_2 \sin(\omega\, x + \varphi))$, $\omega = \frac{2\pi}{T}$ - a frequency component of "relief", T = 200-1900, $c_1 = 20, c_2 = 10$.
3. Aprioristic time of reflecting horizon, t_0 = 1.5 s.
4. The equation of CDP arrival-time curve $t(x) = \sqrt{t_o^2 + x^2/V_0^2}$
5. The system of supervision is flank with repeated overlapping: length of a profile of 10 000 m., quantity of channels 20, distances ΔRP = ΔSP = 50 m, max |RP-SP| = 1.3H, H = 1.5V$_0$/2 = 1462.5, length of arrangement 20*50-50=950, number of reference points from 10 to 100, RP – point of receiving and SP – point of source ;
6. The noise level is set by value mean-square σ (further on schedules it is signed «sigma») regulary distributed in the range of $\left[-\frac{l}{2}, \frac{l}{2}\right]$, $l = \sqrt{12}\sigma$, a random variable with a zero average of distribution.

The software procedure of calculation of times of arrival of the reflected wave (a direct problem) is realized. The inverse problem was solved by iterative Gauss-Zejdel method; on each step of it values of criteria F_1, F_2 and F_3, were estimated at fixed value F_4. The set of "trial" decisions was in such a way formed. The received results will allow to formulate recommendations to construction of global criterion $F = \lambda_1 F_1 + \lambda_2 F_3$, and then pass to Pareto construction of optimum set on (F, F_4). Researches of numerical modeling were spent for differently periodical functions of heterogeneity (item 3. of this paragraph), admissible noise levels (item 7) in the right part of the equation (5) and different quantity of reference points.

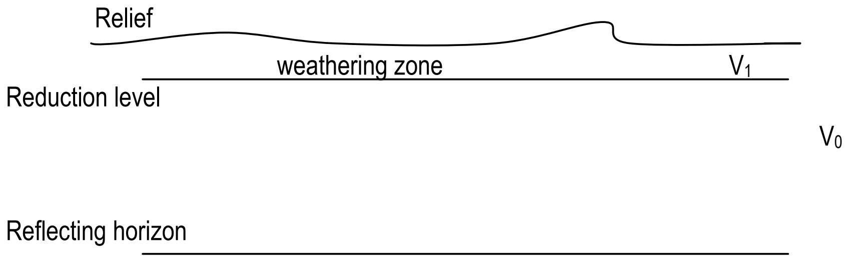


fig. 1. The scheme of single-layered formation

Algorithm Research

The first stage of program testing consisted in check of a program code of algorithm on correctness of realization, i.e. in research on convergence of iterative process (value of criterion F_1) at zero noise level, σ = 0 and various values of quantity of reference points (value of criterion F_4). On fig. 2 efficiency of algorithm is shown. The tendency of increase in speed of convergence and accuracy increase is obvious at bigger number of reference points. On all schedules the norm square in metrics L^2 is considered.

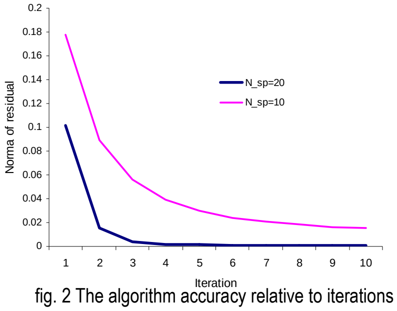


fig. 2 The algorithm accuracy relative to iterations

At the second stage for each iteration the values of criterion F_2, were numerically received. It was necessary to do in order to estimate on an example the relative contribution of each of criteria to the general (scalarized) on the third approach. Results on fig. 2 and fig. 3 show necessity of introduction of the scaling factor for criterion F_2. In the presented example it is equal approximately to one decimal order for number of reference points equal 10. It is clear, that this size will depend on aprioristic knowledge of noise level. For example, at known small enough noise level the value of discrepancy and also value of norm on reference points decreases with increase in quantity of iterations that allows to make the decision already on the fourth iteration.

On synthetic data there is a possibility to receive the error of the solution presented in the form of norm (criterion F_3). The schedule 4 reflects essential reduction of an error in drawing on bigger number of reference points and enough quick stabilization of value of an error (on 3-4 iterations). At comparison of schedules on fig. 3 and fig. 4 a question of possibility of an alternative estimation of quality of the solution arises, i.e. a norm estimation on reference points as on control sample. Passing to statistical estimations $\sqrt{\frac{F_2}{2\dim D_a}}$ and $\sqrt{\frac{F_3}{\dim D_p}}$, mean-square, on an example for 10th iteration it is possible to see their small difference $\sqrt{25\cdot 10^{-5}}$ and $\sqrt{8\cdot 10^{-5}}$.

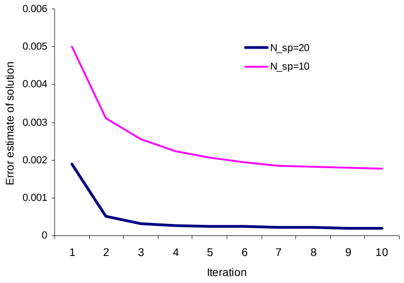


fig. 3 The Error estimate of solution on support points, Var[eps]=0.

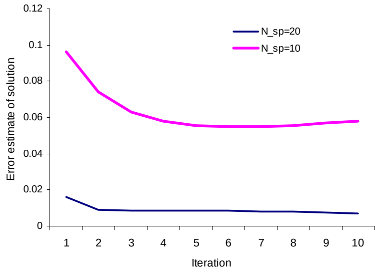


fig. 4 The error estimate of solution, Var[eps]=0.

At the third stage noise of various levels has been entered into model. The schedule in drawing 5 reflects polynomial dependence of accuracy of iterative algorithm in 10th iteration depending on noise level.

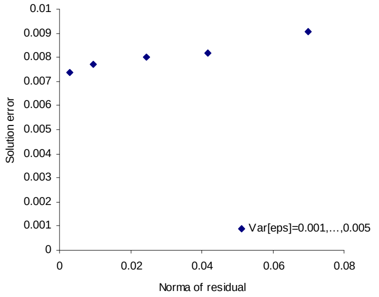


fig. 5 Analysis of stability domains by error on support points, Var[eps]=0.001,…,0.005

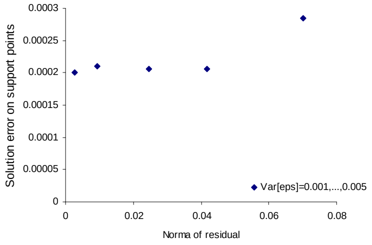


fig. 6 Analysis of stability domains by error on solutions, Var[eps]=0.001,…,0.005

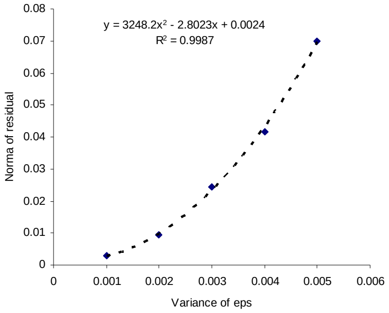


fig. 7 The algorithm accuracy relative to noise level, N_sp=20

Analyzing results of modeling two-dimensional criteria area (F_1 , F_2) on fig. 6 and area (F_1 , F_3) on fig. 7 it is possible to make a conclusion on essential joint increase in values of criteria at noise level bigger than 0.004.

At the fourth stage the influence of «relief form» on convergence and accuracy of the decision was numerically estimated. It was considered while one parameter presented in item 5, sub item 3, frequency component T. In all previous examples T it was equal 700, which make 0.73 from length of arrangement. Results on schedules fig. 8, 9, 10, and 11 illustrate an existing problem of reception of the stable solution at value of T (equal 200, 235) less than a quarter of length of arrangement. The schedule on fig. 8 shows fast stabilization of process on small enough value of discrepancy. For T = 200 iterative process already from the third iteration starts to disperse slowly. Results on fig. 10 and 11 show change of accuracy of the decision and norm of discrepancy at various quantity of reference points. Some law of division of schedules on decimal parity and oddness of quantity of reference points is observed. Probably, that it is connected with uniformity of distribution of reference points on all length of a profile.

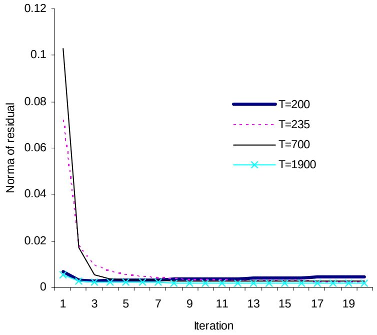


fig. 8 The algorithm accuracy relative to iterations for different wave period and N_sp = 20, Var[eps] = 0.001

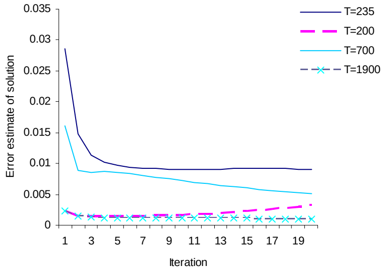


fig. 9 The error estimate of solution relative to iterations for different wave period and N_sp = 20, Var[eps] = 0.001

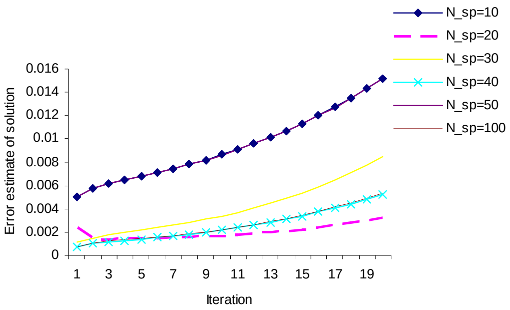


fig. 10 The error estimate of solution relative to iterations for different number of support points and T = 200, Var[eps] = 0.001

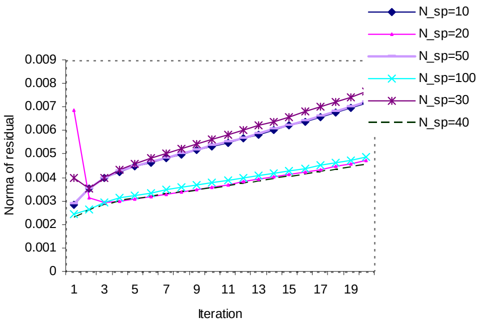


fig. 11 The norma of residual relative to iterations for different number of support point, T = 200 and Var[eps] = 0.001

Basic Results of Researches

With reference to investigated model it is possible to draw following conclusions:

- Fast enough convergence of iterative algorithm at zero noise and a maximum quantity of reference points.
- At mean square noise level smaller than 0.004 it is possible to construct the solution according to the offered approaches. Thus the admissible set of solutions can start to be formed from the third iteration.
- It is possible to draw a conclusion that only at careful enough analysis of aprioristic data about the model and "relief" model (heterogeneity) and the complex approach to estimation of quality of the made decision, it is possible to receive rather stable solutions of a problem of correction of static amendments.

For ours investigation we have created programming instrument. It is algorithm for numerical research, that was realized in the programming language C++ in the environment of Microsoft Visual C ++. The basic program mainframes:

1. Data input;
2. Formation of a matrix of system of supervision with record in a format in a file;
3. Functions of construction of heterogeneity of agent, CDP arrival-time curve, noise.
4. Iterative algorithm of the solution of the inverse problem with a conclusion of numerical values by criteria.

Conclusion

The linear model of residual times of the reflections, considering static and kinematic amendments for a relief with the account of the aprioristic information on the model is presented in the present work. It has been noticed, that such problems remain actual in connection with the big conditionality of their solutions. Four adaptive approaches to decision-making is offered, i.e. as much as possible considering the aprioristic information presented by quantity of reference points on a seismic profile. The offered approaches are based on introduction combined functionals of model quality and decision updating at a stage of iterative formation of the solution.

For the homogeneous single-layered isotropic environment of the set power, the fixed system of supervision of certain length, numerical modeling by the technique presented in work is spent. The results reflecting convergence of iterative process of the solution of inverse problem are received (definition of parameters of correction of statics without kinematics distortion), possibility of application of four offered approaches to decision-making on set of admissible decisions is shown.

In the conclusion we will notice, that the considered problem is “inverse incorrectly set” and settles by construction of « quasi-solution». The decision strongly enough depends on quality of initial data and a class of functions in which the decision is under construction. Except application of the mathematical apparatus used for an accurate substantiation of used methods, any practical problem always demands an individual approach to its solution, not looking on the standard approaches.

The author of work had practical experience in solving inverse problems also in other fields of knowledge (medicine, hydrology, ecology). In general, it would be desirable to notice, that the expert information strongly enough narrows space of possible decisions, less difficult models more often gives the comprehensible solution. Very often behind consideration frameworks there is a question on possibility of application of model to observable data in problems of the big dimension and complexity of model.

Bibliography

[V.S. Kozirev, 2003] V.S. Kozirev, A.P. Zjukov, I.P. Korotkov, A.A. Zjukov, M.B. Shneerson. Taking into account of heterogeneousness of upper part of formation in exploration seismology. Modern technologies, Moscow: “Nedra-Busness center”, 2003, 227 pp.

[A.P. Sisoev, 1988] A.P. Sisoev. Analyze of stability of evaluation static and kinematic parameters in reflection method. Mathematical problems of interpretation data of exploration seismology. Novosibirsk, Nauka, 1988.

[L. Hatton, 1989] L. Hatton, M. Uerdington, Dj. Mejkin. Processing seismic data. Theory and practice. Moscow: Mir, 1989, 216 pp.

[V.M. Glogovskij, 1984] V.M. Glogovskij, A.R. Hachatryn. Static correction in its true value kinematic parameters of reflection. Geology and Geophysics. 1984, №10.

[V.V. Podinovskij, 1982] V.V. Podinovskij, V.D. Nogin. Pareto-optimal decision of multicriterion problem. 1982, 256 pp.

Authors' Information

Tatyana A. Stupina *– Trofimuk Institute of Petroleum-Gas Geology and Geophysics of SBRAS, Koptyug Ave 3, Novosibirsk, 630090, Russia; e-mail: stupinata@ipgg.nsc.ru*

METHOD AND ALGORITHM FOR MINIMIZATION OF PROBABILISTIC AUTOMATA

Olga Siedlecka

***Abstract**: The theory of probabilistic automata is still evolving discipline of theory of information. As in classical theory of automata, it might be a base for computations, can be exploited in design and verification of circuits and algorithms, in lexical analyzers in compilers computers in future. Minimization of any type of automata gives always saving in resources and time, and is important problem that has been analyzed for almost sixty years. Different types of automata are used for modeling systems or machines with finite number of states.*

In this article we show few specific type of probabilistic automata, especially the reactive probabilistic finite automata with accepting states (in brief the reactive probabilistic automata), and definitions of languages accepted by it. We present definition of bisimulation relation for automata's states and define relation of indistinguishableness of automata states, on base of which we could effectuate automata minimization. Next we present detailed algorithm reactive probabilistic automata's minimization with determination of its complexity and analyse example solved with help of this algorithm.

***Keywords**: minimization algorithm, reactive probabilistic automata, equivalence of states of automata, bisimulation relation.*

***ACM Classification Keywords**: F. Theory of Computation, F.1 Computation by Abstract Devices, F.1.1 Models of Computation, Automata; F.4 Mathematical logic and formal languages, F.4.3 Formal Languages*

Introduction

The automata theory is older than any physical computer, after defining abstract machines like Turing machine, scientist searched for equally simple model that resolve problems that doesn't need to write symbols, but only read - they created automata. Like in Turing machine occurred many types of this model: deterministic, nondeterministic, finite, probabilistic, and many others. They could be used for simulation of circuits, algorithms, and every system that have states and read symbols, or react on some action. If we have states as a simulation of real resources, it is welcomed to narrow down their number.

The problem of finite automata minimization appeared in the end of fifties of last century and its main point is to find automata with the minimum number of states accepting the same language as input automata. During last fifty years many algorithms for minimization of finite deterministic automata came into existence, most of which (except Brzozowsky algorithm which is based on derivatives [Brzozowski, 1962]), is based on equivalence of states. One of the most popular minimization algorithms is Hopcroft and Ullman's algorithm with running time $O(|\Sigma|n^2)$ (where $|\Sigma|$ is the number of symbols in the alphabet, n is the number of states) [Hopcroft, 2000]. Another algorithm with the same time complexity, but better memory complexity ($O(|\Sigma|n)$) is Aho-Sethi-Ullman's algorithm [Aho, 2006]. The most efficient deterministic finite automata minimization algorithm is Hopcroft's algorithm [Hopcroft, 1971] with time complexity $O(|\Sigma|n\log n)$.

In the same period of time scientists were searching for another models of computation. They developed probabilistic automata [Rabin, 1963], which are extensions of Markov chains with read symbols [Sokolova, 2004], models of finite automata over infinite words [Thomas, 1990], timed automata [Alur, 1994], hybrid automata [Henzinger, 1998] etc. We can find their ontological review in article: [Kryvyi, 2007]. In last few years new type of automata is researched by scientist - quantum automata, the probabilistic automata is intermediate way to understand them. It became important to find minimization algorithms for new types of automata. So far minimization of reactive probabilistic automata hasn't been described and it also is a step to minimize quantum automata.

Probabilistic Automata

A probabilistic automata, just like nondeterministic, has no consistently specified state, in which it will remain after reading symbol. But for probabilistic automata we have probability of reaching a state.

There exist many types of probabilistic automata which differ with properties, applications or probability distributions (continuous or discrete). Hereunder we itemize few of probabilistic automata's types with discrete probability distribution:

- reactive automata,
- generative automata,
- I\O automata,
- Vardi automata,
- alternating model of Hansson,
- Segala automata,
- bundle probabilistic automata,
- Pnueli-Zuck automata and others.

The algorithm showed in article was formulated for the reactive probabilistic automata.

A Markov chain is a transitive system, which has a probability of reaching state, but has no symbols to read, so it is the middle course to the probabilistic automata.

A **Markov chain** is a tuple *PA=(Q, δ)*, where

- *Q* is the finite set of states,
- *δ* is the transition probability function given by *δ:Q → D(Q)* (where *D(Q)* is the set of all discrete probability distribution on the set Q) [Sokolova, 2004] .

If *q* is a member of Q and *δ(q) = P* with *P(q') = p > 0,* then we say that Markov chain comes from state *q* to state *q'* with probability p (it may be written in many ways: *δ(q) = P(q)* or *δ(q)(q') = p*.

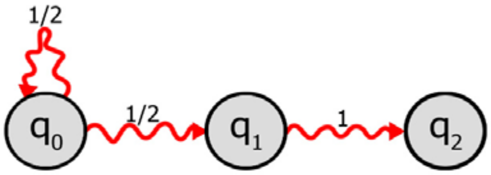


Fig.1. The Markov chain

The example of Markov chain is shown on figure 1, on which we can see probability of going out from state q_0.

The reactive probabilistic automata is a type of automata that react on reading symbol by going to another or the same state with given probability (sometimes we can interpret symbol as an action of simulated system).

A **reactive probabilistic automata** is a triple *PA=(Q, Σ, δ)*, where

- *Q* is the finite set of states,
- *Σ* is the finite set of input symbols (an alphabet),
- *δ* is the transition probability function given by *δ:Q × Σ → D(Q)* (where *D(Q)* is the set of all discrete probability distribution on the set Q) [Sokolova, 2004] .

An **initial reactive probabilistic automata** with accepting states is a five *PA=(Q, Σ, δ, q_0, F)*, in which we have additionally two elements:

- q_0 - a member of Q, is the start state,
- $F \subseteq Q$ is the set of final (accepting) states.

After reading given symbol automata is in state of **superposition** of states:

$$p_0q_0 + p_1q_1 + \ldots + p_nq_n,$$

where $p_0 + p_1 + \ldots + p_n=1$. Henceforth we will use shorter name of probabilistic automata within the meaning of initial reactive probabilistic automata with accepting states. An example of this type of automata we show on figure 2.

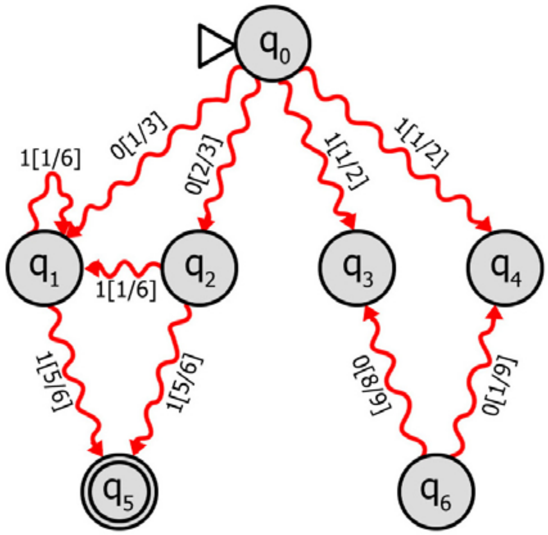


Fig.2. The initial reactive probabilistic automata with accepting states

Language Accepted by PA

Every type of automata is strictly connected with idea of language accepted by it. In deterministic and nondeterministic finite automata we say, that language is accepted by given automata if and only if for all words from this language, automata after reading of those words is always in its final state. In probabilistic automata we must also consider the probability of acceptance.

$$\delta(q_1, w\sigma) = \sum_{q\in Q} \delta(q_1, w)(q) \cdot \delta(q, \sigma).$$

The probability of going from state q_1 to state q_2 after reading symbol σ we denote as $\delta(q_1,\sigma)(q_2)=p$. An extended transition probability function, for given word *v* and prefix *w, so v = w σ,* denoted by the same notation, is given by [Cao, 2006]:

The **language accepted** by the probabilistic automata is defined as function:

$$L_{PA}:\Sigma^{*}\rightarrow[0,1],$$

such that [Cao, 2006]:

$$\forall w \in \Sigma^{*}, L_{PA}(w) = \sum_{q\in F} \delta(q_0, w)(q).$$

We say that language *L* is recognized with bounded error by a probabilistic automata *PA* with interval (p_1,p_1), if $p_1<p_2$ and

$$p_1 = \sup\{P_w|w\notin L\},\ p_1 = \inf\{P_w|w\in L\}$$ [Golovkins, 2002].

We say that language *L* is recognized with probability *p*, if the language is recognized with interval *(1-p,p)* [Golovkins, 2002].

We say that language *L* is recognized with probability $1\ {-}{-}\varepsilon$, if for every $\varepsilon >0$ there exist an automata which recognizes the language with interval $(\varepsilon_1,1{-}{-}\varepsilon_2)$, where $\varepsilon_1, \varepsilon_2 \leq\varepsilon$ [Golovkins, 2002].

Bisimulation and Indistinguishableness

When two automata accept the same language? When they possess equivalent states? Maybe one of them has smaller number of states and accepts the same language? These questions are very important for automata minimization problem. So, if we can find equivalent states, we can minimize some types of automata, but relevant relation is needed. One of the manners is to find first bisimulation relation and on the base of it define indistinguishableness of states.

Firstly we say that two deterministic finite automata are equivalent if they accept the same language, and two states are equivalent, if for every given word, reading this word after going out from these states always will finish for both states in accepting state or finish for both states in nonaccepting state. Automata is called minimal if all its states are nonequivalent.

For two deterministic finite automata: $DFA_1=(S, \Sigma, \delta)$ and $DFA_2=(T, \Sigma, \delta)$ *exists a* strong **bisimulation relation** $R\subseteq S\times T$ if for all $(s,t)\in R$ and for all $\sigma\in\Sigma$:

- if $\delta(s,\sigma)=s'$ then there exists $t'\in T$ such that $\delta(t,\sigma)=t'$ and $(s',t')\in R$ *and*
- if $\delta(t,\sigma)=t'$ then there exists $s'\in S$ such that $\delta(s,\sigma)=s'$ and $(s',t')\in R$ [Kozen, 1997].

The relation of strong bisimulation *R* has such properties as:

- a diagonal $\Delta_S \subseteq S \times S$ is bisimulation on *(S, δ),*
- an inverse relation R^{-1} is bisimulation,
- a sum of bisimulation relations is also bisimulation.

The equivalence relation *R* is a **congruence** on set of automata states for $(q_1,q_2)\in Q$ and symbols $\sigma \in\Sigma$ if q_1Rq_2 and $\delta(q_1,\sigma)R\delta(q_2.\ \sigma)$ [Gecseg, 1986].

The relation of strong bisimulation *R* is a congruence [Milner, 1989].

For two initial deterministic finite automata with accepting states DF$A_1=(S, \Sigma, \delta, q_0, F_S)$ and DF$A_2=(T, \Sigma, \delta, q_0, F_T)$ exists an **indistinguishableness relation** $N\subseteq S\times T$, if for all $(s,t)\in N$ and for all $\sigma\in\Sigma$:

- $(s,t)\in N^0$ if and only if $((s\in F_S\wedge t\in F_T)\vee(s\notin F_S\ \wedge t\notin F_T))$,
- $(s,t)\in N^k$ if and only if $(s,t)\in N^{k-1}$ and
- if $\delta(s,\sigma)=s'$ then there exists $t'\in T$ such that $\delta(t,\sigma)=t'$ and $(s',t')\in N^{k-1}$and
- if $\delta(t,\sigma)=t'$ then there exists $s'\in S$ such that $\delta(s,\sigma)=s'$ and $(s',t')\in N^{k-1}$.

The relation of indistinguishableness *N* is a congruence [Milner, 1989].

For Markov chain the bisimulation relation was defined in article [Sokolova, 2004], and its construction is helpful for defining the same relation for reactive probabilistic automata.

Let *R* be an equivalence relation on the set *S*, and let $P_1,P_2\in D(S)$ be discrete probability distributions. Then

$$P_1\equiv_R P_2 \Leftrightarrow \forall C\in S/R\text{: } P_1[C] = P_2[C],$$

where *C* is an equivalence class [Sokolova, 2004].

Let *R* be an equivalence relation on the set *S*, let *A* be a set, and $P_1,P_2\in D(S)$ be discrete probability distributions. Then:

$$P_1\equiv_{R,A} P_2 \Leftrightarrow \forall C\in S/R,\ \forall a\in A\text{: } P_1[a,C] = P_2[a,C]$$

[Sokolova, 2004].

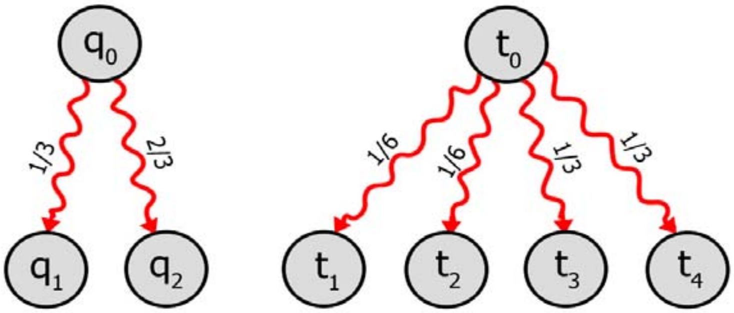


Fig.3. The bisimulation relation on MC

The equivalence relation on the set *Q* of states of Markov chain (Q, δ) will be strong **bisimulation relation** $R\subseteq S\times T$ then and only then when for all $(q,t)\in R$:

if $\delta(q)=P_1$ then there exists a distribution P_2 with $t\in T$ such that $\delta(t)=P_2$ and $P_1\equiv_R P_2$ [Sokolova, 2004].

On figure 3 we have five pairs in bisimulation relation: $\{(q_0,t_0), (q_1,t_1), (q_1,t_2), (q_2,t_3), (q_2,t_4)\}$.

On base of bisimulation relation on Markov chain states we can define the same type of bisimulation for reactive probabilistic automata.

Let $PA_1=(S, \Sigma, \delta)$ and $PA_2=(T, \Sigma, \delta)$ be two reactive probabilistic automata. A **bisimulation relation** $R\subseteq S\times T$ exists if for all $(s,t)\in R$ and for all $\sigma\in\Sigma$:

if $\delta(s,\sigma)=P_1$ then there exists a distribution P_2 with $t\in T$ such that $\delta(t,\sigma)=P_2$ and $P_1\equiv_{R,\Sigma} P_2$ [Sokolova, 2004].

States $(s,t)\in R$ we call bisimilar, what is denoted by $s\approx t$.

On figure 4 we have six pairs in bisimulation relation: $\{(s_0,t_0), (s_1,t_1), (s_2,t_1), (s_3,t_2), (s_4,t_2), (s_5,t_3)\}$.

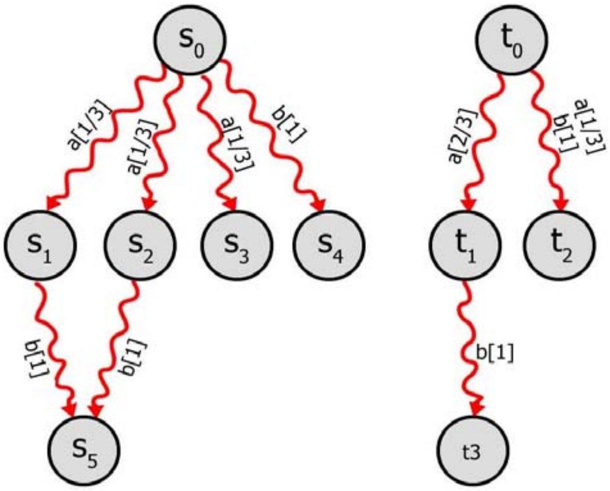


Fig.4. The bisimulation relation on PA

Let $PA_1=(S, \Sigma, \delta, q_0, F_S)$ and $PA_2=(T, \Sigma, \delta, q_0, F_T)$ be two initial reactive probabilistic automata with accepting states. We can define **indistinguishableness relation** $N\subseteq S\times T$, if for all $(s,t)\in N$ and for all $\sigma\in\Sigma$:

$(s,t)\in N^0$ if and only if $((s\in F_S \wedge t\in F_T) \vee (s \notin F_S \wedge t \notin F_T))$,

$(s,t)\in N^k$ if and only if $(s,t)\in N^{k-1}$ and

if $\delta(s,\sigma)=P_1$ then exists the probability distribution P_2 with $t\in T$ such that $\delta(t,\sigma)=P_2$ and $P_1\equiv_{R,\Sigma} P_2$.

For $n=|Q|$, we have

$$N\subseteq N^{n-2}\subseteq N^{n-3}\subseteq\ldots\subseteq N^1\subseteq N^0.$$

States s,t we call indistinguishable, what is denoted by $s \equiv t$, if there exists indistinguishableness relation N, such that $(s,t)\in N$.

On figure 5 we have six pairs in indistinguishableness relation: $\{(s_0,t_0), (s_1,t_1), (s_2,t_1), (s_3,t_2), (s_4,t_2), (s_5,t_3)\}$.

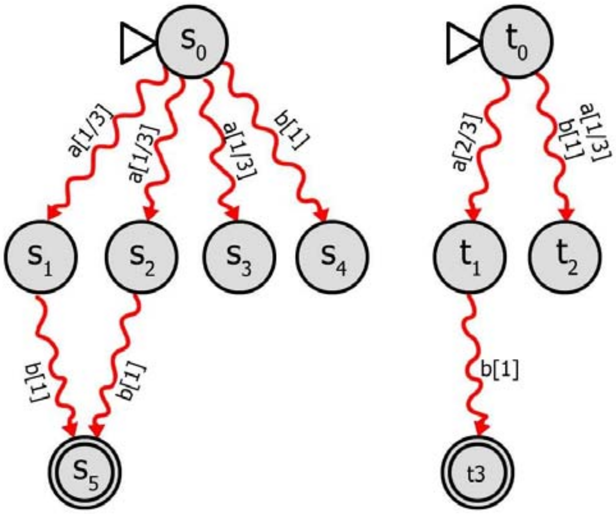


Fig.5. The indistinguishableness relation on PA

The indistinguishableness relation is a congruence in set of states of automata $PA=(Q, \Sigma, \delta, q_0, F)$ - for adequate transition function δ, in analogical way as bisimulation relation, if two results of transition function belong to relation, also states from which we go out belong to relation: for $(q_1,q_2)\in Q$ and symbols $\sigma \in\Sigma$ if q_1Rq_2 and $\delta(q_1,\sigma)R\delta(q_2. \sigma)$.

Minimization of Reactive Probabilistic Automata

The indistinguishableness relation defined in prior section give us possibilty to create minimization methods and algorithms for reactive probabilistic automata.

A probabilistic automata $PA=(Q, \Sigma, \delta, q_0, F)$ recognizing language L with probability p we call minimal, if there doesn't exist automata with smaller number of states recognizing language L with not smaller probability.

In minimal automata there are no two states that could be equivalent in terms of indistinguishableness relation.

A minimization of probabilistic automata parts on two steps:

- elimination of unreachable states (probability to reach those states is 0),
- joining of indistinguishable states (using indistinguishableness relation).

First we show on below code elimination of unreachable states:

Alg.1. Algorithm of elimination of unreachable states:

```
INPUT: $PA=(Q,\Sigma,\delta,q_0,F)$- reactive probabilistic automata.
OUTPUT: $PA'=(Q',\Sigma,\delta',q_0,F')$ - reactive probabilistic automata without
unreachable states, recognizing the same language as $PA$.
FOR ALL {$q\in Q$} DO
  markedStates[$q$]$\leftarrow$0;
END FOR
S.push($q_0$);
markedStates[$q$]$\leftarrow$1;
pr$\leftarrow$0;
WHILE {S$\neq\emptyset$} DO
  p$\leftarrow$S.first();
  S.pop();
  FOR ALL {$\sigma\in\Sigma$} DO
    FOR ALL {$q\in Q$} DO
      pr$\leftarrow$ $\delta(p,\sigma)(q)$;
      IF {pr$\neq$0 $\wedge$ markedStates[$q_0$]=0} THEN
        S.push($q$);
        markedStates[$q$]$\leftarrow$1;
      END IF
    END FOR
  END FOR
END WHILE
FOR ALL {$q\in Q$} DO
```

```
    IF {markedStates[$q$]=1} THEN
      $Q'$.push($q$);
    END IF
  END FOR
  $F' \leftarrow F \cap Q$;
  FOR ALL {$q \in Q$} DO
    IF {markedStates[$q$]=1} THEN
      FOR ALL {$p \in Q$} DO
        IF {markedStates[$p$]=1} THEN
          FOR ALL {$\sigma \in \Sigma$} DO
            $\delta'(q,\sigma)(p) \leftarrow \delta(q,\sigma)(p)$;
          END FOR
        END IF
      END FOR
    END IF
  END FOR
```

In this algorithm *S* is auxiliary stack, on which we put states, which we can reach with non-zero probability going out from the start state q_0. The transition probability function $\sigma(p,\sigma)(q)$ gives probability *pr* of reaching state *q*, going out from state *p*, reading symbol *σ*. The running time of the algorithm time is bounded by:

$$T(n,|\Sigma|) \leq a(7+9n+2|\Sigma|n+2n^2+6|\Sigma|n^2)+c(4+8n+2|\Sigma|n+3n^2+5|\Sigma|n^2) ,$$

where *a* is time of an assignment and *c* is time of comparison, clearly $O(|\Sigma|n^2)$ is the time complexity of this algorithm.

In the algorithm of joining indistinguishable states we use already defined indistinguishableness relation. In one word, states to be indistinguishable, have to be in the same equivalence class, and must have the same probability distribution for symbols and equivalence classes, which can be reach from this states. Inspired by Hopcroft-Ullman's algorithm [Hopcroft, 2000], first we assume that all pairs of states are indistinguishable, above that, that first element of pair is member of final states' set and second isn't. Next analyzing all pair of states and all symbols we find distinguishable states, until the moment that any change is made. Algorithm analyses probability distributions of reaching state from state.

Alg.2. Algorithm of joining indistinguishable states:

```
INPUT: $PA=(Q,\Sigma,\delta,q_0,F)$ - reactive probabilistic automata.
OUTPUT: $PA'=(Q',\Sigma,\delta',q_0',F')$ - minimal reactive probabilistic automata recognizing language $L_{PA}$.
  FOR {i←0; i<|Q|; i←i+1} DO
    FOR {j←0; j≤i; j←j+1} DO
      IF {$(q_i \in F \wedge q_j \notin F) \vee (q_i \notin F \wedge q_j \in F)$} THEN
        $D_{qi,qj} \leftarrow 1$;
      ELSE
        $D_{qi,qj} \leftarrow 0$;
      END IF
    END FOR
  END FOR
  FOR {$i \leftarrow 1$; $i<|Q|$; $i \leftarrow i+1$} DO
    FOR {$j \leftarrow 0$; $j<i$; $j \leftarrow j+1$} DO
      IF {$D_{qi,qj}=0$} THEN
        FOR ALL {$\sigma \in \Sigma$} DO
          $E1 \leftarrow 0$;
          $E2 \leftarrow 0$;
          $N1 \leftarrow 0$;
          $N2 \leftarrow 0$;
          FOR ALL {$p \in Q$} DO
            IF {$D_{qi,p}=0$} THEN
              $E1 \leftarrow E1+\delta(q_i,\sigma)(p)$;
            ELSE
              $N1 \leftarrow N1+\delta(q_i,\sigma)(p)$;
            END IF
            IF {$D_{qj,p}=0$} THEN
              $E2 \leftarrow E2+\delta(q_j,\sigma)(p)$;
            ELSE
              $N2 \leftarrow N2+\delta(q_j,\sigma)(p)$;
            END IF
          END FOR
          IF {$E1 \neq E2 \vee N1 \neq N2$} THEN
            $D_{qi,qj} \leftarrow 1$;
            break;
          END IF
        END FOR
      END IF
    END FOR
  END FOR
  $Q' \leftarrow Q$;
  $F' \leftarrow F$;
  $q_0' \leftarrow q_0$;
  FOR {$i \leftarrow 1$; $i<|Q|$; $i \leftarrow i+1$} DO
    FOR {$j \leftarrow 0$; $j<i$; $j \leftarrow j+1$} DO
      IF {$D_{qi,qj}=0$} THEN
        $Q' \leftarrow Q' \setminus \{q_i, q_j\}$;
        $Q' \leftarrow Q' \cup \{q_{ij}\}$;
        IF{$q_i \in F$} THEN
          $F' \leftarrow F' \setminus \{q_i, q_j\}$;
          $F' \leftarrow F' \cup \{q_{ij}\}$;
        END IF
```

```
50.          IF {j=0} THEN
51.            $q_0 \leftarrow q_j$;
52.          END IF
53.        END IF
54.      END FOR
55.    END FOR
56.    FOR ALL {$q_1 \times q_2 \times \sigma \in Q' \times Q' \times \Sigma$ } DO
57.      IF {$q_1 = q_2 \wedge q_1 \notin Q \wedge q_1 = p_1 p_2 : p_1 p_2 \in Q$} THEN
58.        $\delta'(q_1,\sigma)(q_2) \leftarrow \delta(p_1,\sigma)(q_2) + \delta(q_2,\sigma)(p_2)$;
59.      IF {$q_1 \notin Q \wedge q_1 = p_1 p_2 : p_1 p_2 \in Q$} THEN
60.        $\delta'(q_1,\sigma)(q_2) \leftarrow \delta(p_1,\sigma)(q_2)$;
61.      ELSIF {$q_2 \notin Q \wedge q_2 = p_1 p_2 : p_1 p_2 \in Q$ } THEN
62.        $\delta'(q_1,\sigma)(q_2) \leftarrow \delta(q_1,\sigma)(p_1) + \delta(q_2,\sigma)(p_2)$;
63.      ELSE
64.        $\delta'(q_1,\sigma)(q_2) \leftarrow \delta(q_1,\sigma)(q_2)$;
65.      END IF
66.    END FOR
```

Analyzing algorithm in details: on input we have reactive probabilistic automata; on output we get minimal automata that accept the same language as input automata. In lines 1 to 9 we tentatively fill structure *D*, which is lower triangular matrix of all combination of automata's states. In place where one of the states is final and second isn't, we set value 1, because states are distinguishable. In other case we set 0, providing that all other pairs of states are indistinguishable. In lines 10 to 33 is the main part of algorithm, which decides if states are equal or not, comparing probability distributions. First (line 12) we verify if pair of states is indistinguishable $D_{qi,qj}=0$ (otherwise it makes no sense in analyzing them). For every symbol from alphabet Σ we reset value of auxiliary variables *E1, E2, N1, N2*, in which we will sum probabilities of reaching distinguishable states *N* or indistinguishable states *E*. States will be generally recognized as indistinguishable if values of *E1, E2* and *N1, N2* will be respectively equal. If for two analyzed states, for any symbol of alphabet, we get different values of those variable, loop is interrupted (line 32), because states are distinguishable and we go to next iteration. In the last part of algorithm (from line 38) we create output automata, so we replace indistinguishable states by single states, and calculate values for transition probability function (from line 54). Depending, if we analyze reaching state or going out from new state, values of probability will be summed or copied. The running time of the algorithm is bounded by:

$$T(n,|\Sigma|) \le a(5 + 4.5n - 3.5|\Sigma|n + 7.5n^2 + 2|\Sigma|n^2 + 3n^3 + 1.5|\Sigma|n^3) +$$
$$c(2 + 7n - 2.5|\Sigma|\, n + 7n^2 + |\Sigma|\, n^2 + 7n^3 + 1.5|\Sigma|n^3),$$

so complexity will be $O(|\Sigma|n^3)$.

Lets analyze steps of both algorithms on example from figure 2. First we reset table *markedStates[q_i]*, which size is 7 (automata has 7 states). We push on stack start state. Next we mark with *1* field for this state in table *markedStates[q_0]*. We pop from the stack start state and push those, which we can reach from start state reading symbol *0*, with nonzero probability (those will be q_1, q_2) and for symbol *1*, respectively q_3, q_4, in every case marking them with *1* in table *markedStates[q_i]*. In next iteration we search for states we can reach from states put on the stack. Finally, the only state, which wasn't marked, is q_6. In next steps we exclude it from the set of states of automata (figure 6).

The algorithm of joining indistinguishable states in first part fill structure $D_{qi,qj}$ with *1* in those places where one of states is final, and second isn't – for all combinations of other states with state q_5. Next we check successively all combinations of states and sum probabilities of going out from this states in variables *E1, E2, N1, N2*, for example for states q_1, q_0, values for this variables are *E1=0, E2=1, N1=0, N2=0*, so this pair of states is distinguishable and $D_{qi,qj}$*=1*. Finally structure $D_{qi,qj}$ has value *1* only for pairs: q_1, q_2 and q_3, q_4, which will be replaced by new single states q_{12}, q_{34}. Probabilities for reaching those states will be summed, and for going out from them will be copied (figure 7).

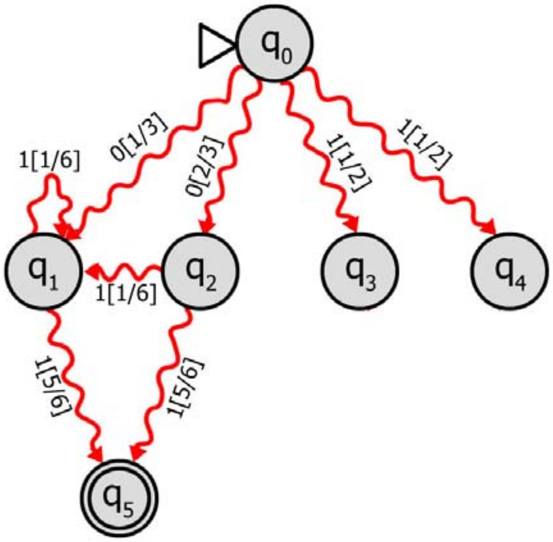


Fig.6. Elimination of unreachable states

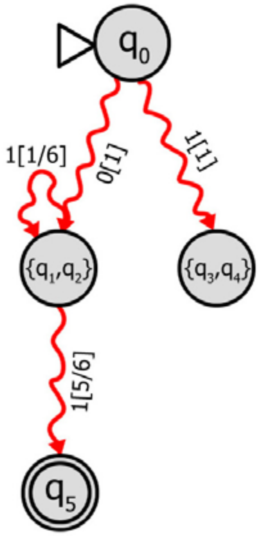


Fig.7. Joining of indistinguishable states

Conclusion

In this article we defined indistinguishableness relation for reactive probabilistic automata, what give us opportunity to build minimization algorithm, with complexity $O(|\Sigma|n^3)$. Algorithms will terminate, because number of states or symbols in alphabet is always limitation for iterations (and we work on finite sets). Probabilities for accepting words don't change because they are respectively summed or copied.

Minimization of any types of automata is important problem that has its practical application – less number of states – less amount of resources. So, this definition of indistinguishableness relation and minimization algorithm is the base for further work on adequate algorithm for quantum automata.

Bibliography

[Aho, 2006] A.V. Aho, M.S. Lam, R. Sethi, J.D. Ullman, Compilers: Principles, Techniques, and Tools (2nd Edition). Addison Wesley, 2006.

[Alur, 1994] R. Alur, D.L. Dill, A theory of timed automata. Theoretical Computer Science 126, 2 (1994), pp. 183 - 235.

[Brzozowski, 1962] J. A. Brzozowski, Canonical regular expressions and minimal state graphs for definite events. In Proceedings of the Symposium on Mathematical Theory of Automata (1962), vol. 12 of MRI Symposia Series, Polytechnic Press of the Polytechnic Institute of Brooklyn, pp. 529 - 561.

[Cao, 2006] Y. Cao, L. Xia, M. Ying, Probabilistic automata for computing with words. ArXiv Computer Science e-prints (2006).

[Gecseg, 1986] F. Gecseg, Products of automata. Monographs on Theoretical Computer Science, Springer-Verlag (1986).

[Golovkins, 2002] M. Golovkins, M. Kravtsev, Probabilistic reversible automata and quantum automata. Lecture Notes In Computer Science 2387 (2002), p. 574.

[Henzinger, 1998] T.A. Henzinger, P.W. Kopke, A. Puri P.Varaiya, What s decidable about hybrid automata? Journal of Computer and System Sciences 57 (1998), pp. 94 - 124.

[Hopcroft, 1971] J.E. Hopcroft, An n log n algorithm for minimizing the states in a finite automaton. In The Theory of Machines and Computations, Z. Kohavi, Ed. Academic Press, 1971, pp. 189 - 196.

[Hopcroft, 2000] J.E. Hopcroft, R. Motwani, J.D. Ullman, Introduction to Automata Theory, Languages, and Computation (2nd Edition). Addison Wesley, 2000.

[Kozen, 1997] D.C. Kozen, Automata and Computability. Springer-Verlag New York, Inc.,

Secaucus, NJ, USA, (1997).

[Kryvyi, 2007] S. Kryvyi, L. Matveeva, E. Lukianova, O. Siedlecka, The Ontology-based view on automata theory. In Proceedings of 13-th International Conference KDS-2007 (Knowledge-Dialog-Solution) (Sofia, 2007), ITHEA, Ed., vol. 2, pp. 427 – 436.

[Milner, 1989] C. Milner, Communication and Concurrency. Prentice-Hall, Inc., upper Saddle River, NJ, USA (1989).

[Rabin, 1963] M.O. Rabin, Probabilistic automata. Information and Controle 6 (1963), pp. 230 - 245.

[Sokolova, 2004] A. Sokolova, E. de Vink, Probabilistic automata: System types, parallel composition and comparison. In Validation of Stochastic Systems: A Guide to Current Research (2004), LNCS 2925, pp. 1 – 43.

[Thomas, 1990] W. Thomas, Automata on infinite objects. Handbook of theoretical computer science: formal models and semantics B (1990), pp. 133 – 191.

Authors' Information

Siedlecka Olga – Institute of Computer and Information Sciences, Czestochowa University of Technology, ul. Dabrowskiego 73, 42-200 Czestochowa, Poland; e-mail: olga.siedlecka@icis.pcz.pl

Major Fields of Scientific Research: Theory of automata, quantum computing, quantum automata

APPLICATION OF INFORMATION THEORIES TO SAFETY OF NUCLEAR POWER PLANTS

Elena Ilina

Abstract: *To this date, strategies aiming at a safe operation of nuclear power plants focused mainly on the prevention of technological breakdowns and, more recently, on the human attitudes and behaviors. New incidents and challenges to safety, however, motivated the nuclear community to look for a new safety approach. The solution became a strong focus on knowledge management and associated theories and sciences as information theories, artificial intelligence, informatics, etc. In all of these, the fundamental role is played by a category of information. This work reviews a number of information interpretations and theories, among which of great relevance are those capturing the fundamental role information plays as a mean to exercise control on the state of a system, those analyzing information communication between agents involved in safety-related activities, and, finally, those which explore the link between information and the limits of our knowledge. Quantitative measures of information content and value are introduced. Completeness, accuracy, and clarity are presented as attributes of information acquired by the receiver. To conclude, suggestions are offered on how to use interpretations and mathematical tools developed within the information theories to maintain and improve safety of nuclear power plants.*

Introduction

It is widely recognized that safety of nuclear power plants is a problem of great relevance for society. If it is not properly managed, the increase in power and complexity of the plants can lead to a catastrophic release of energy and dangerous materials and pollute the environment. To prevent this to happen, the nuclear community always put significant efforts into finding new ways to improve safety.

As a matter of fact, the nuclear accidents in Chernobyl and in Three Mile Island lead to initiation of extensive research activities within the nuclear society. A consensus quickly emerged that the breakdowns could not be explained exclusively from the perspective of technological failures but indeed required new, holistic views on safety. A solution seemed to be a strong focus on human attitudes and behaviors. The concept of safety culture, first introduced by the International Atomic Energy Agency (IAEA), rapidly became increasingly popular [1].

These efforts notwithstanding, incidents continued to occur indicating that important aspects of plants' safety remain unsolved. In fact,

- In 2000, at the Davis Besse nuclear power plant, an extensive material degradation was detected in an area around a nozzle of the reactor pressure vessel. Commentators have suggested that a minor additional propagation of the crack would have lead to the rupture of the pressure vessel. Obviously, the plant owners did not possess accurate and updated information of current condition of the reactor pressure vessel [2].
- In 2006, at the Forsmark nuclear power plant, a sudden disruption from an external power supply lead to the failure of the house turbine operation, the internal battery secured power supply, and of 2 of the 4 emergency diesel generators. Luckily, the 2 diesel generators that started were able to supply sufficient power to the cooling system of the reactor core, and, thus, were able to maintain the whole system in functioning conditions. This dangerous incident has resulted from errors in modifications of old components, caused by insufficient understanding of the consequences of introduced changes [3].

These examples indicated that the established safety frameworks were in need of further improvement if the occurrence of severe accidents is to be prevented. Recognizing this need of further improvement, the IAEA has recently declared a strong focus on knowledge management and built a new knowledge management group for assisting member states in associated questions [4, 5].

However, the scope of knowledge management is known to be broad and overlap, for instance, information management, information theories, artificial intelligence, systems theories, synergetic, informatics, etc. What all of these theories and sciences have in common is the information orientation. With other words, the category of information plays the fundamental role here whilst other categories as data, knowledge, intelligence, etc. can be derived from the category of information. Therefore, a review of main information theories and interpretations seems to be a reasonable start.

Review of main information theories and interpretations

The concept of information acquires meaning only with respect to the context within which it is used. Specifically, different definitions of information can be provided depending on whether information is:

1. Used as a mean for regulation and control.
2. *Transferred* (communication).
3. Or *generated* (acquisition).

(1) Information as a tool for regulating (controlling, steering) activities was emphasized first within the science of cybernetics. This perspective is often denoted as *functional* (also known as cybernetic, external, active, or relation-based). Norbert Wiener, who may be considered as the father of cybernetics, claimed that all goal-oriented actions of human beings are based on information [6]. An opposite, *structural* (also known as attributive, or internal) perspective believes that information mirrors an objectively existing diversity in the reality [7,8,9]. The structural perspective emphasizes that information is an overall property of the reality from its simplest forms to the human brain or complex engineered facilities.

That is to say, the concept of information is the bi-polar concept that arrives into two shapes: a functional perspective and a structural perspective. Both perspectives are necessary for an effective and rigorous management of safety. From the functional point of view, all decisions on safety must to be based on accurate and updated information. From the structural point of view, safety of an engineered facility is determined by its structural organization, i.e. its components, subsystems, systems, and connections between them.

A measure of regulation of complex systems was proposed by another prominent figure in the field of cybernetics, Ross Ashby, who stated that “only variety can destroy variety” [8,9]. The meaning of this statement, which is also known as the *Law of requisite variety*, is that the survival of a system depends on the regulator’s ability to master the diversity of external impacts and to block the flow of information. When *essential variables* start going outside the acceptable range, the regulator must take actions until the essential variables are stabilized and the safe condition is reached. A notion of essential variables (also known as *order parameters* within the science of synergetic) denotes those variables that govern the entire system. Ashby’s view of information as the variety agrees with the *structural* perspective on information.

Information as an instruction (algorithm, program) is emphasized within a non-probabilistic approach, also known as *algorithmic information theory*, developed by Andrej Kolmogorov [10]. Kolmogorov considers information as an *instruction* that has to be executed in order to transform a system from state *A* into state *B*. The larger the difference between the states *A* and *B*, the longer (more complex) the transformation instruction.

Although an exact mathematical measure of Kolmogorov’s complexity has not yet been provided by the community of mathematicians, the idea itself allows fruitful discussions on how to manage instructional information, which is contained in norms, standards, instructions, and other types of documentation. Another insight that follows is that successful accomplishment of the goal is dependent on the quality in a program, which describes what measures need to be taken to achieve the goal.

(2) Information is regarded *as a transferred message* within a *communication model*, which was proposed by Claude Shannon [11]. This model includes at least the following elements: a source (a sender), a channel, and a receiver. The channel is always *noisy*, and noise leads to the loss or misinterpretation of information. Ashby [8,9] was first to point out that the distinction between message and noise depends on what the receiver regards as important. The receiver tends to ignore information that does not promote the achievement of his *goals and objectives.* Furthermore, the receiver cannot understand information that considerably exceeds his *background knowledge*. For this reason, all concepts involving information communication should be formulated accounting first and foremost the information acquired by the receiver, while the information sent by the source should play a lesser role, as pointed out by David Harrah in his *model of rational communication* [12,13].

Illustrativeness of Shannon's and Harrah's communication models helps to understand the role of individual objectives, values and knowledge for information perception and making choices. It is therefore increasingly important for top managers to clarify for all the employees the overall goal and values of the entire organisation. In case of conflicts between the subjective objectives of the individual employees and the overall goal and values of the entire organization, the latter have precedence.

As mentioned above, Shannon's *communication model* [11], considers information as a *message* that is transferred from a source to a receiver via a channel. The *information content* of that message was defined by Hartley [14], Shannon [11], and Wiener [6] as the *uncertainty* that can be eliminated upon reception of the message. Ralph Hartley [13] appears to be first with providing an explicit mathematical way to determine the information content of an event in the simplest case in which an event has *N* possible outcomes with equal probability of occurrence $p = (1/N)$:

$$I = -\log p = \log N$$

Note that according to this interpretation, it is possible to speak about information only when *several alternatives are available*. For situations of certainty (determinism) a number of available alternatives shrinks to 1 and information content shrinks to 0. For situation of complete randomness all alternatives have equal probability and information content goes to maximum. The smaller probability of a chosen alternative to happen, the larger the uncertainty it removes and thus the larger amount of information it *generates*.

The great advance of information theories lies in highlighting relationships between information, uncertainty, presence of alternatives, choice, and, in the end, decision making. A decision, i.e. a choice of one among available alternatives, may need to be made although the available information is insufficient. Each choice is thus associated with a risk of making wrong decision and resulting unwanted consequences. Looking from that perspective, insufficient information can be interpreted in terms of the existence of several alternatives to act. In case of decision making by a group of people, a lack of consensus indicates that more information is needed in order to clarify the best choice.

Hartley's equation was limited to the simplest case of complete randomness and later Shannon proposed a more general equation [11]. The following mathematical expression gives the average information which is available when the knowledge about an alternative is expressed by the probabilities p_i. Information content is then measured by averaging over *n* groups:

$$I = -\sum_{i=1}^{n} p_i \log p_i$$

This mathematical expression is identical to that of entropy, as it is defined in statistical mechanics, and plays a fundamental role in several applications of information theories. Of interest to this review, Jaynes [15] suggested the information content as a fundamental quantity from which the probability values, p_i, weighting the possible outcomes of that event, could be recovered. To this end, the values p_i must be chosen so that information content has a maximum constrained by the available knowledge about a given event. This procedure is known as *a principle of maximum entropy*.

Prescription that follows from the principle of maximum entropy is to use the probability distribution, which maximizes the information content with respect to the available knowledge. This procedure allows making the least biased conclusions under situations, when available information is not sufficient for making certain conclusions. Figuratively speaking, one needs to choose a broad probability distribution that comprises all known events.

(3) Information generation is addressed by a *dynamical information theory* [16], which has recently being developed within the frame of science of synergetic. A central notion of *information value* estimates to what extend the information helps to accomplish *goals and objectives* of the user.

Information without a *meaning* has certainly no *value*. Therefore, it is in many cases convenient to study information value and meaning at the same time, using semantic-pragmatic information theories. The fundamental works of Bar-Hillel and Carnap [17] suggested using logic probability to measure semantic information. Logic probability describes to which degree a hypothesis has been confirmed and, from the practical point of view, resembles probabilistic equations of Shannon and Wiener.

The equation for estimate of information value, *V,* was proposed in an early work of Alexander Harkevich [18]:

$$V = \log\frac{p_1}{p_0} = \log p_1 - \log p_0$$

where p_0 and p_1 - are probabilities of goal accomplishment before (a priori) and after (a posteriori) the information has been acquired.

Disinformation leads to decrease in probability of goal accomplishment and information value becomes negative. In the opposite situation, when the goal has actually been accomplished, the value of information becomes equal maximal information content for the system:

$$V = \log\frac{p_1}{p_0} = \log p_1 - \log p_0 = \log 1 - \log\frac{1}{N} = 0 - (-\log N) = \log N = I$$

In the above equation the posteriori probability p_1 is equal 1 because the goal has been accomplished. The a priori probability p_0 is equal *1/N* because under the conditions of limited knowledge all alternatives are considered being equally probable.

A proposal on how some information theories and interpretations can be applied to safety of nuclear plants

It needs to be emphasized that the nuclear community including the nuclear operators, the government regulators, the international organizations such as the IAEA, and other involved actors, has always had a strong focus on safety and a strive for continuous improvement. However, the nuclear power plants belong to a class of complex dynamic systems, which can be difficult to fully overview, understand and control. This work suggests to use information theories and interpretations to solve some safety issues of nuclear plants, in particular those associated with nuclear containments, maintenance schedule, configuration management, and incident explanation.

Configuration management

As a matter of fact, the amount of information that is handled at nuclear power plants increases steadily with time. The amount of information at the operational start was quite limited and well-structured as so called safety analysis reports (SAR). With time, the plants undergo modifications and the old components and systems are being replaced by new items. Changes in technological processes, variations in environmental parameters, new-employments, and other factors contribute to the need to update the information.

The vital role of information for safety management of nuclear power plants was recognized by the IAEA [19] within the concept of configuration management. The IAEA structures the information into following categories:

- The documentation of the entire life-time of the nuclear power plant comprising its design, manufacturing, construction, pre-operational testing, operation, maintenance, testing, and further modification.
- The information contained in safety standards, codes, norms, etc.
- The personnel files and work instructions.

In all three cases, the IAEA requires information to be *complete* and *accurate*. However, measures of information completeness and accuracy are not discussed. Similarly, potential problems arising from excess of information are not addressed, disregarding the fact that extraction of relevant information from abundant sources is as problematic as the lack of information. Crucial issues connected with quantifying the amount of transferred information and assessing the impact of the receiver's background knowledge on the successful completion of this process do not receive the much needed attention. Finally, the need for time optimization of information generation is mentioned in wordings like "right information at right time", but no indication is offered on appropriate strategies to achieve this objective. An additional remark, which demands consideration by any satisfactory approach to safety management but is not addressed adequately by the IAEA's document, concerns the IAEA's requirement for information be clear. It should be stressed that, although *clarity* is a necessary requirement for achieving safety, it does not suffice. In fact, clear information is not necessarily true, or, alternatively, disinformation can also be clearly communicated.

It is to be noted that according to the communication theories of Shannon and Harrah the information that is received by the receiver is not the same information that has been sent by the source. Information acquired by the receiver is conditioned to the receiver's background knowledge as well as subjective goals, objectives and values. That is why all measures (such as information completeness, accuracy, clarity) must be estimated from the perspective of receivers (users) of information.

Degradation of nuclear containments

It is known that containments in nuclear power plants constitute the last barrier between the dangerous radiation and the environment. In case of a nuclear accident, the containment is supposed to confine the radiation and, by doing so, to protect people and the environment. However, containments were constructed for decades ago and are subject to a long-term ageing deterioration. The original design lifetime of containments has been exceeded in many cases [20]. At the same time, the established testing and inspection practices are limited mostly to visual inspections of containments' accessible surfaces and pressure tests, and are not capable of providing needed information about the state and safety of containments.

From the informational-functional point of view, all conclusions on containments' safety must be based on the accurate and updated information. In this regard, two major questions must be answered: "What material parameters are needed to be measured to assess the state and safety of containments?" and "What methods shall be used to perform needed measurements?"

Here, the earlier mentioned concept of "essential variables" by Ashby (corresponds to a more recent term is "order parameters" in the science of synergetic) is useful. Though the containments are complex structures, composed of a reinforced and prestressed concrete and a steel liner vessel, it might be sufficient to use a few essential variables that govern the state and safety of the entire structure. The previous study [20] has indicated that measurements of four major variables as concrete strength, concrete fracture toughness, prestressing force, and corrosion of steel members provides a solid ground for overall assessment of containments state and safety.

Once essential variables have been identified, one needs to decide what method to use for the measurements. It is known that concrete structures may be tested by means of various methods such as taking cores, Schmidt hummer, visual inspections, radar, radiography, fiber-optical method, acoustic emission, and other destructive and non-destructive (NDT) testing methods. The information approach offers an appropriate tool for choosing the best method with regard to methods' informativeness. According to the previous study [20] the best (the most informative) method seems to be a quantitative acoustic emission (QAE) NDT method. The QAE method can reliably and independently:

- Monitor the entire reinforced concrete structure under the specific conditions of nuclear power plants.
- Reveal specific defects or combination of defects in the entire structure and differentiate between undamaged and damaged parts of the structure.
- Distinguish between different types of defects.
- Assess flaws in terms acceptable for fracture mechanics analysis. Particularly, the real stress state in the prestressed concrete structure.

An additional remark concerns the algorithmic information and its impact on containments' safety. The first generation of containments in the Swedish nuclear power plants was designed and constructed in 60s and 70s, when there were no specific standards or norms available for the task. The second generation of containments was designed and constructed in 80s after the American standards were developed and published. Because of the insufficient instructional (algorithmic) information at the time of design and construction, the old containments are very probably not as strong as the new containments [20].

Optimization of maintenance schedule

Each nuclear power plant contains several thousands of components, most of which are needed to be maintained over time. A question that arises is: "How to determine an optimal time point for maintenance activities?" Support for the optimization of maintenance schedule is provided by the dynamical information theory. As previously discussed, the information theories highlight the intimate relationship between decision making and information.

In case of maintenance, a maintenance engineer must choose (decide) when to perform maintenance activities. The overall goal of the maintenance is to find defects, if any, and to repair them in order to restore the state of the facility. If maintenance activities are scheduled too late, a defect will likely cause failure of the component. Then the decision to repair becomes obvious, the posterior probability p_1 goes to 1, and the information value *V* goes to maximum:

$$V = \log(p_1 / p_0) = \log(1 / p_0) = V_{\max}$$

At the same time, the information content *I* goes to zero. Once the component has broken down, the choice becomes obvious. That is it to say, this is the situation of certainty with one outcome:

$$I = \log N = \log 1 = 0$$

If maintenance activities are scheduled too early, the defect will very probably not be detected, which means that available alternatives of the component's state (the presence of defect alternatively the absence of defect) have equal probabilities. In this situation of uncertainty the information content *I* goes to maximum:

$$I = \log N = I_{\max}$$

At the same time, the information value *V* goes to zero because the maintenance actions do not increase probability of goal-achievement:

$$V = \log(p_1 / p_0) = \log 1 = 0$$

To sum up, when choosing an optimal time point for maintenance, one needs to consider and maximize both information content and value. In many cases, the optimal time point for maintenance will lie close to the end of the life-time, when the defect if large enough to be reliably detected, but is not so large that it can cause a sudden failure.

Incident explanation

It has almost become a common practice for nuclear regulators to blame poor safety culture each time a degradation is observed in a nuclear plant. For instance, after the well known incident in Forsmark in July 2006 [3], the regulator explained the incident by deficiencies in the plants' safety culture. As a result, the following programme of corrective measures had a strong focus on safety culture and associated questions as attitudes to safety, existence of written instructions, etc.

This work believes that the focus on safety culture is necessary but not sufficient for explanation of occurred incidents and prevention of new incidents to happen. Indeed, one needs to take a look from the information perspective and to identify and correct deficiencies in information acquisition, information communication, knowledge creation, decision making, and other stages in information processes.

In case of the Forsmark incident, it is known that the incident was initiated by a severe distortion in the external power grid. This distortion propagated far into the plant and lead to the failure of several power supply systems and safety systems. As a matter of fact, the employees of the plant lacked experiences of similar distortions and could not foresee possibility of such distortions to happen. Furthermore, the incident revealed that the employees of the plant did not fully realize that recently performed modifications of old electrical equipments have changed systemic interactions of the plant. In particular, the sensitivity to distortions increased or with other words, vulnerability of the plant degraded.

As a matter of fact, modification works and exchange of old equipments take place in several operating nuclear plants. As a result, the nuclear community may face situations when decisions need to be taken when the available information is not sufficient to take certain decisions. Therefore, it is important to highlight the role of information aggregation and the need to consider all types of knowledge including practical experiences, modeling results, expert judgments, calculations, tests, etc. for making plausible decisions under conditions of uncertainty.

Conclusions

From the review of main information theories and perspectives, it emerges that information can be understood

- As a *message* which is transferred from a source to a receiver via a communication channel.
- As an *instruction* (program, algorithm) which, once it is carried out, allows the transformation of a system from state *A* to state *B*.
- As *uncertainty* regarding the present state or evolution of a system, which can be reduced upon reception of the message.
- To mirror the diversity of reality (*structural* perspective).
- To provide the ground for regulation, control, steering, and other goal-directed activities (*functional* perspective).

Furthermore, information can be *quantified*, and therefore *measured*. The relevance of its potentially fundamental role as a tool to verify whether transfer of crucial knowledge has properly occurred cannot be overestimated. In addition, as it is emphasized in the communication models, issues concerning completeness, accuracy, and clarity of information should mainly concern the information *acquired by the receiver*. The role of background knowledge, objectives and values of the receiver is an additional point of concern which emerges from this review. Finally, a mathematical model has been developed which allow for current information to be updated without running the risk of unjustified bias in favour of a particular alternative/outcome.

The diversity in interpretations of the concept of information mirrors potential instances in which problems for the safe operation of nuclear plants may arise. The promotion of a proper understanding this concept within the specific context in which it is used, and of the mathematical tools developed within information theories, will hopefully help preventing accidents to occur in the future.

References

[1] "Safety culture in nuclear installations. Guidance for use in the enhancement of safety culture." IAEA, Vienna, 2002.

[2] "Davis Besse reactor pressure vessel head degradation." US NRC, NUREG/BR-0353, rev 1, August 2008.

[3] Wikdahl CE. "Forsmark incident on the 25th of July 2006." Analysis group at KSU, N 4, Sweden, 2006.

[4] 2002 IAEA General Conference, IAEA, Vienna, 2004.

[5] "Knowledge management now seen as a priority." IAEA, Nuclear news, Vienna, September 2007.

[6] Wiener N. "Cybernetics: control and communication in the animal and the machine." New York, 1948.

[7] Ursul A. "Information: a philosophical study." Berlin, Dietz, 1970.

[8] Ashby R. "An introduction to cybernetics." J Wiley, New York, 1956.

[9] Ashby R. "Design for a brain." Chapman and Hall, London, 1952.

[10] Kolmogorov A. "Three approaches to the definition of information content." Problems of information communication, N 1, pages 3-11, Moscow, 1965.

[11] Shannon C. "A mathematical theory of communication." The Bell system technical journal, N 27, pages 379-423 and 623-656, 1948.

[12] Harrah D. "The psychological concept of information." Philosophy and Phenomenological Research, Vol 18, N 2, pages 242-249; December 1957.

[13] Harrah D. "A model of communication." Philosophy of science, Vol 23, N 4, pages 333-342, October 1956.

[14] Hartley RVL. "Transmission of information." Bell system technical journal, N 7, pages 335-363, 1928.

[15] Jaynes E. “Information theory and statistical mechanics.” Physical review, Volume 106, N4, pages 620-630, May 1957.

[16] Chernavsky D. “Synergetic and information.” URSS, Moscow, 2004.

[17] Bar-Hillel Y., Carnap R. “Semantic information.” British Journal of science, N 4, pages 147-157.

[18] Harkevich A. “On information value.” Cybernetics problems, N 4, Physmathgiz, Moscow, 1960.

[19] “Configuration management in nuclear power plants.” IAEA-TECDOC-1335, Wien, 2003.

[20] Österberg E. (changed name to Ilina E.). “Revealing of age-related deterioration of reinforced concrete containments in nuclear power plants – Requirements and NDT methods.” The licentiate research thesis, The royal institute of technology, Stockholm, 2004.

Author information

Elena Ilina- *Present employment 1:System analyst at the Swedish radiation safety authority, Department of nuclear plants safety, System assessments, Solna Strandväg 122, 17 154 Solna, Sweden, email: elena.ilina@ssm.se*

Present employment 2: PhD student at the Royal institute of technology, Department of structural engineering, Brinellvägen 34, 100 44 Stockholm, Sweden, email: elena.ilina@byv.kth.se

Major fields of scientific research: Application of information and knowledge theories to the safety assessments of complex technologies

INCREASING RELIABILITY AND IMPROVING THE PROCESS OF ACCUMULATOR CHARGING BASED ON THE DEVELOPMENT OF PCGRAPH APPLICATION

Irena Nowotyńska, Andrzej Smykla

***Abstract**: The article presents the software written in Builder C++ that monitors the process of processor impulse charger. Protocol, interface, components used and the future research are presented.*

***Keywords**: PCgraph, developing software, charging process, C++ Builder*

***ACM Classification Keywords**: C.3 SPECIAL-PURPOSE AND APPLICATION-BASED SYSTEMS*

Introduction

The article presents the software written in Builder C++ language that monitors the process of charging through COM port. Pulsar is impulse charger made by Elprog. It is a professional fast impulse charger to charge all kinds of cells available on the market. The product won the prestigious prize “Polish Market” in 2004. Among its qualities are:

1. the speed of charging
2. the reliability of the charger process
3. the regeneration of old cells
4. the monitoring of charging process.

Computer program that read the data directly from charger is a useful tool to analyze the quality of aku pack. The charger system and the graphical interface of PCGraph program is shown in Fig. 1. and Fig. 2.

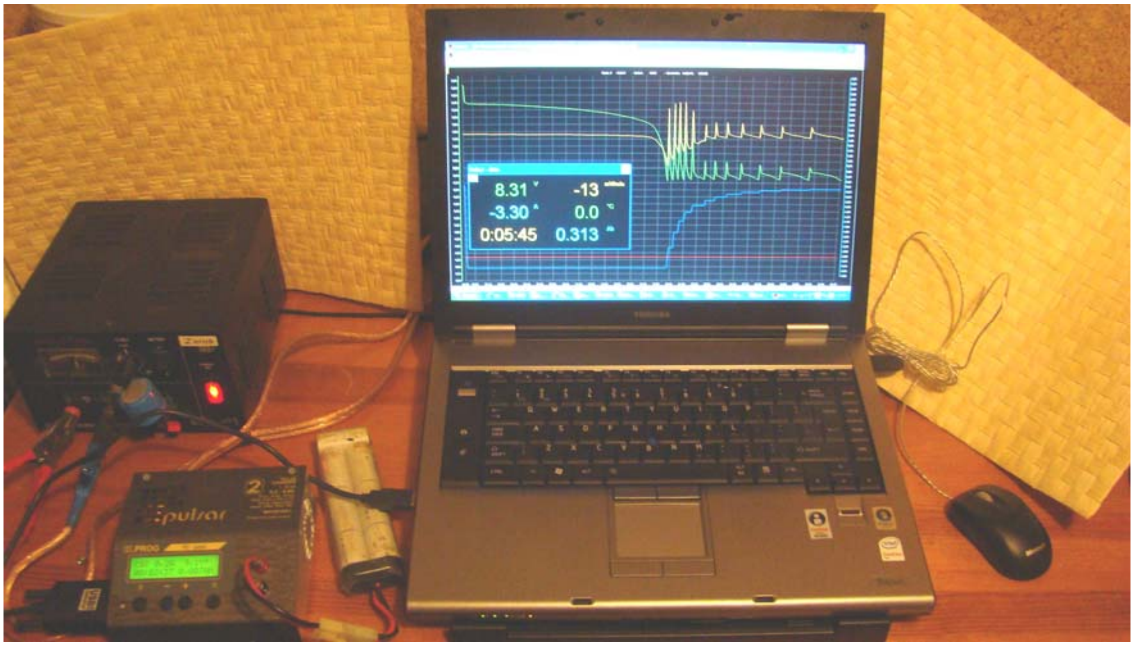


Fig. 1. The charger system during work

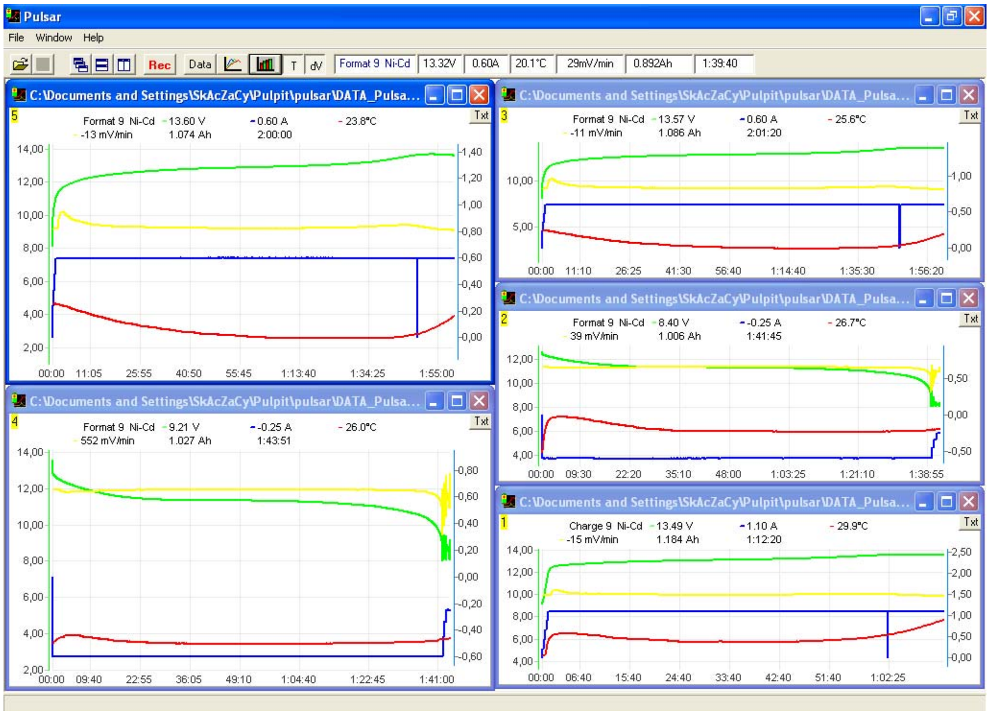

Pulsar
File Window Help
Rec Data T dV
Format 9 Ni-Cd 13.32V 0.60A 20.1°C 29mV/min 0.892Ah 1:39:40
C:\Documents and Settings\SkAcZaCy\Pulpit\pulsar\DATA_Pulsa...
5 Format 9 Ni-Cd 13.60 V -13 mV/min 0.60 A 1.074 Ah 2:00:00 23.8°C
3 Format 9 Ni-Cd 13.57 V -11 mV/min 0.60 A 1.086 Ah 2:01:20 25.6°C
2 Format 9 Ni-Cd 8.40 V 39 mV/min -0.25 A 1.006 Ah 1:41:45 26.7°C
4 Format 9 Ni-Cd 9.21 V 552 mV/min -0.25 A 1.027 Ah 1:43:51 26.0°C
1 Charge 9 Ni-Cd 13.49 V -15 mV/min 1.10 A 1.184 Ah 1:12:20 29.9°C
Txt

Fig. 2 The graphical interface of PCGraph program (the colors determine: blue – amperes, red – temperature, green – volts, yellow – dV/dt). Figure shows three cycles of charging process and two cycles of discharging

The system is used especially when high reliability is required. The systems work for example in European Space Agency and Polish Polar Station (Spitsbergen).

The program is written in Builder C++ as an MDI application. Components used are presented in Figure 3.

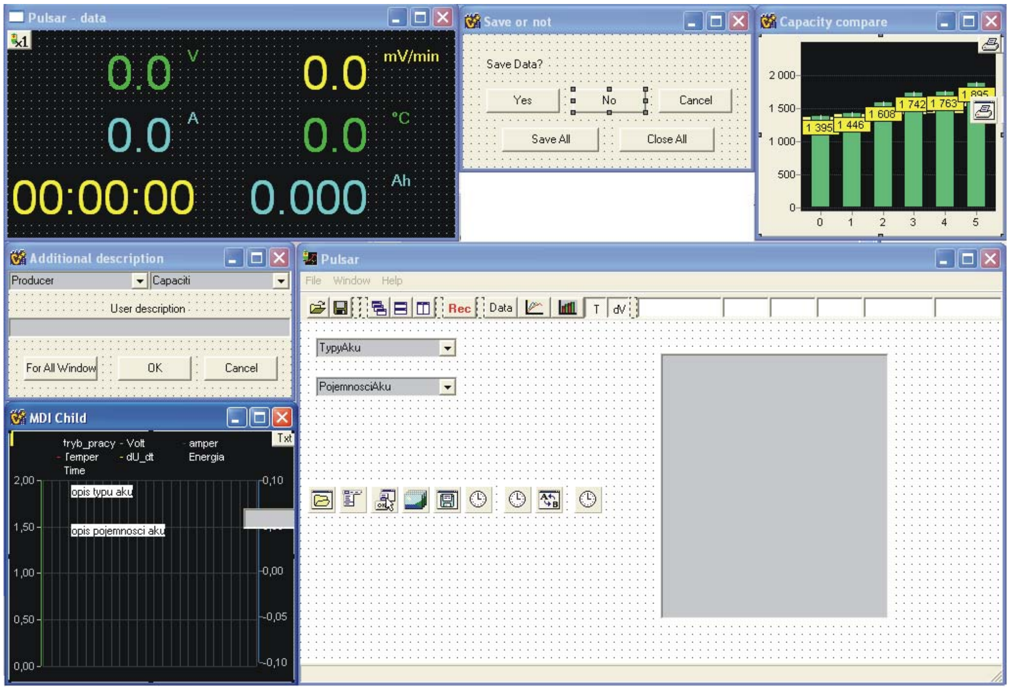


Fig. 3. The view of chosen forms with components used.

The data flows diagram between the most important components are presented in Figure 4.

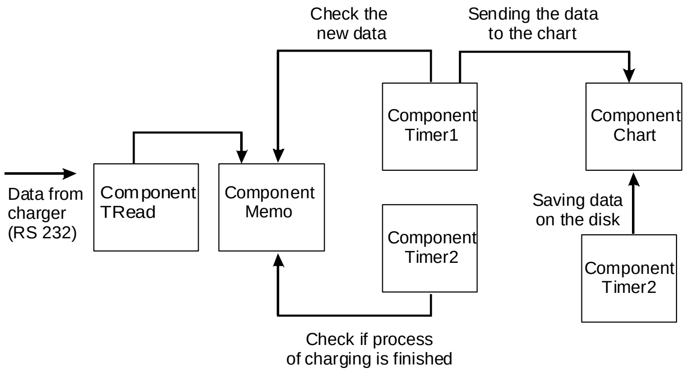


Fig. 4 Data flow diagram of applications.

A serial port is used for communication. The format of protocol (32 bytes long) is presented in Figure 5.

#C	3	5	D	00035,	12029,	+0199,	000,	00001
Begin frame	Number of cells	Cell type	Char. Type	Time[s]	Volts[mV]	Curent[mV]	Temperature[°C]	Energy [mAh]

Fig. 5. The charger communications protocol data format

The type of calls determined the kind of process charging (Table 1).

Table 1

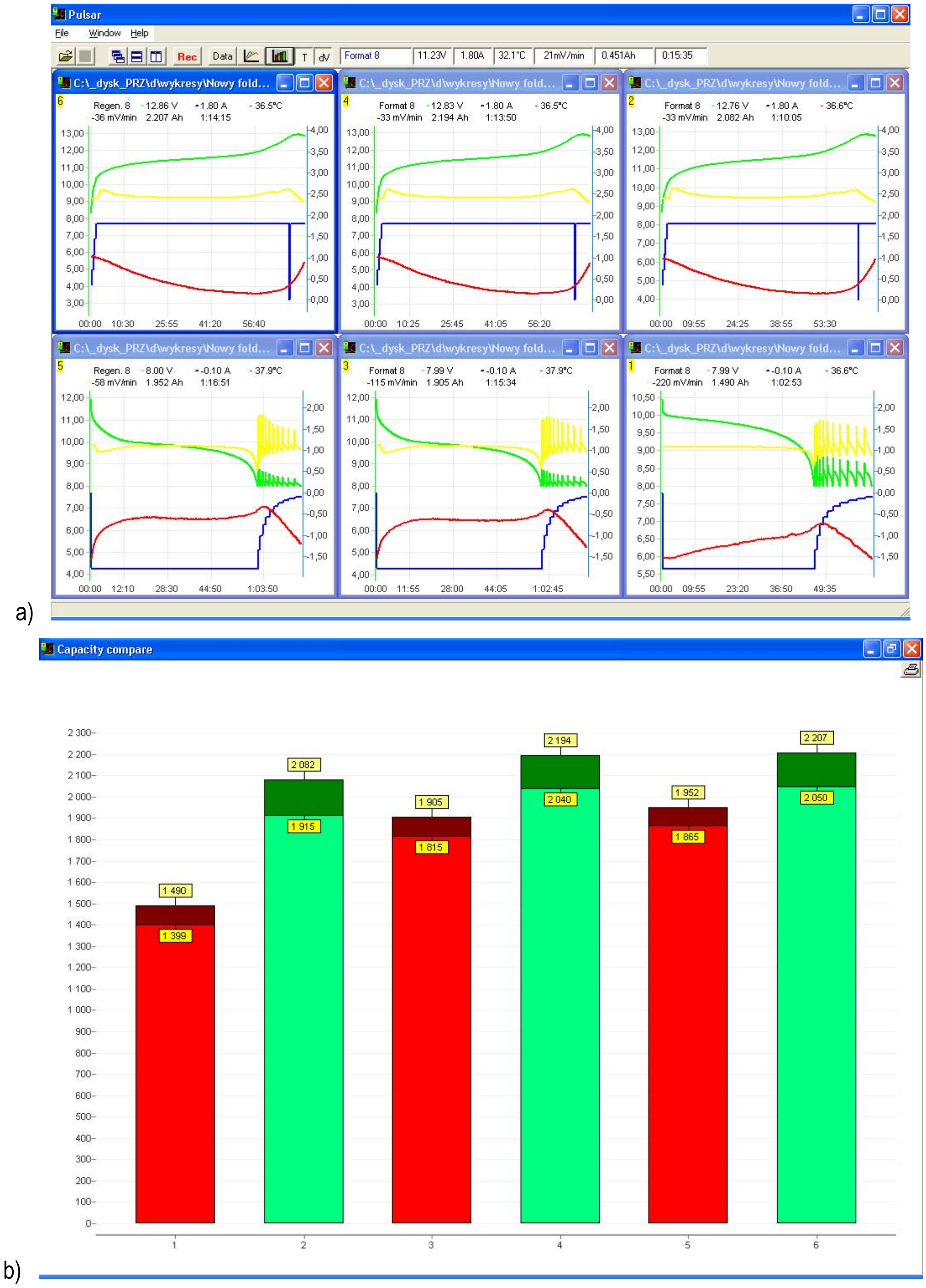
Type of calls	Kind of accumulator
1	Ni-Cd
2	Ni-MH
3	Pb-bat
4	RAM
5	Li-Ion
6	Li-Pol
7	Li-TA
8	Li-S

The charge mode is described with letters – for their meaning see Table 2. The plus and minus sign before the ampere's value means charging and discharging.

Table 2.

Charge/ Discharge mode	Charge mode
D	Discharge
S	CH. Simple
R	CH. Reflex
P	CH. Pb-bat
L	CH. Lith
C	Regen.
C	Charge
D	Disch
F	Format

The program presented enables the recording of several discharging/recharging cycles and to present the data in graph form. The possibility of such analysis is valuable during the package's regeneration as well as during the determination of its consumption in time (see Figure 6).



Format 8 | 11.23V | 1.80A | 32.1°C | 21mV/min | 0.451Ah | 0:15:35

6 Regen. 8 12.86 V -1.80 A 36.5°C -36 mV/min 2.207 Ah 1:14:15

4 Format 8 12.83 V -1.80 A 36.5°C -33 mV/min 2.194 Ah 1:13:50

2 Format 8 12.76 V -1.80 A 36.6°C -33 mV/min 2.082 Ah 1:10:05

5 Regen. 8 8.00 V -0.10 A 37.9°C -58 mV/min 1.952 Ah 1:16:51

3 Format 8 7.99 V -0.10 A 37.9°C -115 mV/min 1.905 Ah 1:15:34

1 Format 8 7.99 V -0.10 A 36.6°C -220 mV/min 1.490 Ah 1:02:53

a)

b)

Fig. 6. A three-cycle discharging/recharging process in which the capacity of an accumulator package was raised from 1400mAh to 1900 mAh

a) the graph of separate cycles b) the graph comparing capacity

The monitoring program that collects data and presents them in the form of graphs is therefore extremely useful to all that increases the functionality of the charger. The reading and proper interpretation of graphs is possible only after gaining some essential skill by the user which often requires some time and effort.

The analyze of received data

The reading and correct interpretation of charging and discharging process characteristics that are obtained enables us to determine the actual quality of accumulator. However, some experience in reading the characteristics obtained is required.

A correct charging cycle of a twelve-cell NiMH accumulator package is shown in Figure 7.

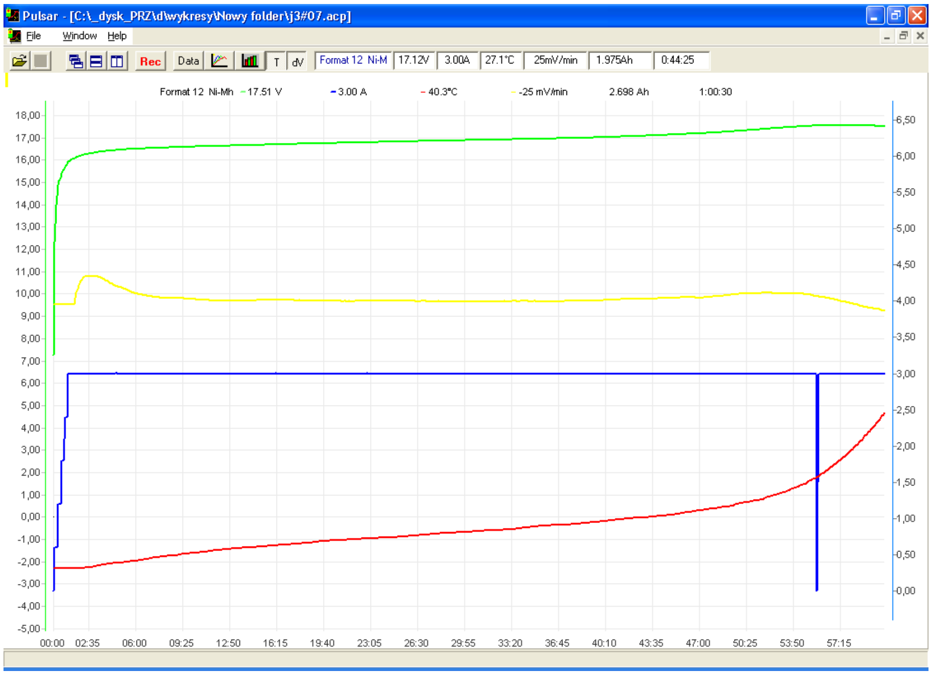


Fig. 7. A correct charging cycle of a twelve-cell accumulator package.

The list of typical damage of accumulator packages during the charging process is shown in Figures 7-12.

Figure 8a shows how faults of two cells – 1:12 min and 1:45 min - were revealed during the charging process. Data analysis does not enable us to tell which cells are malfunctioning. This information is to be obtained only after measuring the voltage on individual cells.

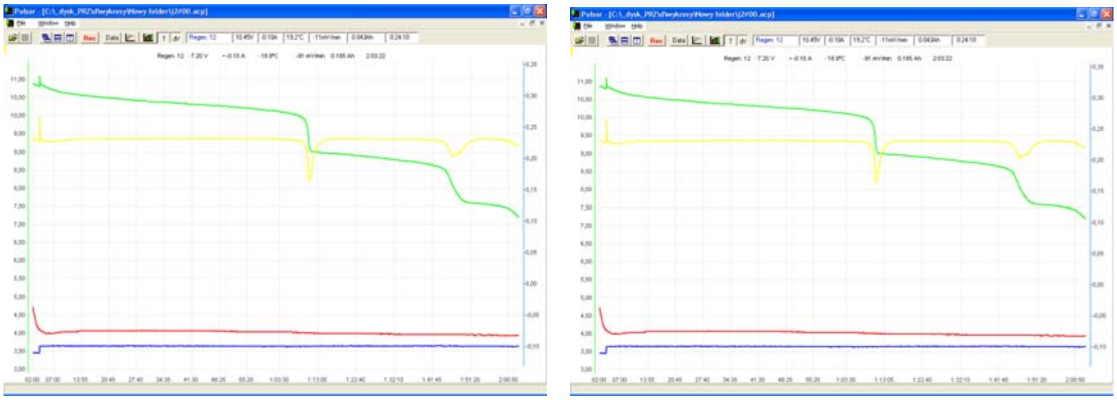
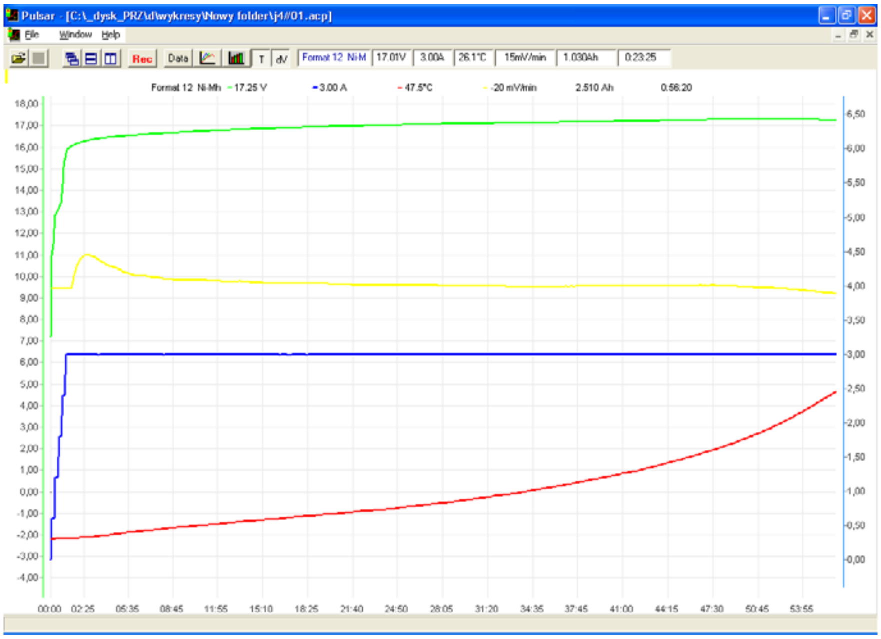


Fig. 8. The examples of packages of accumulators' with two weaker cells

Figure 9 shows a charging process characteristic that is too flat. The process of voltage rising cannot be observed. Huge differences in charging of individual cells could be the reason – in this case package regeneration might help. The graph might also indicate the wear of the package.



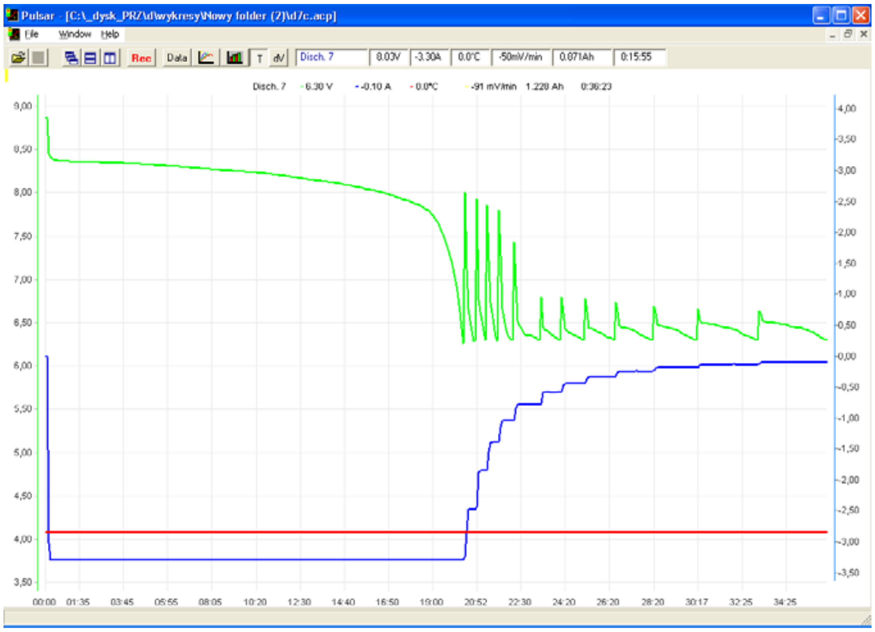
Pulsar - [C:_dysk_PRZ\d\wykresy\Nowy folder\j4//01.acp]

Format 12 NiM | 17.01V | 3.00A | 26.1°C | 15mV/min | 1.030Ah | 0:23:25

Format 12 Ni-Mh - 17.25 V - 3.00 A - 47.5°C - -20 mV/min 2.510 Ah 0:56:20

Fig. 9. “Flat” charging characteristic.

Figure 10 shows a case in which faulty cell in the package is revealed (time 0:23:25).



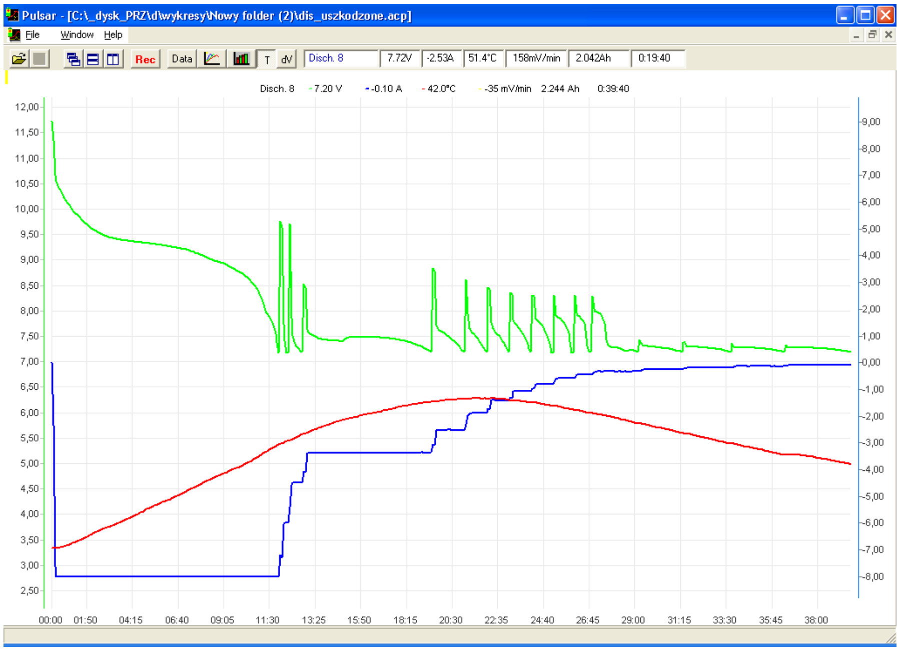
Pulsar - [C:_dysk_PRZ\d\wykresy\Nowy folder (2)\d7c.acp]

Disch. 7 | 0.03V | -3.30A | 0.0°C | -50mV/min | 0.071Ah | 0:15:55

Disch. 7 - 6.30 V - -0.10 A - 0.0°C - -91 mV/min 1.220 Ah 0:36:23

Fig. 10. Charging characteristic indicating the impact of faulty cell.

The process of package discharging with too high internal resistance is shown in figure 11. After 2 minutes 28 seconds the program limited the current (one cell was eliminated). Between 13 and 18 minutes the voltage was partially regenerated on the eliminated cell.



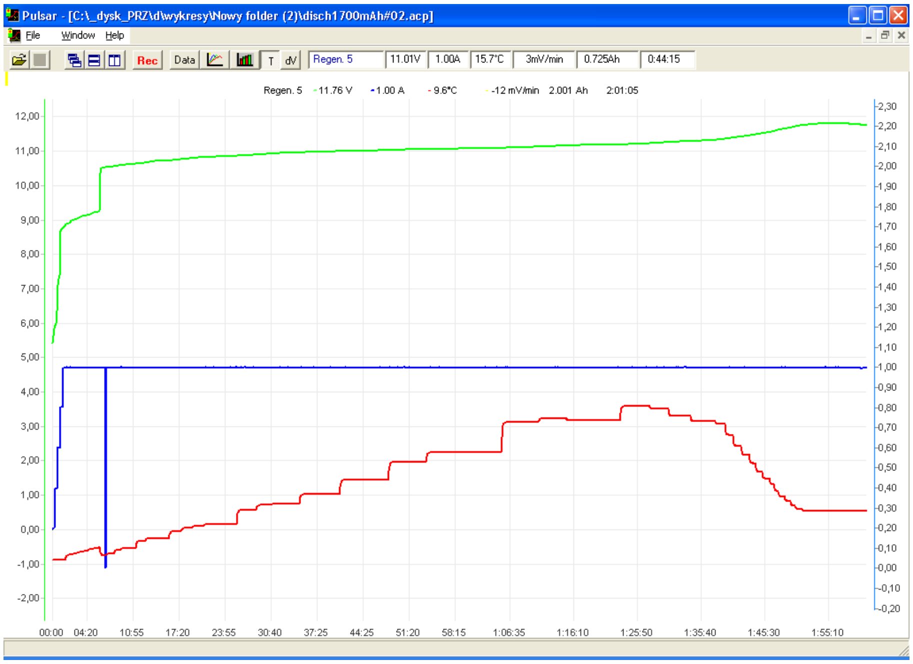
Pulsar - [C:_dysk_PRZ\d\wykresy\Nowy folder (2)\dis_uszkodzone.acp]

Disch. 8 | 7.72V | -2.53A | 51.4°C | 158mV/min | 2.042Ah | 0:19:40

Disch. 8 - 7.20 V - -0.10 A - 42.0°C - -35 mV/min 2.244 Ah 0:39:40

Fig. 11. The partial regeneration of the faulty cell during the discharging process.

Figure 12 shows the effect of package regeneration where one of the cells back to life (6 minutes). The cell was discharged to 0V.



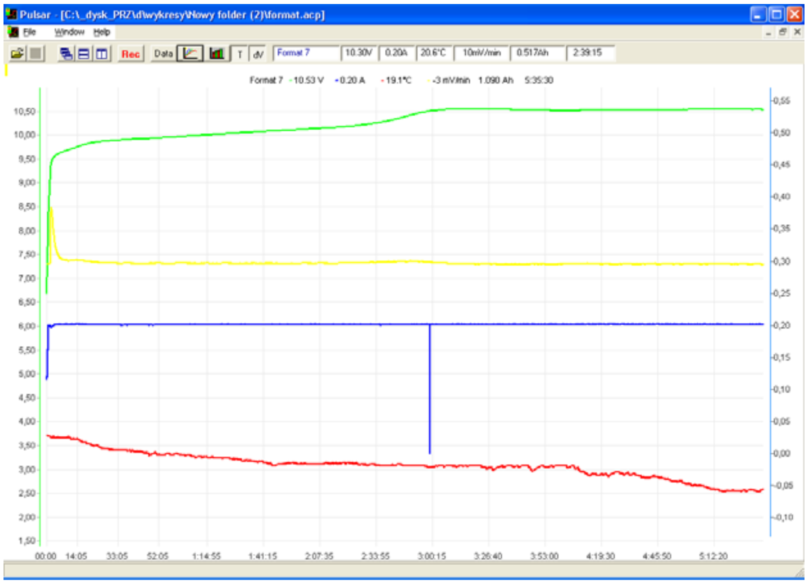
Pulsar - [C:_dysk_PRZ\d\wykresy\Nowy folder (2)\disch1700mAh#02.acp]

Regen. 5 | 11.01V | 1.00A | 15.7°C | 3mV/min | 0.725Ah | 0:44:15

Regen. 5 - 11.76 V - 1.00 A - 9.6°C - -12 mV/min 2.001 Ah 2:01:05

Fig. 12. The charging process of a package with one totally discharged cell.

The next figure (Fig. 13) shows a worn-up package (a case analogical to Fig. 9). The voltage during the charging process increased only a little. However, “Inflex” was correctly revealed. Inflex is a moment in the charging process when a significant temperature rise of the charged accumulator begins. The detection of this point enables us to finish the charging process and therefore the accumulator packages could be usable for a longer time period. The Inflex detection is marked on the voltage graph by a vertical line. At present there is a very small amount of this kind of equipment available on the market. Also, Figure 13 does not depict “Delta Peak” – a point of abrupt voltage change during the charging process – which indicates a significant exploitation of the package.

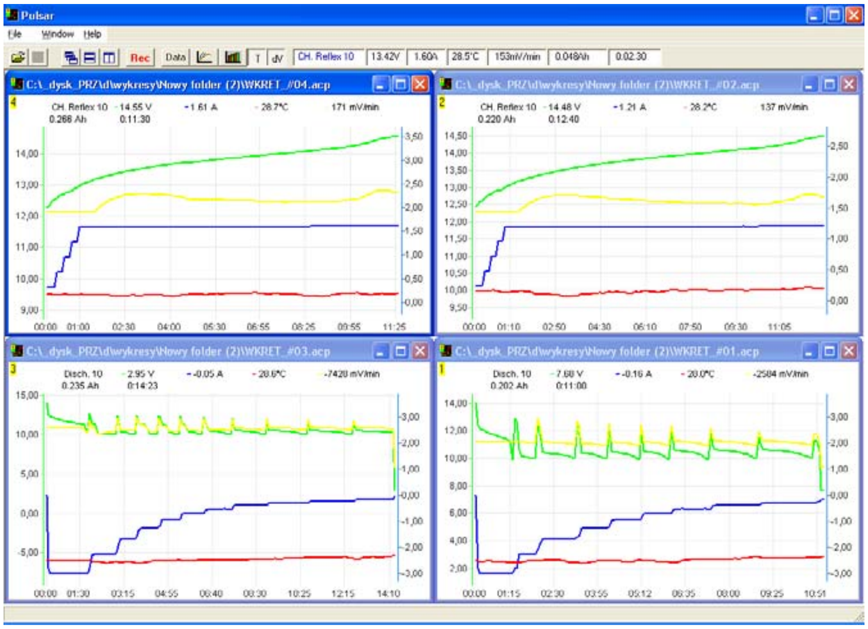

Pulsar - [C:_dysk_PRZ\d\wykresy\Nowy folder (2)\format.acp]

Format 7 | 10.30V | 0.20A | 20.6°C | 10mV/min | 0.517Ah | 2:39:15

Format 7 10.53 V 0.20 A 19.1°C -3 mV/min 1.090 Ah 5:35:30

Fig. 13. The charging process of exploited package.

An example of an unsuccessful attempt at package regeneration is shown in Figure 14. The cycle consisted of two accumulator charging and discharging processes. No improvement in the cell’s parameters was observed. The graph indicates the exploitation of the package.


Pulsar

CH. Reflex 10 | 13.42V | 1.60A | 28.5°C | 153mV/min | 0.048Ah | 0:02:30

C:_dysk_PRZ\d\wykresy\Nowy folder (2)\WKRET_#04.acp

CH. Reflex 10 14.55 V 1.61 A 28.7°C 171 mV/min
0.268 Ah 0:11:30

C:_dysk_PRZ\d\wykresy\Nowy folder (2)\WKRET_#02.acp

CH. Reflex 10 14.48 V 1.21 A 28.2°C 137 mV/min
0.220 Ah 0:12:40

C:_dysk_PRZ\d\wykresy\Nowy folder (2)\WKRET_#03.acp

Disch. 10 2.95 V -0.05 A 20.6°C -7420 mV/min
0.235 Ah 0:14:23

C:_dysk_PRZ\d\wykresy\Nowy folder (2)\WKRET_#01.acp

Disch. 10 7.60 V -0.16 A 20.0°C -2584 mV/min
0.202 Ah 0:11:00

Fig. 14. An unsuccessful attempt at package regeneration.

A successful attempt at package regeneration was shown in Figure 6.

In case shown in Figure 6 the effective accumulator capacity increased from 1490 mAh to 1952 mAh (about 50%). This information could be gained from the comparative column graphs (Fig. 6b.). In this case the package efficiency was regained in almost 100%.

An ideal charging process is presented in Figure 15 where a correctly revealed Inflex and the end of the charging process at the point of package's temperature rise are shown.

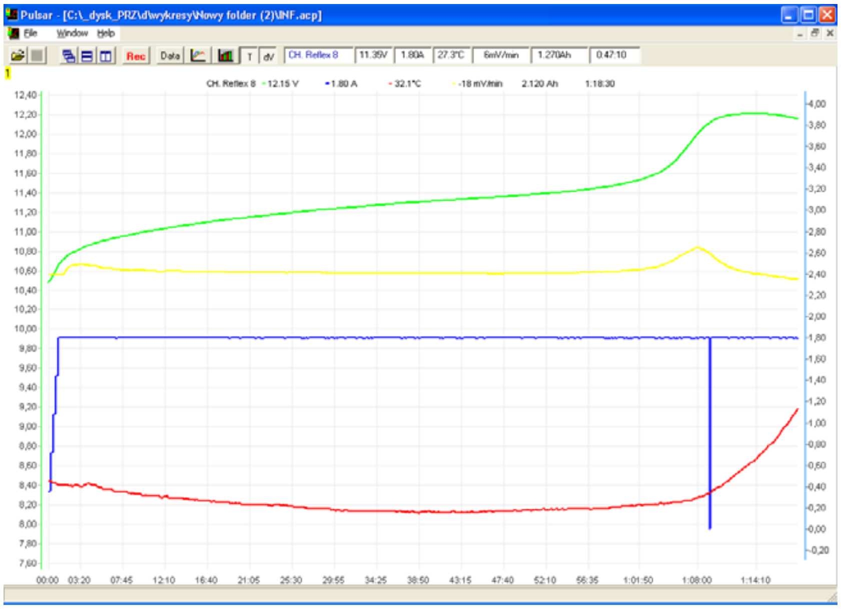


Fig. 15. A correct charging process.

Figure 16 shows a case of self-regeneration of a package during the charging process. After 1 hour, when the current and voltage were limited, the voltage of batter rises. Such situations might happen if a package had not been used for a long time.

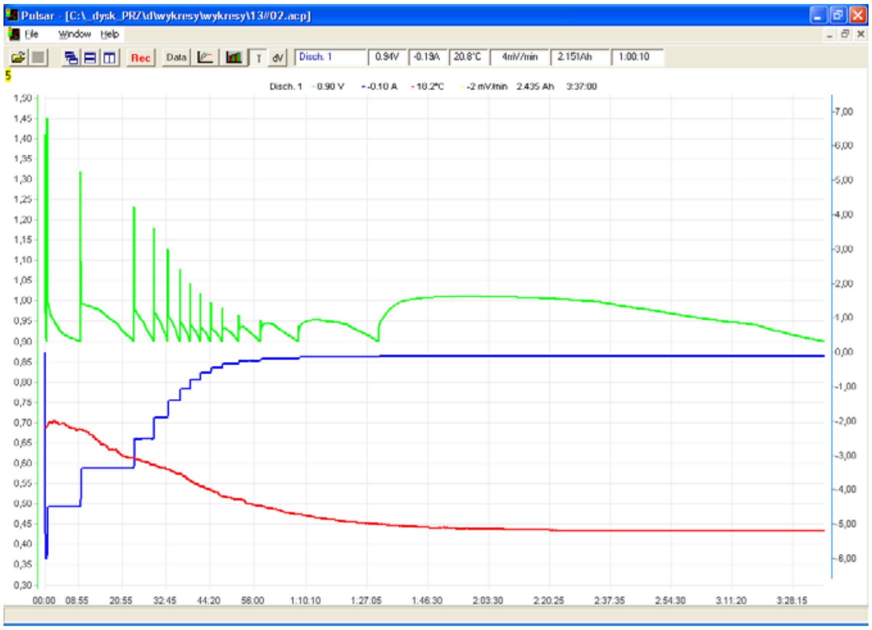


Fig. 16. Self-regeneration of a package during the charging process.

Figure 17 depicts a potentially dangerous situation. Data Peak and Inflex were not detected ant the temperature of the accumulator rose. If the package was charged with a big current an inflammation or even an explosion of the package might occur.

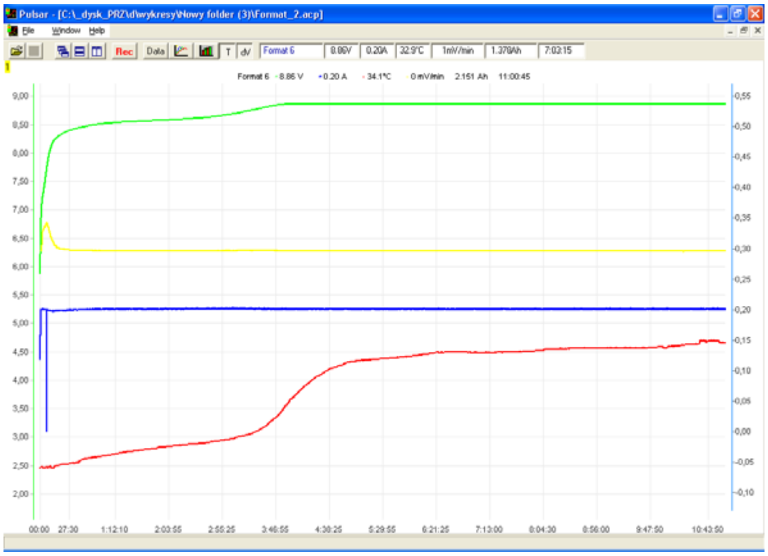


Fig. 17. A package reacting to charging not with a voltage rise but with a temperature increase. A potentially dangerous situation.

Equalization of Li-xxx cells

At present lithium cells are more commonly used. Long life-span of this kind of power sources is possible in the case of correct charging of individual cells in the package. It is important that every cell in the package was charged exactly to a certain characteristic voltage. Taking the whole package consisting of several cells into consideration, the entire voltage equals the sum of voltages on individual cells:

$$V_{package}=V_{1\ cell}+V_{2\ cell}+\ldots+V_{N\ cell}$$

The level of package charging is regulated in most chargers in order to protect the unit from overcharging. However, if one of the cells is undercharged, the others take some part of energy and its voltage level is higher. Protection against incomplete charging can be used that partially protects the unit from overcharging. For example, for cells of LiPol type the maximum charging level is set to 4,15V instead of 4,2V. However, with too huge differences in charging levels of individual cells it is only a partial solution. This approach is especially dangerous for cells of LiPol type where overcharging literally causes the danger of package explosion. Therefore, in these cases it seems necessary to use packages leveling devices – balancers. Each cell is then plugged to a unit which brings the individual voltage levels to one level. It is often accomplished by charging the specific cell with a system dissipating the energy in the form of heat. A scheme of such situation is presented in Figure 18.

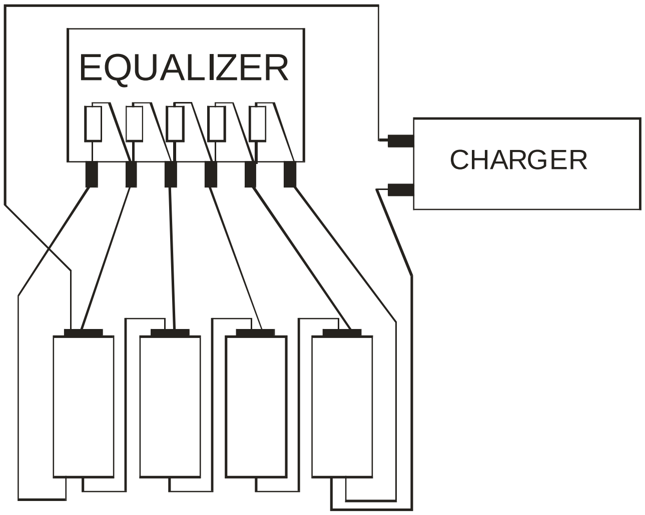


Fig. 18. A scheme of a system balancing the package with simultaneous charging.

Equalizers currently produced by Elprog company are shown in Figure 19.

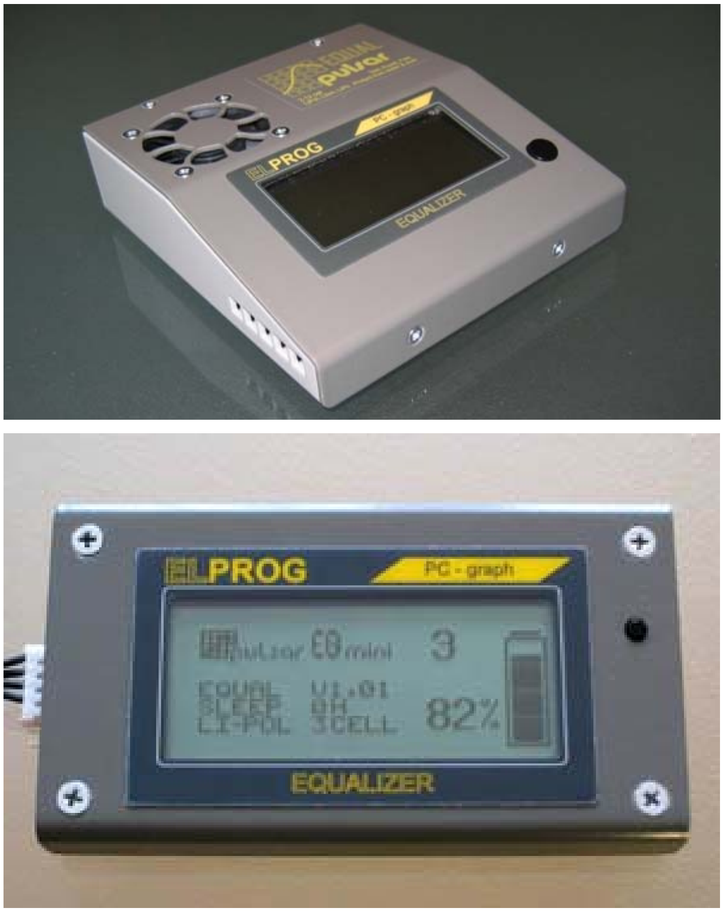


Fig. 19. Two types of equalizers produced in Elprog company.

As in the case of a charger, both devices provide record and data analysis by a computer system. PCGraph software was adopted to work with these devices. The main window of the program is shown in Figure 20.

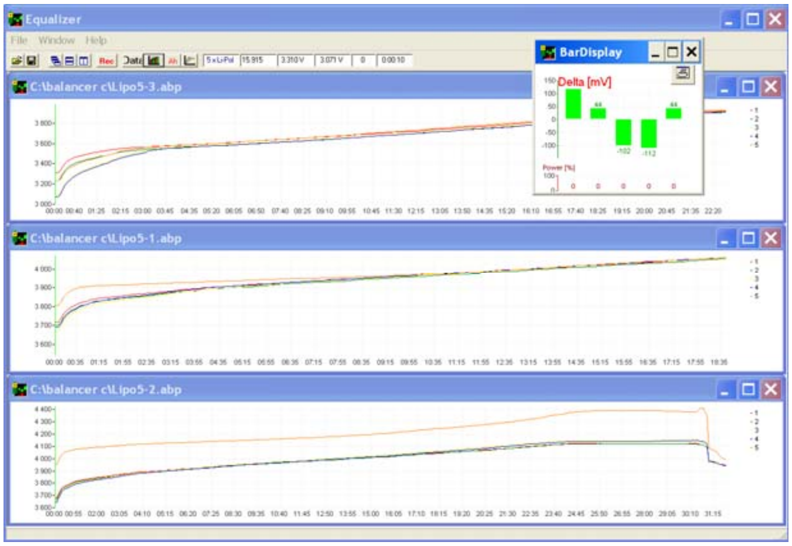

Equalizer
File Window Help
BarDisplay
Delta [mV]
Power [%]
C:\balancer c\Lipo5-3.abp
C:\balancer c\Lipo5-1.abp
C:\balancer c\Lipo5-2.abp

Fig. 20. The PCGraph program - version for EQUALIZER during work.

BarDisplay window seen in the right angle of Figure 20 allows the preview of momentary values of voltage differences on individual cells. As in the case of a charger, the software was written with the use of C++ Builder. A basic unit of components is shown in Figure 21.

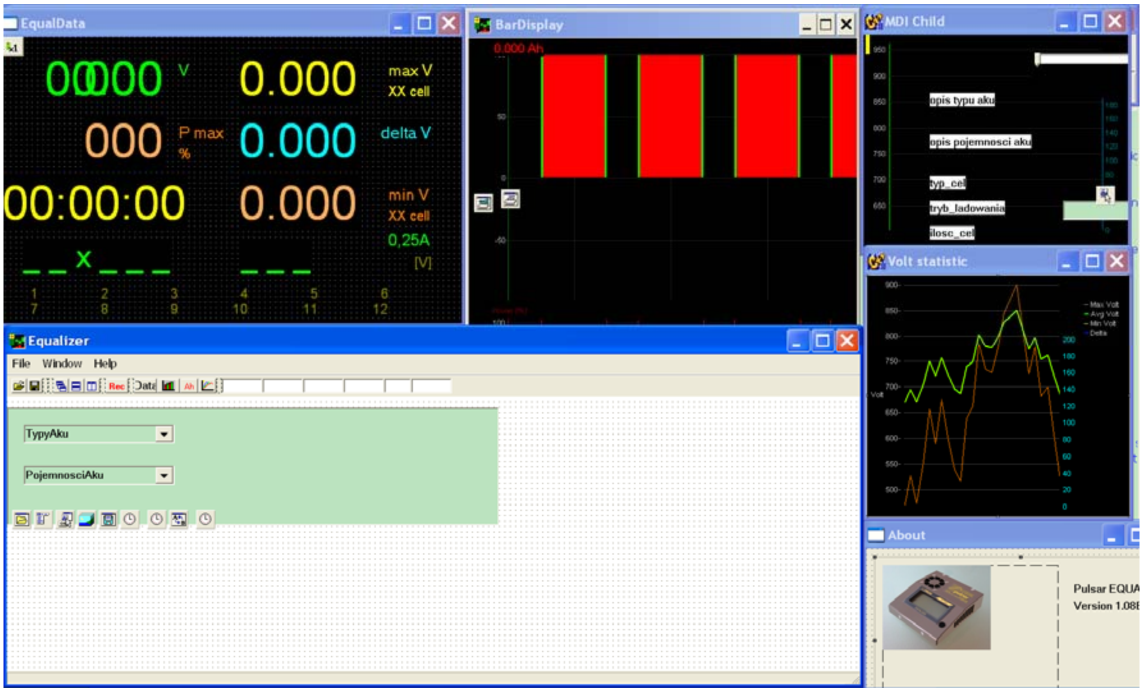


Fig. 21. A basic unit of components of application building in the EQUALIZER version.

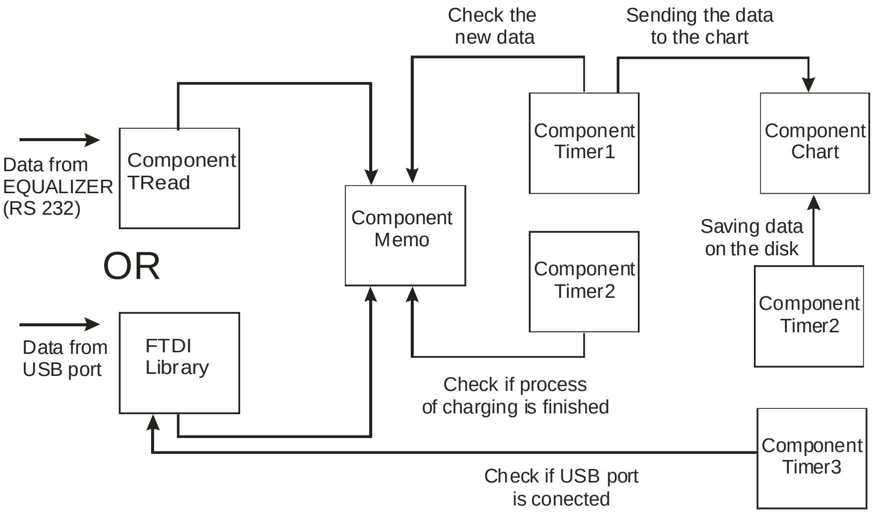


Fig. 22 Data flow diagram of applications.

FTDI library has a set of functions to connect a PC computer with an USB port through an FTDI chip. The set of functions of the chip support is delivered by the producer in the form of DLL library. The format of protocol (80 bytes long) is presented in Figure 23.

#	A	B	C	D	Volt1,	Adjustment1,	Volt2,	Adjustment2,	...,	Volt12,	Adjustment12

Rys 23 The format of protocol for equalizer device

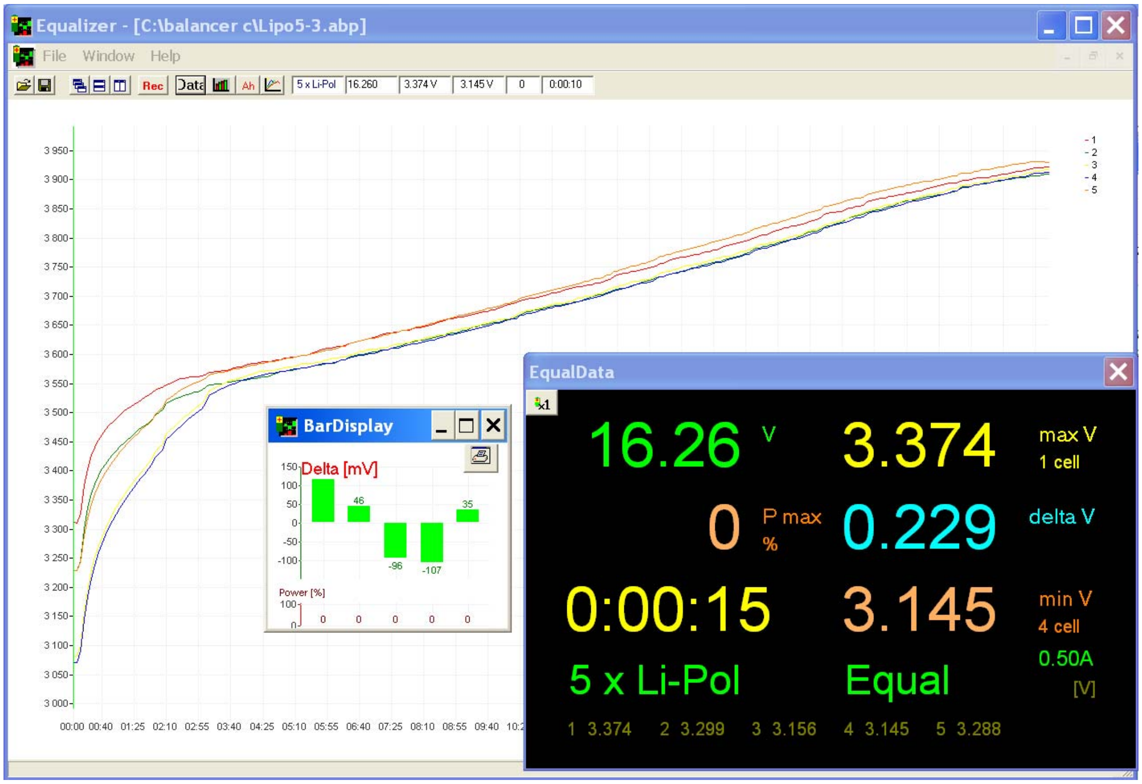
The meaning of fields:

# - begin frame

A,B,C,D - 4 char
A - mode ASCII
t - test
e - equal
f - fast
B - 1-12 - number of cells HEX
C - 5-7 - type of cell HEX
5 – LiIon
6 – LiPol
7 – LiPh (Fe)
D - reserve {now „B" char is sending}
[Volt] - {0000} 4 char ASCII [mV]
[Adjustment] {0} one char from 0 to 10 in ASCII {0,1, .. 9,:}
[end transmision] -{0D 0A} 2 chars HEX

The data transmission is sent every 5 seconds.

A leveling of package session is presented in Figure 24. The voltage of the whole unit rises which means that a charger was also plugged during the leveling process. There is a possibility of using the equalizer without package charging. The device in this mode is powered by the leveled package so the entire voltage is going to decrease. The big black window shows the read values characteristic for the process. BarDisplay window presents the momentary charging of individual cells (depending on the location of the cursor on the graph). The package shown in Figure 24 was successfully leveled. The values set on individual cells differ no more than 1mV at the end of the process.



16.26 V

3.374 max V 1 cell

0 P max %

0.229 delta V

0:00:15

3.145 min V 4 cell

5 x Li-Pol

Equal

0.50A

[V]

1 3.374 2 3.299 3 3.156 4 3.145 5 3.288

Fig. 24. The correct cycle of package leveling.

The current research

Research on the process of creation the expert system that would make reading graphs not essential for the user is being conducted. Initial analyses using the artificial neural network were satisfactory. The network (for the scheme see Figure 25) was prepared as an expert system stating whether the discharging process of an accumulator package finished successfully.

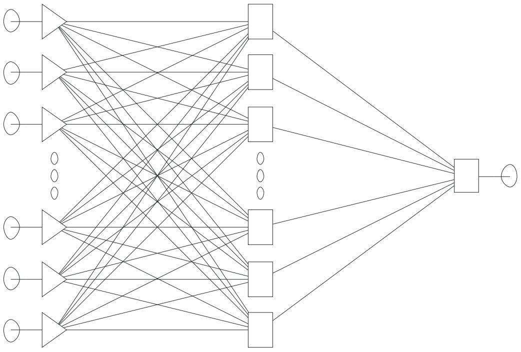
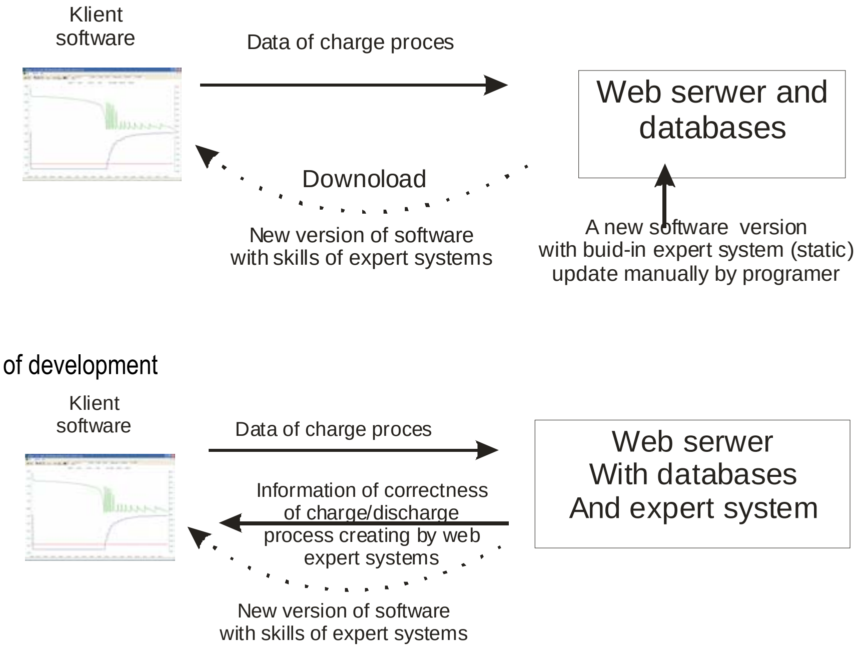


Fig. 25. The scheme of neural network used to identify the fault of accumulator packages.

The data received during the charging process was normalized in table of volting level [U_1,U_2,..,U_n]. The output vector consisted of two elements (the correct package and the wrong package). Coincidence of the network's learning process was observed during the experiment. The modeling of neural network was made in STATISTICA NEURAL NETWORKS system. At this stage collecting more actual data including information about the charging process is crucial. In order to achieve these purposes the current program version includes the record of extra information about the package.

The future

It is probable that the next program version would enable sending the collected characteristics directly to the network server that contains a database. Building an internet server including the expert system is another aim that should be possible to achieve. It would be attainable for the user to check whether their package is correct using the current knowledge of server's expert system and the data could increase sources available. Several possible scenarios of development are shown in Figure 26.



a) Next version of software

Klient software
Data of charge proces
Web serwer and databases
Downoload
New version of software with skills of expert systems
A new software version with buid-in expert system (static) update manually by programer

b) Possibility of development

Klient software
Data of charge proces
Web serwer With databases And expert system
Information of correctness of charge/discharge process creating by web expert systems
New version of software with skills of expert systems

c) Future software bases on agent oriented technology.

Fig. 26. The strategy of development of the software presented

The possibility of building evaluation system that would not only find the package faults but also indicate the type of damage, e.g. damage of a single cell (or a few cells), complete discharging of one cell in the package, wrong charging method used (e.g. Ni-Cd accumulators were charged as Ni-MH), wear–out of the package etc should be examined during future research. The significance of such research is emphasized by the fact that the increase of the number of mobile devices was observed in recent years and therefore there will be more demand for evaluating and repair of accumulator. The charger is also currently utilized by certain services such as police, fire brigades where the reliability of technical devices is of great importance. The use of data-mining for huge amount of data might result in development of routine accumulator exchange or inspection standards. The experiments indicate that the chargers delivered with the equipment often cannot make full use of the device. Also they can damage the accumulator packages even though they were destined for the concrete type of equipment as it often happens in different kinds of inexpensive mobile tools. The user exchanges the devices or accumulator packed which causes the increase of their number on garbage dumps and causes the pollution of environment. Therefore, the conduction of this research and further development of software presented seems to be appropriate.

Conclusion

1. The use of application to analyze the charger mode expands the device's possibilities.
2. The analysis of characteristic obtained requires experience in their interpretation.
3. The use of expert system to analyzing chosen signals will be possible in the future.
4. The creation of databases including the characteristics' history during the usage of separate models of accumulators might make it possible to forecast the wear–out of packages. In older to do that an analysis using data-mining should be performed. The current program version includes the mechanism of accumulator description suitable for such research.

The creation of internet service collecting data from the user that might be automatically integrated with the program for expert system learning and data-mining seems appropriate.

Author Information

Andrzej Smykla – *Rzeszow Univesity of technology; Departure of Computer Science –,ul. W. Pola 2 Rzeszow 35-959, Poland; e-mail: asmykla@prz.edu.pl*

Major Fields of Scientific Research: General theoretical information and mechanical research,

Irena Nowotyńska *- Researcher; Department of Computer Engineering in Management, Rzeszow University of Technology, W. Pola 2, 35-959 Rzeszów, Poland; e-mail:* i_nowot@prz.edu.pl *Major Fields of Scientific Research: FEM methods, Artificial Neural Networks, Mechanical Working*

TRANSLITERATION AND LONGEST MATCH STRATEGY

Dimiter Skordev

***Abstract**: A natural requirement on transliteration systems is that transliteration and its inversion could be easily performed. To make this requirement more precise, we consider a text transduction as easily performable if it can be accomplished by a finite transducing device such that all successful tokenizations of input words are compliant with the left-to-right longest-match strategy. Applied to inversion of transliteration this gives a convenient sufficient condition for reversibility of transliteration.*

***Keywords**: left to right, longest match, transliteration, reversible transliteration, sequential transducer.*

***ACM Classification Keywords:** E.4 Data: Coding and information theory — Formal models of communication*

Introduction

When considering a transliteration system, it is natural to impose on the transliteration and on its inversion the requirement to be easily performable (cf., for instance, [7, Section 7] or [8, Subsection 25.2]). This requirement can be made more precise in different ways. We present one in the following spirit: a text transduction is regarded as easily performable if it can be accomplished by a finite transducing device such that all successful tokenizations of input words are compliant with the left-to-right longest-match strategy (cf. e.g. [1] for some other applications of this strategy). The requirement that both transliteration and its inversion could be easily performed is understood in this sense, but with one device and its inverse used for transliteration and for its inversion, respectively. A convenient sufficient condition for reversibility of transliteration is obtained in this way.

Some definitions and examples

Let Σ and Δ be two alphabets, and let T be a mapping of Σ^* into the set $P(\Delta^*)$ of the subsets of Δ^*, where Σ^* consists, as usually, of all finite strings of symbols from Σ (including the empty string ε), and similarly for Δ^*. We may intuitively regard T as a mathematical description of some transliteration system from Σ to Δ, and for any ω in Σ^* consider the strings belonging to $T(\omega)$ as the admissible transliterations of ω in this system (clearly each actual transliteration system from Σ to Δ can be supplied with a description of this kind having some specific features). For any element ω of Σ^* the elements of $T(\omega)$ will be called *images* of ω under T, and ω will be called a *pre-image* under T of each of them. The mapping T will be called *total* if each element of Σ^* has an image under T, *surjective* if each element of Δ^* has a pre-image under T, *single-valued* if no element of Σ^* has two distinct images under T, and *injective* if no element of Δ^* has two distinct pre-images under T. The *inverse mapping* T^{-1} is the mapping of Δ^* into $P(\Sigma^*)$ defined as follows: for any τ in Δ^* the set $T^{-1}(\tau)$ consists of all pre-images of τ under T. Clearly T is total if and only if T^{-1} is surjective, and T is injective if and only if T^{-1} is single-valued.

The following will be assumed in the three examples below: $\Sigma\backslash\Delta$ consists of the capital and the small Russian letters, $\Delta\backslash\Sigma$ consists of the capital and the small Latin letters, $\Sigma\cap\Delta$ contains the space character, the digits and other characters commonly used both in Russian and in English (for instance punctuation marks).

Example 1. The transliteration system proposed in [8] can be described by a mapping T such that for any ω in Σ^* the set $T(\omega)$ has as its only element the string from Δ^* obtainable from ω by means of replacements of the following kinds (for being easier distinguishable, all Russian letters will be given in boldface):

А→A, **а**→a, **Б**→B, **б**→b, **В**→V, **в**→v, **Г**→G, **г**→g, **Д**→D, **д**→d, **Е**→E, **е**→e, **Ё**→Yo, **ё**→yo,
Ж→Zh, **ж**→zh, **З**→Z, **з**→z, **И**→I, **и**→i, **Й**→Yj, **й**→yj, **К**→K, **к**→k, **Л**→L, **л**→l, **М**→M. **м**→m,
Н→N, **н**→n, **О**→O, **о**→o, **П**→P, **п**→p, **Р**→R, **р**→r, **С**→S, **с**→s, **Т**→T, **т**→t, **У**→U, **у**→u,
Ф→F, **ф**→f, **Х**→Kh, **х**→kh, **Ц**→C, **ц**→c. **Ч**→Ch, **ч**→ch, **Ш**→Sh, **ш**→sh, **Щ**→Th, **щ**→th,
Ъ→Jh, **ъ**→jh, **Ы**→Ih, **ы**→ih, **Ь**→J, **ь**→j, **Э**→Eh, **э**→eh, **Ю**→Yu, **ю**→yu, **Я**→Ya, **я**→ya

(for instance, if ω_0 is the sentence "**Гармонический ряд расходится**." then $T(\omega_0)$ has as its only element the string "Garmonicheskiyj ryad raskhoditsya."). The mapping T is total and single-valued by its definition, and it is injective, but not surjective because of the fact that no Russian letter is replaced by "Y" or "y", and there is no Russian letter whose replacement string begins with "h".

Example 2. We shall add to the transliteration system considered in Example 1 some replacements used by a system that is suggested in a document accessible from http://www.metodii.com. Let us consider a mapping T such that for any ω in Σ^* the set $T(\omega)$ consists of all strings from Δ^* obtainable from ω by means of replacements of the kinds considered in Example 1 and of the following additional ones: **Ж**→X, **ж**→x, **Ч**→Q, **ч**→q, **Ш**→W, **ш**→w (now the set $T(\omega_0)$ for the concrete ω_0 from Example 1 will have two elements – the string indicated there and the same string with "Garmoniqeskiyj" instead of "Garmonicheskiyj"). The mapping T is again total, injective and non-surjective, but it is not single-valued.

Example 3. Let us define a single-valued mapping T in the same way as in Example 1, except that we take now **Й**→Yy, **й**→yy, **Х**→Hh and **х**→hh instead of **Й**→Yj, **й**→yj, **Х**→Kh and **х**→kh, respectively (thus the set $T(\omega_0)$ for the concrete ω_0 from that example will have as its only element the string "Garmonicheskiyy ryad rashhoditsya."). Then T is still total, single-valued, injective and not surjective, but its injectiveness is seen in a somewhat more complicated way.

The mappings T from the above examples have the property that $T(\varepsilon)=\{\varepsilon\}$ and for any ω_1 and ω_2 in Σ^* the equality $T(\omega_1\omega_2)=T(\omega_1)T(\omega_2)$ holds (its right-hand side denotes the set of all concatenations $\tau_1\tau_2$, where τ_1 belongs to $T(\omega_1)$ and τ_2 belongs to $T(\omega_2)$). Any mapping T with this property will be called *homomorphic*. Each mapping C of Σ into $P(\Delta^*)$ can be extended in a unique way to a homomorphic mapping T of Σ^* into $P(\Delta^*)$; the mapping T in question will be said to be *generated by C*. Intuitively, we may regard the elements of $C(\sigma)$ for any σ in Σ as the admissible code strings for σ, and regard the mapping T generated by C as a description of transliteration done by replacing the symbols from Σ with admissible code strings for them. If T is generated by C then T is total if and only if all sets $C(\sigma)$ are non-empty, T is single-valued if and only if each set $C(\sigma)$ has at most one element, and T is injective if and only if the following two conditions are satisfied:

(i) $C(\sigma_1)$ and $C(\sigma_2)$ have no common element, whenever σ_1 and σ_2 are distinct symbols of Σ;

(ii) no string in Δ^* can be represented in two different ways as a concatenation of strings belonging to the union of all $C(\sigma)$ corresponding to symbols σ of Σ.

In the practically important case when the above-mentioned union is finite (the mapping T will be called *finitary* in that case) a check for the condition (ii) can be always done by means of a theorem of Sardinas and Patterson [5].

In particular, the injectivity of the mappings T from Examples 1, 2 and 3 can be established also in this way (taking as C the restriction of T to Σ).

The possibility of proving injectivity by means of the Sardinas-Patterson theorem does not make pointless the search for other convenient injectivity criteria, namely such ones that give only sufficient conditions for injectivity but guarantee a better quality of the injectivity. There are at least two reasons for this.

1. For the convenience of a transliteration system not only the injectivity matters, but also the quality of the injectivity. An injective mapping T of Σ^* into $P(\Delta^*)$ could allow to find easily some element of $T(\omega)$ for any ω in Σ^*, but the problem to find ω when an element of $T(\omega)$ is given could be much more difficult. To have an example of such a situation, let us consider the mapping T from Example 3. Suppose that a string τ is given whose first symbol is "s", all other ones being "h", and we look for a string ω in Σ^* such that τ belongs to $T(\omega)$. Such a string ω exists, and its first symbol is "**ш**" if the length of τ is even and "**с**" otherwise. Evidently, it is not possible to determine the first symbol of ω on the base of knowing only some proper prefix of τ, thus in the case of a long τ it would be not easy to find ω by reading τ only once from left to right and writing consecutive symbols of ω.
2. Among the numerous transliteration systems proposed until now there are some whose corresponding mappings T are non-homomorphic due to context-dependent encoding of some letters. Such is the case for example with the second of the two systems proposed by Uspensky in [7] – in that system the Cyrillic letter "**Й**" has the encoding "Jh" when followed by a vowel or by a soft sign and has the encoding "J" otherwise, the letter "**й**" being treated in a similar way (however, any other symbol from the corresponding alphabet Σ has a unique code string not depending on the context).

The finitary homomorphic mappings of Σ^* into $P(\Delta^*)$ are a particular case of mappings accomplished by means of sequential transducers in the sense of [2, Section 3.3]. By the definition accepted there, a *sequential transducer* with input alphabet Σ and output alphabet Δ is any quintuple of the form (K,Σ,Δ,H,s_0), where K is a finite set (the set of the *states*), s_0 is an element of K (the *start state*), H is a finite set of quadruples (called *moves*) with first and last components in K, and second and third components in Σ^* and Δ^*, respectively.[1] The mapping T *accomplished* by such a sequential transducer is defined as follows: for any ω in Σ^* the set $T(\omega)$ consists of the elements τ of Δ^* such that for some non-negative integer k, some $\omega_1, \ldots, \omega_k$ in Σ^*, some $\tau_1, \ldots, \tau_k$ in Δ^* and some $s_1, \ldots, s_k$ in K the quadruples $(s_{i-1},\omega_l,\tau_i,s_i)$, $l=1,\ldots,k$, belong to H, and the equalities $\omega=\omega_1\ldots\omega_k$, $\tau=\tau_1\ldots\tau_k$ hold.

If T is a finitary homomorphic mapping of Σ^* into $P(\Delta^*)$ then T can be accomplished by a sequential transducer (K,Σ,Δ,H,s_0) such that $K=\{s_0\}$, and H consists of all quadruples (s_0,σ,θ,s_0) with σ in Σ and θ in $T(\sigma)$. We shall denote by S_1, S_2 and S_3, respectively, sequential transducers constructed in this way for the mappings T considered in Examples 1, 2 and 3.

To have an example of a transliteration system needing a more complicated sequential transducer for the accomplishment of the corresponding mapping, we shall consider in more detail the already mentioned second transliteration system from [7].

Example 4. Let Σ and Δ be as in the previous examples except that the Russian alphabet is supposed to be without capital hard and soft signs, and $\Delta\backslash\Sigma$ contains also the apostrophe besides the Latin letters. The

[1] This terminology is not universally adopted. For example the same term means something else in [3], and its present meaning is somewhat closer to the notion of finite transducer considered there (cf. Section 1.3.3 of that book, but note that there are certain omissions in the definitions of both notions in that section).

corresponding mapping T can be described for instance as follows: for any ω in Σ^* the set $T(\omega)$ has as its only element the string from Δ^* obtainable from ω by the application of a normal algorithm (in the sense of [6] and [4]) such that its scheme begins with the substitution formulas $\sigma\sigma_1\to\theta\sigma_1$, where either σ=“**Й**”, θ=“Jh”, or σ=“**й**”, θ=“jh”, and σ_1 is a Cyrillic vowel or a soft sign, and the further substitution formulas do the replacements listed in Example 1 except that

(a) “J” and “j” are used instead of “Y” and “y”, respectively, in the strings for the letters “**Ё**”, “**ё**”, “**Ю**”, “**ю**”, “**Я**”, “**я**”;

(b) there are no substitution formulas for “**Ъ**” and “**Ь**”;

(c) the substitution formulas for “**Й**”, “**й**”, “**Щ**”, “**щ**”, “**ъ**”, “**Ы**”, “**ы**” and “**ь**” are

Й→J, **й**→j, **Щ**→Xh, **щ**→xh, **ъ**→j’, **Ы**→Y, **ы**→y, **ь**→’

(of course such a normal algorithm would be not practically convenient, since its execution would require to read one and the same symbol many times; a more appropriate normal algorithm for the same transliteration system can be indicated whose substitution formulas contain an auxiliary symbol and whose execution actually performs a letter-by-letter transliteration from left to right).The mapping T can be accomplished by a sequential transducer $(\{k_0,k_1\},\Sigma,\Delta,H,k_0)$, where $k_0\neq k_1$. We shall indicate two such sequential transducers (both of them get into the state k_1 after reading “**Й**” or “**й**” and only in that case). The first one closely corresponds to the brief description we gave of the transliteration system in question. The set H of this sequential transducer consists of all quadruples of the following forms:

(i) $(k_0,\sigma,T(\sigma),k_0)$, where σ is in $\Sigma\backslash\{$“**Й**”,“**й**”$\}$;

(ii) $(k_i,\sigma,T(\sigma),k_1)$, where σ is “**Й**” or “**й**” (I=0,1);

(iii) $(k_1,\sigma,T(\sigma),k_0)$, where σ is in $\Sigma\backslash\{$“**Й**”,“**й**”$\}$, and σ is neither a vowel, nor a soft sign;

(iv) $(k_0,\sigma\sigma_1,T(\sigma)$“h”$T(\sigma_1),k_0)$, where σ is “**Й**” or “**й**”, and σ_1 is a Cyrillic vowel or a soft sign.

The set H of the second one consists of all quadruples of the forms (i) and (ii) above, as well as of all quadruples (k_1,σ,θ,k_0), where σ is in $\Sigma\backslash\{$“**Й**”,“**й**”$\}$, θ is $T(\sigma)$ if σ is not a vowel and not a soft sign, otherwise θ is $T(\sigma)$ preceded by “h”. (Note that the second components of all quadruples in this set are one-symbol strings, whereas it is not so for the set H of the first of the considered sequential transducers.). We shall denote the first and the second transducers considered in this example by $S_{4,1}$ and $S_{4,2}$, respectively.

As a further example on the application of sequential transducers to transliteration we shall indicate an extension of Uspensky’s transliteration system considered in Example 1. The extension in question can be used for reversible Russian-Latin transliteration of mixed texts — possibly containing not only Russian, but also Latin letters.[1]

Example 5. Let Δ consist of the capital and the small Latin letters and of characters commonly used both in English and in Russian, including the apostrophe, and Σ be obtained from Δ by adding to it all capital and small Russian letters. Let C be the mapping of Σ into Δ^* defined as follows: for any Russian letter σ, $C(\sigma)$ is its corresponding string from Example 1, $C(\sigma)=\sigma\sigma$ if σ is an apostrophe, and $C(\sigma)=\sigma$ for all other symbols σ in Σ. Let $D(\sigma)$ be $C(\sigma)$ preceded by an apostrophe. We consider a sequential transducer $S_5=(\{k_0,k_1\},\Sigma,\Delta,H,k_0)$, where $k_0\neq k_1$ and H consists of the following quadruples:

(i) all quadruples $(k_i,\sigma,C(\sigma),k_i)$, where I = 0 and σ is not a Latin letter, or I =1 and σ is not a Russian letter;

[1] A document accessible from http://www.metodii.com indicates another reversible Russian-Latin transliteration system with some similar features.

(ii) all quadruples (k_i,σ,D(σ),k_{1-i}), where I = 0 and σ is a Latin letter, or I =1 and σ is a Russian letter.

Let T be the mapping accomplished by this transducer. Evidently T is total and single-valued. This mapping will be shown also to be injective, but we prefer to postpone the corresponding proof to the last section.To illustrate the action of this mapping, let us apply it to the string

“**Операционная система** Windows 2000 **создана раньше системы** Windows XP.”

The image of this string looks as follows:

“Operacionnaya sistema **’**Windows 2000 **’**sozdana ranjshe sistemih **’**Windows XP.”

As a second illustration, we note that the string “**О’Нил** (O’Neill)” has the image “O**’’**Nil (**’**O**’’**Neill)” under T.

Remark 1. In each of the examples 1, 2 and 3, the mapping T^{-1} is not homomorphic although T is homomorphic and finitary. For instance T^{-1}(“sh”)={”ш”}, whereas T^{-1}(“s”)T^{-1}(“h”) is empty, in any of these examples. However, if a mapping T is accomplished by a sequential transducer S then the corresponding mapping T^{-1} is accomplished by the inverse sequential transducer S^{-1} (cf. Exercise 5 in [2, Section 3.3]). In particular, T^{-1} is accomplished by some sequential transducer in any of the above examples. Therefore the complexity of transliteration and of the inverse transformation can be considered in a uniform way by studying the complexity of using an arbitrary sequential transducer.

Input and Output Tokenizers of a Sequential Transducer

We shall need a notion that is similar to the notion of sequential transducer, but is somewhat simpler. We shall call a *tokenizer* any quadruple (K,Γ,G,s_0), where Γ is an alphabet, K is a finite set (the set of the *states*), s_0 is an element of K (the *start state*), G is a finite set of triples (the *moves*) with second components in Γ* and first and third components in K. For any t_0 in K, we shall call a *path of* (K,Γ,G,s_0) *starting at* t_0 any finite sequence

$$t_0,\psi_1,t_1,\psi_2,t_2,\ldots,t_{m-2},\psi_{m-1},t_{m-1},t_{m-1},\psi_m,t_m \qquad (1)$$

such that the triples (t_0,ψ$_1$,t_1), (t_1,ψ$_2$,t_2), …, (t_{m-2},ψ$_{m-1}$,t_{m-1}), (t_{m-1},ψ$_m$,t_m) belong to G (the case of m=0, i.e. of the one-term sequence consisting only of t_0, is also admitted). The string $\psi_1\psi_2\ldots\psi_{m-1}\psi_m$ will be called *the result* of this path (in the case of m=0 the result is empty). A path of (K,Γ,G,s_0) starting at the state s_0 will be called a *tokenization by* (K,Γ,G,s_0) of its result.

To any sequential transducer (K,Σ,Δ,H,s_0) two tokenizers will be made to correspond — its *input tokenizer* (K,Σ,H_1,s_0) and its *output tokenizer* (K,Δ,H_2,s_0), where H_1 and H_2 consist of the triples (k,ω,k') and (k,τ,k'), respectively, corresponding to the quadruples (k,ω,τ,k') in H. The input and the output tokenizers of any sequential transducer S will be denoted by IN S and OUT S, respectively.

Example 6. The strings “**тайна**”, “**рай**” and “**район**” have, respectively, the following tokenizations by the tokenizer IN $S_{4,1}$:

k_0,“**т**”,k_0,“**а**”,k_0,“**й**”,k_1,“**н**”,k_0,“**а**”,k_0,

k_0,“**р**”,k_0,“**а**”,k_0,“**й**”,k_1,

k_0,“**р**”,k_0,“**а**”,k_0,“**йо**”, k_0,“**н**”,k_0.

Example 7. T he string “O’’Nil (’O’’Neill)” has the following tokenization by OUT S_5:

k_0,“O”,k_0,“’”,k_0,“N”,k_0,“i”,k_0,“l”,k_0,“ “,k_0,“(”,k_0,“’O”,“’’”,k_1,“N”,k_1,“e”,k_1,“i”,k_1,“l”,k_1,“l”,k_1,“)”,k_1.

Next statement is obvious, and it indicates a way for applying the input and output tokenizers to the problems of single-valuedness and injectivity of the mapping accomplished by a sequential transducer.

Main sufficient conditions for single-valuedness and for injectivity. Let T be the mapping accomplished by a sequential transducer (K,Σ,Δ,H,s_0). Then T is surely single-valued if the following two conditions are satisfied:

(i) no string in Σ^* has two different tokenizations by IN(K,Σ,Δ,H,s_0);

(ii) for any s, s' in K and any ω in Σ^* there is at most one τ in Δ^* such that (s,ω,τ,s') belongs to H.

The mapping T is surely injective if the following two conditions are satisfied:

(iii) no string in Δ^* has two different tokenizations by OUT(K,Σ,Δ,H,s_0);

(iv) for any s, s' in K and any τ in Δ^* there is at most one ω in Σ^* such that (s,ω,τ,s') belongs to H.

The condition (i) is clearly satisfied in the case when (K,Σ,Δ,H,s_0) is the one-state sequential transducer that corresponds to a finitary homomorphic mapping of Σ^* into $P(\Delta^*)$, and the condition (ii) is equivalent in this case to the non-existence of σ in Σ with more than one element in $T(\sigma)$. In particular, both conditions are satisfied for the sequential transducers S_1 and S_3. These conditions are satisfied also for the sequential transducers $S_{4,1}$, $S_{4,2}$ and S_5, but the verification of (i) needs some care for the first of them. The conditions (iii) and (iv) are satisfied for all considered concrete sequential transducers S_1, S_2, S_3, $S_{4,1}$, $S_{4,2}$, S_5, however the verification of (iii) is somewhat cumbersome in all these cases.

The considerations below can make all above-mentioned verifications easier, except the one for the sequential transducer S_3 (but it corresponds just to the example of transliteration with a more difficult inverse transformation).

We shall introduce the left-to-right longest-match strategy (LRLMS) as an algorithm for transforming strings into tokenizations of them. Let a tokenizer (K,Γ,G,s_0) and a string θ from Γ^* be given. We shall consider a partial operation on the tokenizations by (K,Γ,G,s_0) of proper prefixes of θ, namely if

$$s_0,\varphi_1,s_1,\varphi_2,s_2,\ldots,s_{k-2},\varphi_{k-1},s_{k-1},\varphi_k,s_k \tag{2}$$

is such a tokenization then we look for a triple (s_k,φ,s) from G such that $\varphi_1\varphi_2\ldots\varphi_{k-1}\varphi_k\varphi$ is a prefix of θ with the maximal possible length, and if there is exactly one such triple then we append its components φ and s to the sequence (2). The following algorithm will be called *the LRLMS-algorithm*: given a string θ from Γ^*, we start with the one-term sequence consisting only of s_0, and we apply the above-mentioned partial operation until a tokenization of θ is obtained or no further application of the operation is possible. A termination of this process is considered as successful if a tokenization of θ is obtained.

A tokenization (2) by the tokenizer (K,Γ,G,s_0) will be said to be *compliant with the left-to-right longest-match strategy* (*LRLMS-compliant*, for short) if this tokenization can be obtained by applying to its result the LRLMS-algorithm. The condition is trivially satisfied if k=0, and for k≠0 it is equivalent to the following requirement: φ_k is non-empty, and there are no l in $\{1,\ldots,k\}$ and no triple (s_{l-1},φ,t) from G distinct from (s_{l-1},φ_l,s_l) such that φ is a prefix of $\varphi_l\varphi_{l+1}\ldots\varphi_{k-1}\varphi_k$, and φ_l is a prefix of φ.

Example 8. The two sequences below are tokenizations by OUT S_3 (of the strings “suhhoyy” and “ishhod”, respectively), but the first one is LRLMS-compliant, whereas the second one is not (due to the prefix “sh” in the string “shhod”):

$$s_0,\text{“s”},s_0,\text{“u”},s_0,\text{“hh”},s_0,\text{“o”},s_0,\text{“yy”},s_0,$$

$$s_0,\text{“i”},s_0,\text{“s”},s_0,\text{“hh”},s_0,\text{“o”},s_0,\text{“d”},s_0.$$

Remark 2. If there are no triples in G with empty second components, and we apply the LRLMS-algorithm to some string θ from Γ* that has no LRLMS-compliant tokenization, then the application of the LRLMS-algorithm terminates unsuccessfully. For instance, if (K,Δ,G,s_0) is OUT S_3,.then the application of the algorithm to the string “ishhod” terminates unsuccessfully at the sequence s_0,“i”,s_0,“sh”,s_0.

A tokenizer will be called *compliant with the left-to-right longest-match strategy* (*LRLMS-compliant*, for short) if all tokenizations by this tokenizer are LRLMS-compliant. Of course this implies the non-existence of two distinct tokenizations of one and the same string. It is easy to check that all considered concrete sequential transducers S_1, S_2, S_3, $S_{4,1}$, $S_{4,2}$, S_5 have input tokenizers that are LRLMS-compliant (the situation is not completely obvious only in the case of the sequential transducer $S_{4,1}$). Example 8 shows that OUT S_3 is not LRLMS-compliant. However, the output tokenizers of all other sequential transducers in question are LRLMS-compliant. This will be verified in the last section by means of corollaries of the necessary and sufficient condition below, where a state of a tokenizer is called *accessible* if it is the last term of some tokenization by this tokenizer.[1]

Necessary and sufficient condition for LRLMS-compliance of a tokenizer. A tokenizer (K,Γ,G,s_0) is LRLMS-compliant if and only if it has the following properties:

(i) there is no triple in G with accessible first component and empty second one;

(ii) no two triples in G with accessible first component exist that differ from one another only in their third components;

(iii) for any (t_0,φ,t) in G with accessible t_0 there is no path (1) in (K,Γ,G,s_0) with m>1 such that $\psi_1\psi_2\ldots\psi_{m-1}$ is a proper prefix of φ and φ is a prefix of $\psi_1\psi_2\ldots\psi_{m-1}\psi_m$.

Proof. Let (K,Γ,G,s_0) be an arbitrary tokenizer. For the proof of the necessity, suppose (K,Γ,G,s_0) is LRLMS-compliant. Let t_0 be an accessible element of K, and (2) be a tokenization by (K,Γ,G,s_0) such that $s_k=t_0$. There is no triple (t_0,φ,t_1) in G with empty φ – otherwise

$$s_0,\varphi_1,s_1,\varphi_2,s_2,\ldots,s_{k-2},\varphi_{k-1},s_{k-1},\varphi_k,t_0,\varphi,t_1 \qquad (3)$$

would be a tokenization that is not LRLMS-compliant. There are also no φ in Γ* and distinct t and t_1 in K such that both (t_0,φ,t) and (t_0,φ,t_1) belong to G – otherwise again (3) would be a tokenization that is not LRLMS-compliant. Finally, it is not possible that there are (t_0,φ,t) in G and a path (1) in (K,Γ,G,s_0) with m>1 such that $\psi_1\psi_2\ldots\psi_{m-1}$ is a proper prefix of φ and φ is a prefix of $\psi_1\psi_2\ldots\psi_{m-1}\psi_m$ – this would contradict the LRLMS-compliance of the tokenization

$$s_0,\varphi_1,s_1,\varphi_2,s_2,\ldots,s_{k-2},\varphi_{k-1},s_{k-1},\varphi_k,t_0,\psi_1,t_1,\psi_2,t_2,\ldots,t_{m-2},\psi_{m-1},t_{m-1},\psi_m,t_m$$

[1] All concrete tokenizers we mentioned have only accessible states. On the other hand, from an arbitrary tokenizer we can get one having only accessible states by elimination of the states that are not accessible and of the moves that contain such states as first or third components. It is easy to see that the reduction in question will not affect the set of the tokenizations by the tokenizer.

(since ψ_1 is a proper prefix of φ). For proving the sufficiency, suppose (K,Γ,G,s_0) is not LRLMS-compliant. Then some tokenization (2) by (K,Γ,G,s_0) is not LRLMS-compliant, hence $k\neq 0$ and either φ_k is empty or there are some l in $\{1, \dots, k\}$ and some triple (s_{i-1},φ,t) from G distinct from (s_{i-1},φ_i,s_i) such that φ_i is a prefix of φ and φ is a prefix of $\varphi_i\varphi_{l+1}\dots\varphi_{k-1}\varphi_k$. In the first case the triple (s_{k-1},φ_k,s_k) violates (i). In the second case, either $\varphi=\varphi_i$, and then the pair of triples (s_{i-1},φ,t) and (s_{i-1},φ_i,s_i) violates (ii), or φ_i is a proper prefix of φ. If φ_i is a proper prefix of φ, then $k>i$ and there is some j in $\{l+1, \dots,k-1,k\}$ such that φ is a prefix of $\varphi_i\varphi_{l+1}\dots\varphi_{j-1}\varphi_j$, whereas $\varphi_i\varphi_{l+1}\dots\varphi_{j-1}$ is a proper prefix of φ. In this case we can violate (iii) by setting $m=j-l+1$, $t_0=s_{i-1}$ and $t_r=s_{l-1+r}$, $\psi_r=\varphi_{l-1+r}$ for $r=1,\dots,m$. ∎

Corollary 1. Let (K,Γ,G,s_0) be a tokenizer such that no triple from G has an empty second component and there are no two triples in G that differ from one another only in their third components. Let no triples (t_0,φ,t), (t_0,ψ_1,t_1) and (t_1,ψ_2,t_2) exist in G such that ψ_1 is a proper prefix of φ and some of the strings $\psi_1\psi_2$ and φ is a prefix of the other one. Then (K,Γ,G,s_0) is LRLMS-compliant.

Proof. Suppose there are a triple (t_0,φ,t) in G and a path (1) in (K,Γ,G,s_0) with $m>1$ such that $\psi_1\psi_2\dots\psi_{m-1}$ is a proper prefix of φ and φ is a prefix of $\psi_1\psi_2\dots\psi_{m-1}\psi_m$. Then ψ_1 is also a proper prefix of φ, and some of the strings $\psi_1\psi_2$ and φ is a prefix of the other one. Since (t_0,ψ_1,t_1) and (t_1,ψ_2,t_2) belong to G, this is a contradiction. ∎

Next corollary is actually a particular instance of Corollary 1.

Corollary 2. Let $(\{s_0\},\Gamma,G,s_0)$ be a (one-state) tokenizer, and let W be the set of the second components of the triples from G. Suppose that all strings from W are non-empty, and there are no φ, ψ_1 and ψ_2 in W such that ψ_1 is a proper prefix of φ and some of the strings $\psi_1\psi_2$ and φ is a prefix of the other one. Then $(\{s_0\},\Gamma,G,s_0)$ is LRLMS-compliant.

Some concrete applications

Suppose two alphabets Σ and Δ are given, and T is the mapping of Σ^* into $P(\Delta^*)$ describing a given transliteration system. We shall call this transliteration system *easily usable* if T is an injective mapping that can be accomplished by some sequential transducer with LRLMS-compliant input and output tokenizers. The transliteration systems mentioned in Examples 1, 2, 4 and 5 are easily usable in the above sense, and this will be shown by verifying that any of the sequential transducers S_1, S_2, $S_{4,1}$, $S_{4,2}$, S_5 has LRLMS-compliant input and output tokenizers and satisfies item (iv) from the main sufficient conditions for single-valuedness and for injectivity (of course it would be enough to do this for one of the sequential transducers $S_{4,1}$ and $S_{4,2}$ instead of doing it for both of them).

The verification of (iv) for each of the above-mentioned sequential transducers is almost immediate (even in the case of S_5, although S_5 has moves with distinct second components and one and the same third one – for instance the quadruples (k_0,“**Д**”,“D”,k_0) and (k_1,“D”,“D”,k_1) or the quadruples (k_0,“D”,“'D”,k_1) and (k_1,“**Д**”,“'D”,k_0)).

By their construction, the sequential transducers in question have no moves with empty second or third components, hence the corresponding input and output tokenizers have no moves with empty second components.

The LRLMS-compliance of the input tokenizers of the considered sequential tokenizers was already characterized as more or less obvious. However, there is no problem to verify it also by means of Corollary 1 or (for the case of S_1 and S_2) Corollary 2. For S_1, S_2, $S_{4,2}$ and S_5 the verification is trivial thanks to the fact that all second components of their moves have length 1. Now let us suppose that $(t_{0,}\varphi,t)$, (t_0,ψ_1,t_1) and (t_1,ψ_2,t_2) are moves of the tokenizer IN $S_{4,1}$ such that ψ_1 is a proper prefix of φ and some of the strings $\psi_1\psi_2$ and φ is a prefix of the other one. Then $(t_{0,}\varphi,t)$ must have the form $(k_0,\sigma\sigma_1,k_0)$, where σ is "**Й**" or "**й**", and σ_1 is a Cyrillic vowel or a soft sign, hence ψ_1 is "**Й**" or "**й**", t_1 is k_1, and ψ_2 begins with a Cyrillic vowel or a soft sign. From the fact that t_1 is k_1 the conclusion follows that ψ_2 is a symbol of Σ which is neither a Cyrillic vowel nor a soft sign, and this is a contradiction.

The LRLMS-compliance of OUT S_1 and OUT S_2 can be shown again by means of Corollary 2. Since all moves of S_1 are also moves of S_2, it would be sufficient to check the assumption of Corollary 2 only for OUT S_2. Suppose the set W for this tokenizer has elements φ, ψ_1 and ψ_2 such that ψ_1 is a proper prefix of φ and some of the strings $\psi_1\psi_2$ and φ is a prefix of the other one. The assumptions that φ and ψ_1 belong to W and ψ_1 is a proper prefix of φ imply the equality $\varphi = \psi_1$"h". From here, taking into account also the assumptions that ψ_2 belongs to W and some of the strings $\psi_1\psi_2$ and φ is a prefix of the other one, we get a contradiction by firstly concluding that ψ_2 begins with "h".

As to the LRLMS-compliance of the output tokenizers of $S_{4,1}$, $S_{4,2}$ and S_5, it can be shown by means of Corollary 1. It is straightforward (although somewhat tedious) to see that no of these tokenizers has two moves differing from one another only in their third components. Now suppose for some of this tokenizers the existence of moves $(t_{0,}\varphi,t)$, (t_0,ψ_1,t_1) and (t_1,ψ_2,t_2) such that ψ_1 is a proper prefix of φ and some of the strings $\psi_1\psi_2$ and φ is a prefix of the other one. We shall consider the cases of $S_{4,1}$, $S_{4,2}$ and S_5 one by one. By reasoning in two steps, in any of these three cases we shall get a contradiction to the assumption that (t_1,ψ_2,t_2) is a move of the corresponding output tokenizer – in the first step we shall make some conclusions from the assumptions that $(t_{0,}\varphi,t)$, (t_0,ψ_1,t_1) are moves of the tokenizer in question and ψ_1 is a proper prefix of φ, whereas in the second step we shall take into account also that some of the strings $\psi_1\psi_2$ and φ is a prefix of the other one.

In the case of $S_{4,1}$ we conclude in the first step that either the first symbol in φ after its prefix ψ_1 is "h" or we have the equalities $t_1= k_1$, $\varphi=\psi_1\sigma$, where σ is some of the letters "a", "o", "u" or an apostrophe. Making use of this conclusion, we get a contradiction in the second step by inferring that either the string ψ_2 begins with "h" or in the presence of the equality $t_1= k_1$ this string begins with some of the letters "a", "o", "u" or with an apostrophe.

The first step in the case of $S_{4,2}$ is to conclude that we have either the equalities $t_1=k_0$, $\varphi=\psi_1$"h" or the equalities $t_1= k_1$, $\varphi=\psi_1\sigma$, where σ is some of the letters "a", "o", "u". We get a contradiction in the second step by inferring that ψ_2 begins with "h" in the presence of the equality $t_1=k_0$ or with some of the letters "a", "o", "u" in the presence of the equality $t_1= k_1$.

In the case of S_5 we firstly conclude that $t_1=k_0$, $\varphi=\psi_1$"h", and then we get a contradiction by inferring that ψ_2 begins with "h".

Remark 3. Instead by reasoning as above, each of the considered conditions could be verified by straightforward inspection of all finitely many possible cases. Of course, it would be better to do this by using some appropriate computer program.

Remark 4. As we observed (by using Example 8), the tokenizer OUT S_3 is not LRLMS-compliant. This statement can be strengthened as follows: the mapping accomplished by S_3 cannot be accomplished at all by a sequential transducer with a LRLMS-compliant output tokenizer (hence the transliteration indicated in Example 3 is not easily usable in our sense). In fact, if we suppose that such other sequential transducer can be constructed then we can get a contradiction by considering the application of the LRLMS-algorithm for the corresponding output tokenizer to strings consisting of one "s" followed by arbitrarily many "h". Namely, we can show then the existence of a non-empty string in Σ* whose image will be a prefix of all sufficiently long strings of the above-mentioned form, and obviously such a string cannot exist.

Acknowledgments

The author thanks V. A. Uspensky and M. R. Pentus for some very useful discussions and very helpful information.

Bibliography

[1] Gerdemann, D., van Noord, G. Transducers from rewrite rules with backreferences. In: Ninth Conference of the European Chapter of the Association for Computational Linguistics (8-12 June 1999, Univ. of Bergen, Bergen, Norway). San Francisco, Morgan Kaufmann Publishers, 1999, 126-133. http://acl.ldc.upenn.edu/E/E99/

[2] Ginsburg, S. The Mathematical Theory of Context-Free Languages. McGraw-Hill, 1966.

[3] Lothaire, M. Algebraic Combinatorics on Words, Cambridge University Press, 2002.

[4] Markov, A. A., Nagorny, N. M. The Theory of Algorithms. Dordrecht etc., Kluwer Academic Publishers, 1988.

[5] Sardinas, A., Patterson, C. A necessary and sufficient condition for the unique decomposition of coded messages. IRE Intern. Conv. Record, 8:104-108, 1953.

[6] Марков, А. А. Теория алгорифмов. Труды Математ. инст. им. В. А. Стеклова, 42, Москва-Ленинград, Изд. АН СССР, 1954.

[7] Успенский, В. А. К проблеме транслитерации русских текстов латинскими буквами. In: Научно-техническая информация, серия 2, Информационные процессы и системы. 1967, № 7, 12-20 (also in [9], 390-412).

[8] Успенский, В. А. Невто́н-Ньюто́н-Нью́тон, или Сколько сторон имеет языковой знак? In: Русистика. Славистика. Индоевропеистика. Сборник к 60-летию Андрея Анатольевича Зализняка. Москва, "Индрик", 1996, 598-659 (also in [9], 483-561).

[9] Успенский, В. А. Труды по нематематике. Москва, ОГИ, 2002. http://www.mccme.ru/free-books/usp.htm

Author's Information

Dimiter Skordev – Sofia University, Faculty of Mathematics and Informatics, blvd. J. Bourchier 5, Sofia 1164, Bulgaria; e-mail: skordev@fmi.uni-sofia.bg

TABLE OF CONTENTS OF IJ ITA, VOLUME 16, NUMBER 1