ITHEA

International Journal INFORMATION MODELS & ANALYSES

2014 Volume 3 Number 3

International Journal
INFORMATION MODELS & ANALYSES
Volume 3 / 2014, Number 3

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

IJ IMA rules for preparing the manuscripts are compulsory.
The **rules for the papers** for ITHEA International Journals are given on www.ithea.org .
The camera-ready copy of the paper should be received by ITHEA® Submission system http://ij.ithea.org .
Responsibility for papers published in IJ IMA belongs to authors.

International Journal "INFORMATION MODELS AND ANALYSES" Volume 3, Number 3, 2014

Edited by the **Institute of Information Theories and Applications FOI ITHEA**, Bulgaria, in collaboration with
Institute of Mathematics and Informatics, BAS, Bulgaria,
V.M.Glushkov Institute of Cybernetics of NAS, Ukraine,
Universidad Politechnika de Madrid, Spain,
Hasselt University, Belgium
Institute of Informatics Problems of the RAS, Russia,
St. Petersburg Institute of Informatics, RAS, Russia
Institute for Informatics and Automation Problems, NAS of the Republic of Armenia,

Publisher: **ITHEA®**

Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org, e-mail: info@foibg.com
Technical editor: **Ina Markova**

Printed in Bulgaria

ISSN 1314-6416 (printed) **ISSN 1314-6432 (Online)**

ANALYSIS FINITELY SMALL VALUES IN COMPUTER SIMULATION

Oleksii Voloshyn, Vladimir Kudin

Abstract*: We review the development of the infinity concept and continuous mathematics from ancient times to the present day. To the critical analysis of the application of the infinitesimal in the mathematical modeling in XVII-XVIII centuries is given special attention. It is noted necessity of development methods for the analysis of finite small disturbances in computer simulations. The problem of the adequacy of the linear approximation of mathematical and computer models is under discussion. We investigate the influence of finite small disturbances into elements of model to quality of localization solutions of the model which is based on the methodology of options sequential analysis. The results of numerical experiments and taken into account the level of systems conditionality and algorithms (the mantissa length in the numbers representation) are given.*

Keywords*: continuous and infinite, infinitesimal analysis, mathematical and computer simulation, linear approximation, localization, successive analysis variants, analysis finitely small values*

ACM Classification Keywords*: H.4.2 Information Systems Applications: Types of Systems: Decision Support.*

Introduction

Together with the invention of the Pythagorean School of incommensurability quantities in mathematics included the concept of infinity [Yushkevich, 1970]. The study of infinite sets and infinite sequences causes the need to solve two major problems that faced the ancient mathematics. This is - the problem of the real number (the modern name of the problem of incommensurability). And the problem measures (problems of actual infinity). Aporia Zeno of Elea (V century BC) raised profound and complex matters that are contained in terms of continuous and infinite, and that never cease to attract the attention of modern scientists and philosophers.

Leave aside all aspects of the problem except one - the problem of the relation of the mathematical model and the real physical world, having agreed with David Hilbert and P. Barnice ("Fundamentals of Mathematics", 1934r.): "... infinity was not given to us, and there was only interpolated or extrapolated by some intellectual process".

Eventually therefore for us the most important issue is the adequacy of the model and the simulated object (process) as well as an intermediate is - a question the correctness of the choice of means of a mathematical model, and in this context - is it permissible to use actual infinite quantities (infinitely large and infinitely small).

Antique mathematics rejected actual infinity - "... the denial of actual infinity does not take away their mathematicians theory, because they do not need such an infinite and do not use it ..." (Aristotle); "... At least there is a something small, but there is always an even smaller" (Anaxagoras). General teaching about relationships and rigorous methods of limiting transitions were created by Eudoxus of Cnidus (406-355 years. BC). "Banishment" of actual infinite quantities was carried out following the axiom (known as the Archimedean axiom): "The quantities are related to each other if they are taken Multiples can outdo each other". Eudoxus relations theory is used by mathematics essentially to the end of the nineteenth century, after announcing on Newton relations numbers [Yushkevich, 1970]. In the second half of the nineteenth century Eudoxus theory was developed by R.Dedekind. Between these theories there is so profound analogy. In one of the letters R.Lipshitz asked R.Dedekind [Yushkevich, 1970, P.97] what he did that was new in comparison with the ancient

mathematicians (unlike relations Eudoxus of the modern notion of a real number is a system of relations that does not form a field but the group in terms of formal mathematics).

Difficulties associated with the actual infinity were not overcome the theory of G.Cantor sets (70th years of the nineteenth century.) They only changed the form and appeared as the paradoxes of set theory. In modern times there is a point of view according to which the free handling with the actual infinite sets, even denumerable one, is illegal (authors hold to this point of view as well). Extension of the concept of number was explicitly in Islamic countries in XI-XIII centuries (in Europe in XVI-XVII centuries). In both cases, it is in close connection with the development of computing technology and computers [Yushkevich, 1971]. Principal means of mathematical research in the natural sciences "New time" (XVI-XVII centuries) became infinitesimal methods, although understanding of the problems associated with their use by that time mathematicians.

Some prominent mathematicians concluded (apparently implying G.Leibniz thesis - "Useless paradoxes do not exist"), and that "... the error compensation is the driving force of the infinitesimal analysis and is at the same time strictly scientific method of cognition" [Yushkevich, 1973]. Extreme position in the debate about the infinitesimal was occupied by English philosopher Bishop George Berkeley (1685-1753 years): "... a careful study of affect that in any way is not necessary to use the infinitely small parts or quantities represent them smaller than the smallest felt by." Scientists such rank as Eyler, Lagrange and others had to reckon with witty and largely fair criticism Berkeley [Yushkevich, 1972, P.259].

From the perspective of Berkeley, infinite, as sensually imperceptible, have no real existence, such as one ten-thousandth of a piece of inch, and the more there can be no question of the infinite divisibility of any extended value. Two centuries later, in quantum physics, the notion "Planck level" - the smallest physically defined distance (equal to 10^{-35} m). It is known view of Academician A.N. Kolmogorov that in principle there is no number 10^{-100}.

In XIX-th century to construction of models of "wildlife" at research of biological, economic and social processes the analysis of the infinitesimal beginnings to be applied. And if at modeling of objects of "the lifeless nature" (in the physicist and, first of all, in the mechanic) by means of the differential equations the concept of influence on values of function of infinitesimal changes of argument "more or less adequately displays "physical" essence of process at modeling of social and economic processes it basically is incorrect. So, research of influence of infinitesimal change of the price of the goods (10^{-5}\$ or 10^{-35}\$) on demand for it looks as absurdity. Therefore differential models of change of economic parameters are confidants, very "rough". Failures in forecasting of development of social and economic processes appreciably speak this circumstance.

The second factor, which forces to reconsider the relation to infinitesimal models and methods, is computer facilities development in second half XX-th century. Wide use of simulating modeling of the numerical decision of analytical mathematical models have led to that the mathematics has returned "into place" and became, as well as 25 centuries ago, on the substance of final, accuracy of calculations is defined now by length of a mantissa of representation of numbers the computer.

Also does not play a special role what it is - 32, 64 or 1024 digit. Language „ε - δ" here does not work - that from this, what if for $\varepsilon = 0.01$, exists $\delta = 10^{-100}$? How numerically to check up criterion of Teatet (IV century BC) in commensurabilities of two segments [Yushkevich, 1970, C.77]: „If two segments are commensurable, the algorithm of Euclide for final number of steps finds their greatest general measure that is if the algorithm appears infinite segments are incommensurable"? What operating time of the computer to accept for "infinite" - 10, 10^{10}, 10^{100} seconds?

Therefore last decades "the classical" scheme of mathematical modeling "Object - Mathematical model - the Numerical decision" (O-MM-ND) extends at the expense of introduction of "Computer model" (CM) which basically is finit: "O-MM-CM-ND". In 80th years of XX century academician N. N. Moiseev asserted [Moiseyev,

1981] that "… the next decades it is necessary to give the basic attention to research to computer (simulation) models". Now researches of computer models are in a stage of development and are far from successful end.

The analysis finitly small influences in linear models

At the heart of differential and integral calculus (based on infinitesimal methods) the idea of linearization - replacement so difficult as far as possible (nonlinear) change of investigated process by linear one (tangent) is laid. Speaking in images, I.Newton and G.Leibniz „linearized the world". Therefore research of linear models (first of all, systems of the linear algebraic equations and inequalities) was and remains to one of the basic problems in mathematical and computer modeling, and the decision of practical problems causes of consideration of systems of the large dimension, systems with incomplete, inexact, indistinct parameters.

On the other hand, linear models (LM) are only the first approach, local approximation, at research of processes of the real world which is basically nonlinear. For well conditioned problems (even in "a smooth" situation, and in certain cases - and in "rough" [Clark, 1988]) the existing mathematical apparatus allows to estimate influence of small indignations of parameters on properties of model only locally.

However, in many cases approximating linear models are described in a class of systems poorly conditioned (ill-posed) problem, in particular, systems of the linear algebraic equations (SLAE) with the square and strongly filled matrix of restrictions [Samarskiy, 1989; Metjuz, 2001; Demmel, 2001]. Even at small dimension of model and in the absence of structural singularities in a matrix of restrictions of association of a solution from perturbations are approximated insufficiently adequately (It is localized by means of ellipses with axes of essentially various length or parallelepipeds with essentially different boundaries on variables). It is known that the category of poor conditionality is defining in construction of solutions associations (and localization areas) from perturbations in model elements [Samarskiy, 1989].

Experience of a solution of practical problems shows that for mathematical modeling inadequacy of model and real modeled process (appearance simplification, an inaccuracy of the representation of parameters), and for computing experiment - inadequacy of mathematical and computer model (ineradicable errors, errors of digitization, an error of a method, rounding off, truncations, losses of significant figures is characteristic, at performance of operations). The rounding off of numbers to within the fixed value at evaluations is connected with the approached representation of numbers with the final (truncated) digit grid (numbers with fixed and a floating point). "The classical" continuity guarantees existence local somehow a small neighborhood of a solution. (Problems of necessity of definition of a solution with any degree of accuracy now are not considered). In case of computer (discrete) model the solution neighborhood cannot be less than in advance set number so, is not correct local approximation. Round-off errors for such evaluations can be characterized by relative error. During performance of the basic arithmetical operations of an error of rounding off can collect (in большей or a smaller measure). There are certain differences in the scheme of accumulation of errors at performance of evaluations with the fixed and floating numbers. In aggregate they are summarized as an error of evaluations [Samarskiy, 1989; Metjuz, 2001; Demmel, 2001].

Algorithms which realize concrete methods, can be steady (errors during evaluations collect slightly) and, accordingly, otherwise - unstable. To algorithms which realize methods, it is possible to show two groups of criteria: criteria of adequacy of models (discrete computer and mathematical continuous) and criteria of convergence of computing algorithms at magnification of number of the equations in exposition of mathematical model, speed, accuracy of evaluations. „A phenomenon of Runge" [Metjuz, 2001] gives representation about problems of inadequacy of mathematical and machine models. It consists of volume that at interpolating the approximation error grows at magnification of an amount of knots [Voloshyn, 2010].

Such concept is connected with convergence of algorithms as a correctness of a numerical method of solution (continuous association on input datas, uniform concerning an amount of the equations). In particular, a correctness as the stability at modifications of input datas in linear systems, is characterised as a stability from input datas of right members (conditionality) and from all elements of model. The measure of a correctness of a problem is quantitatively described by a condition number [Metjuz, 2001]. It is necessary to notice that the category of a correctness of linear system (conditionality of problems) is exhibited on each pitch of a repetitive process.

In a continuous case the basic criterion of convergence is the monotonicity and boundedness of estimations of a solution, in case of computer model the "numerical" monotonicity and boundedness of "computer" time does not guarantee lack of "ejections" in "the real [Voloshyn, 2011].

In the conditions of an incorrectness (for the Hadamard) [Samarskiy, 1989] small inaccuracies of representation of mathematical model, in particular, at level of computing algorithm, can essentially influence quantitative and qualitative performances of a received solution at use of a concrete method (algorithm) [Voloshyn, Kudin, 2010; Bogaenko, Kudin, 2010]. Such inaccuracies are caused by boundedness of length of a mantissa at representation of numbers from a floating point (an error of truncation, a rounding off) more often. Despite of presence of the COMPUTER with the effective organisation of operation of a rounding off, to the full to avoid them, or to improve known theoretical estimations, it is not possible. It is important to notice that carrying out of evaluations with low accuracy ("roughly") smoothes ("hides") some important details.

In such cases there is an illusion of non-uniqueness of a solution (an incomplete rank) for SLAE or compatibility of area of solutions for SLAU. High-precision evaluations uncover details and more often "give" a unique solution (a full grade of a matrix) for Slough and incompatibility for SLAU. At carrying out of the analysis of evaluations it is important to co-ordinate with different degree of accuracy (to co-ordinate among themselves) properties of rough and high-precision evaluations.

Sequential analysis of variants

In the given work outcomes of research of association of quality of localization of area of a membership of solutions from magnitude of perturbations in model elements, from a condition number (as defining factor) and the organizations of evaluations (from length of a mantissa representation of numbers) are reduced. The attention to complexities of construction of the sets localizing set of solutions is focused. The methodology of a sequential analysis of variants (SAV) [Voloshyn, 1987] and a method of basis matrixes (MBM), as its concrete realization [Kudin, 2002] are in the heart of researches. The general formalism of SAV is offered by V.S.Mihalevich in [Mihalevich, 1965] where the general scheme of SAV on the basis of generalization of idea of the theory of consecutive solutions of A.Vald and R.Bellman dynamic programming is described. The scheme of SAV is reduced to the following sequence of procedures which repeat:

1. The partition of set of variants of solutions of a problem on some subsets, each of which has any specific properties;
2. Use of these specific properties for determination of logic inconsistencies in exposition of separate subsets;
3. Elimination of the further reviewing of those subsets of variants of a solution, in which exposition logic inconsistencies.

The technique of a sequential analysis and elimination of variants consists in such mode of construction of research which allows to "eliminate" initial segments of variants before their full construction are unpromising.

Thus, the considerable economy of computing expenditures as at elimination of unpromising initial parts of variants all possible sets sub variant their prolongations are eliminated also is carried out.

The general scheme of SAV is rendered concrete for various classes of multiple problems in the form of algorithms „a sequential analysis, eliminations and designing of variants" [Voloshyn, 1987] widely known algorithm „the Kiev broom" [Mihalevich, Shor, 1961]. „The principle of monotone recursiveness", related to criterion of an optimality of a dynamic programming of R.Bellman was the key rule of elimination of unpromising variants in these algorithms.

Simultaneously with known advantages, algorithms of step by step designing have also certain shortages. In development of the general concepts of SAV in a series of works of V.S.Mihalevich, V.L.Volkovich and A.F.Voloshyn procedures of parallel elimination подвариантов, in particular, known algorithm W (Volkovich-Voloshyn) [are offered Volkovich, Voloshyn, 1978].

Thus there is a problem of designing of a full variant which dares by consecutive input of restrictions on values of an objective functional and the problem of designing of a full variant is reduced to construction of procedures of the analysis and elimination under variants. Efficiency of the algorithms based on offered principles, proves to be true computing experiments, theoretical estimations and a solution of practical problems [Voloshyn, 2013].

The basic concept in application of algorithmic schemes, are based on decomposition schemes of SAV [Voloshin, 1987] which can be described in spaces of the various nature (values of variables and/or the criteria, the limited amount of variants of the solution, the limited amount of restrictions of a problem, etc. [Voloshyn, 2013]) consists in "localization" of area of an optimum of a target functional.

The method of basis matrices (MBM)

MBM allows to consider the effect of small perturbations on the solving the SLAE. In contrast to the classical iterative methods (such as the simple method) MBM finds a solution in two stages: 1) selection of basis matrix; 2) finding the appropriate selection of basis matrix solving the linear analysis and conditioning system that allows the situation includes the accumulation of errors.

Discusses the linear model:

$$Au = C, \tag{1}$$

where the matrix A (with rows $a_1, a_2, \dots, a_m$, $a_j = (a_{j1}, a_{j2}, \dots, a_{jm})$, $j = \overline{1,m}$) is square with dimension $(m \times m)$, in which the vectors of variables $u = (u_1, u_2, \dots, u_m)^T$ and constraints $C = (c_1, c_2, \dots, c_m)^T$ have dimension m. Constraint matrix of the SLAE are characterized, in general, large sizes, heavily stocked and poor conditioning. MBM is based on the idea of the base matrix. Basis matrices during the iterations vary sequential substitution lines of an auxiliary base matrix (auxiliary SLAE) rows limitations of the model (1). Consider the vectors $a_{i1}, a_{i2}, \dots, a_{im}$ it's normal limits (referred to as strings), $a_j u^T \le c_j$, $j \in J_б$, which form a matrix $A_б$ where $J_б = \{i_1, i_2, \dots, i_m\}$ is the indices of restrictions.

Definition 1. Square matrix $A_б$ consisting of m linearly independent rows of an auxiliary linear algebraic equation (SLAE) is called an artificial base and the solution corresponding system of equations $A_б u = C^0$, where $C^0 = (c_{i_1}, c_{i_2}, \dots, c_{i_m})^T$ is artificial basic solution.

Definition 2. Two basic matrices that differ only in one line are called adjacent.

Let e_{ri} is the elements of the matrix A_{δ}^{-1}, which is inverse to A_{δ} and $e_k = (A_{\delta}^{-1})_k$ is k- th column of the inverse matrix. $u_0 = (u_{01}, u_{02}, ..., u_{0m})^T$ is basic solution. $\alpha_r = (\alpha_{r1}, \alpha_{r2}, ..., \alpha_{rm})$ is a decomposition vector of normal restrictions $a_r u \le c_r$ on $a_r u \le c_r$ the basic lines of the matrix A_{δ}. $\Delta_r = a_r u_0 - c_r$ is the residual r-th constraints u_0. The line a_l (normal limits $a_l u \le c_l$, $l \notin J_{\delta}$). $\alpha_l = (\alpha_{l1}, \alpha_{l2}, ..., \alpha_{lm})$ is a vector of decomposition a_l row by rows of basis matrix A_{δ} , which is a vector of relations $a_l = \alpha_l A_{\delta}$.

We have the relations between the expansion coefficients of the normal restrictions (1) of the basic lines of the artificial matrix and elements: of inverse matrices, of basic solutions, of residuals of constraints (1) in two adjacent basis matrices replacing k-th row of the base matrix A_{δ} , and of a_l string [Voloshyn, Kudin, 2010]:

$$\overline{\alpha}_{rk} = \frac{\alpha_{rk}}{\alpha_{lk}}, \ \overline{\alpha}_{ri} = \alpha_{ri} - \frac{\alpha_{rk}}{\alpha_{lk}}\alpha_{li}, \quad r = \overline{1,m}; \ i = \overline{1,m}; \ i \ne k; \tag{2}$$

$$\overline{e}_{rk} = \frac{e_{rk}}{\alpha_{lk}}, \ \overline{e}_{ri} = e_{ri} - \frac{e_{rk}}{\alpha_{lk}}\alpha_{li}, \quad r = \overline{1,m}; \ i = \overline{1,m}; \ i \ne k; \tag{3}$$

$$\overline{u}_{0j} = u_{0j} - \frac{e_{jk}}{\alpha_{1k}}\Delta_l, \ j = \overline{1,m}, \tag{4}$$

$$\overline{\Delta}_k = -\frac{\Delta_l}{\alpha_{lk}}, \ \overline{\Delta}_r = \Delta_r - \frac{\alpha_{rk}}{\alpha_{lk}}\Delta_l, \ r = \overline{1,n}; \quad r \ne k. \tag{5}$$

It should be noted condition of a regularity basis matrix by replacing a_l line to k-th row of the base matrix A_{δ} is under condition of the inequality $\alpha_{lk} \ne 0$. For the existence of a unique solution of (1) is necessary and sufficient to be carried out the condition $\alpha_{lk}^{(i)} \ne 0, \ i = \overline{1,m}$; where $\alpha_{lk}^{(i)}$ is the key elements of the operation of substitution strings basis matrix normal constraints (1). A matrix of the system (1) is non-degenerate if $\alpha_{lk}^{(i)} \ne 0, \ i = \overline{\overline{1,m}}$.

In the vector form of the formula (3) is: $\overline{e}_k = \frac{e_k}{\alpha_{lk}}$, $\overline{e}_i = e_i - \frac{e_k}{\alpha_{lk}}\alpha_{li}$, $\quad i = \overline{1,m}; \ i \ne k$.

Key stage algorithmic scheme finding the value of the machine rank basis matrix and solutions (1) based on the known properties of the trivial linear algebraic equation (SLAE) of the same dimension it has:

1) Perform simplex iteration substitution strings trivial basis matrix elements with known limitations of the method limited system (1) according to the relations (2) - (5);
2) Check the performance of conditions of regularity at each iteration;
3) Find the corresponding elements of the vector decomposition method: basis matrices by rows limitations of (1), the inverse basis matrix, the discrepancy limitations, artificial basic solutions $u_0^{(r)}$ where k is the number of iteration;
4) Control the number of iterations of substitution strings auxiliary system of the host (1), for which the nonsingularity conditions.

If the number of iterations for which $\alpha_{kk}^{(k)} \neq 0$, as well it is equal m, the only solution is got from the ratio of $A_{\sigma}^{-1} \cdot C = u^0$. If not, then if the conditions $k < m$ for the linear algebraic equation (1) is not satisfied the condition uniqueness of the solution. In this case, the model needs further analysis.

The computational process of translation of the inverse matrix is holding two stages of matrix operations:

5) Dividing the leading k-th column $e_k = (A_{\sigma}^{-1})_k$ on the value of the leading element $\alpha_{kk}^{(k)} \neq 0$ $e_k^{k+1} = e_k^k / \alpha_{kk}$;
6) To calculate on the k-th iteration for the i-th column of the inverse matrix ($i \neq k$, $i \in I$) of a new column $e_i^{k+1} = e_i^k - e_k^{k+1} \times \alpha_{ki}$, the quality of the localization field perturbation solutions (or approximations) essentially depends on the condition number (as a determining factor) and of algorithms (the length of the mantissa).

Features representations of numbers in the computer

Recently, the most commonly used standard IEEE. This is a standard binary arithmetic. This standard provides a well-defined data types along the length of the mantissa and the value of the order. For example, the mantissa 23, the order of 8 (32 characters), the mantissa order 52 11 (64 characters). This imposes a restriction on the threshold of "machine (computer) zero" and "machine (computer) infinity." For example, in the first case we have the machine zero threshold 2^{-126}, the threshold machine infinity 2^{128}.

In accordance with [Demmel, 2001] regardless of the length of the mantissa and the order of the following quality representation of real numbers in a computer:

- Machine zero (2^{-126});
- Subnormalized numbers $(-2^{-128}, 2^{-126})$;
- Normalized positive nonzero numbers $(2^{-128}, 2^{126})$;
- Normalized negative nonzero numbers $(-2^{-128}, -2^{-126})$;
- Positive overflow threshold (2^{127});
- Negative overflow threshold (-2^{128}).

The number is called normalized if its mantissa significant bit of nonzero.

Let is $a \odot b$ accurate results of calculations using the operation $\odot$, $fl(a \odot b)$ approximately computed floating-point number. If the number is closest to the calculated, then we believe that arithmetic rounding is correct (IEEE rounding correctly). For equidistant from rounding to less (the last sign of the mantissa is 0) - round to nearest even number.

When absenteeism beyond the allowable index order and correct rounding can be written: $fl(a \odot b) = (a \odot b)(1 + \delta)$, where - $|\delta|$ does not exceed the number of ε (machine epsilon) or machine precision (ε). You can read that the maximum relative error is equal to a number. Then the machine zero is maximum of $|\delta|$, at which the above mentioned ratio is carried out.

Subnormal number is calling, when it is located between zero and the smallest normalized floating-point number (smallest order number). For any two such numbers x, y *value* $fl(x-y)$ cannot be machine zero. An important feature is the fact that only in this case.

In the IEEE standard also provides for the number NAN (Not a number) overflow (as a result of division by zero). It follows that depending on the embodiment of submission to IEEE standard real accurate zero (as a natural number) in its calculations will respond zone subnormal numbers (fixed neighborhood of zero).

In calculations with different accuracy, we will have their own versions of the quantitative values of the zero-non-null, which can significantly affect the accuracy of the calculations.

Computer experiment in LM of low-dimensional

The effect of small perturbations, in the "small" LM quality localization solutions, is illustrated on a test study. It was organized as a structured multi-step algorithmic procedure. It consists in constructing a sequence of low-dimensional model problems that have properties irregularities starting with some simple steps. Just find items based on MBM method, calculate the condition numbers and graphical presentation of the main functional dependencies [Voloshyn, 2011].

Below are the resulting experiment symbols sets accessories making (localization).

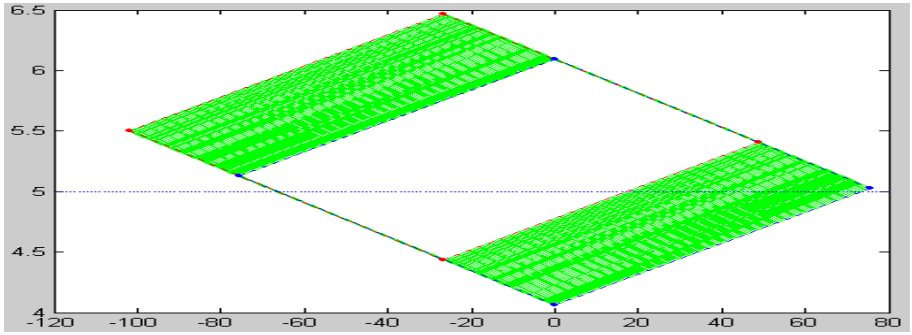


Figure 1. Rhomboid shape evaluation sets accessories decisions on iteration method (axes components of solutions) under perturbations of "parallel transport"

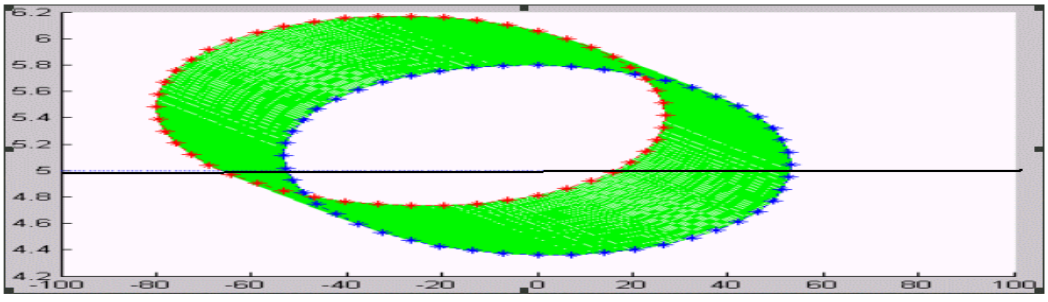


Figure 2. Ellipses accessories solutions on iteration method under perturbations vector of the right sides of sphere limitations of small diameter

Graphic images of ellipsoid shaped figures and rhombs (localization) are given disproportionate respect minimum and maximum abscissa and ordinate axes. Ellipsoid shaped pieces and rhombs images perturbation set of solutions (for small perturbations of the model elements) are elongated relative to the variety solutions boundary system of not full rank. Ellipsoid shaped pieces and rhombs images perturbation set of solutions (for small perturbations of the model elements) are elongated relative to the variety solutions boundary system full rank.

Figures include" major axis "(diagonal) and "minor axis" structurally. Major axis corresponds to a larger eigenvalue along the horizontal axis, is small (with small values of the diagonal axis of ordinates) - responds to small eigenvalues.

Figures 1 and 2 to a certain extent confirms the conjecture one of the co-authors [Voloshyn, 2006], on the validity of the Heisenberg uncertainty principle in computer research. This is of course the influence of small perturbations, namely the reduction of uncertainty ("improvement" localization) in one variable leads to an increase in uncertainty ("deterioration" localization) on the other.

Computer experiment on the analysis of small perturbations in the middle dimension LM

For computational experiment were elected three procedures MBM: without specifying solutions (0), with a one-step (1) and two-stage specification (2) [Bogaenko, 2012; Voloshyn, Kudin, 2010]. Having used different types of data: floating-point numbers (double precision (Double), 128-bit number (Dd) and 256 - bit numbers (Qd), and computation of exact numbers. Worth noting that on the same computing platform the ratio of the speed of algorithms that use different types of data will be permanent.

These computational experiments showed that the test platform for the chosen algorithm, which uses a 128-bit number, was about 35 times slower than an algorithm which uses 64-bit number, and if the 256-bit number - ~ 450 times in slowly. Such a significant slowdown due to the fact that operations with floating point High resolution is not implemented in hardware, in contrast to the operations on 64-bit numbers.

To build up relationships between the heuristic solutions and condition number was carried out a series of computational experiments. Conducted decision SLAE dimension 256x256 different algorithms MBM. As a criterion for the accuracy of machine precision was taken u0 solutions of (1) compared with the analytical (exact) solution. The experimental data presented in [Voloshyn, 2010].

Computer experiment established close to a linear dependence:

1) The accuracy of the solution (and matrix inversion) of the condition number (for fixed data type and algorithm) and the dimension of the model;
2) The accuracy of the solution on the data type (for a fixed number of conditionality and algorithm).

This allows you to build:

1) Interpolation polynomials - depending on the accuracy of the solution of conditionality;
2) Approximating the accuracy of solutions sets accessories (including ellipses minimum area);
3) Depending on extrapolation (for accuracy making algorithms) of the condition number.

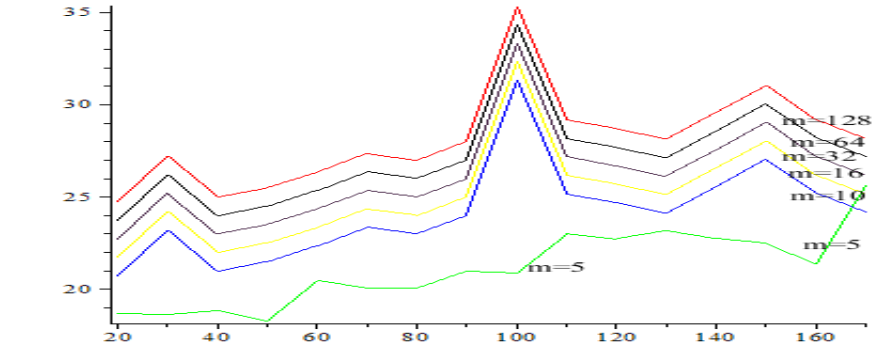
Methodology SAV [Voloshyn, 1987] aims to conduct research to establish the status of the component model, the analysis of inclusions (exclusions) component model on a bounded and closed (i.e., the quality of the localization). As follows from the results of computer simulation, even minor quantitative changes in the components of such models can qualitatively change the status of the component model, and as a consequence, the structure of the set of solutions of the problem.

In the study of stability of the problem is important to identify a quantitative measure of changes in the source data (correctness) in which this property is stored (lost). Identify the parameter values for which there are qualitative changes in the properties of the system. "Catch and evaluate" such structural quality parameters

depending on the quantity (in the small) between the practical problems of mathematical description of the model and engine. For example, in the language" $(\varepsilon - \delta)$" - of the deviation or change solutions - from data changes within the (category of stability) is not only difficult, but more often and it is impossible even for ill-conditioned problems of small dimension.

In conclusion, it can be noted the need to: inclusion in the process of numerical experiment of the decision; analyze the impact of different strategies for computing (data types, algorithms, level of system conditioning) on the basic parameters of the solution (the accuracy of the solutions and matrix inversion, speed, volume calculations); develop efficient computational methods and algorithms schemes evaluating conditioning system during the experiment on the given criteria for models of different dimension.

In particular, we consider the properties of the basic methods and algorithms for finding the exact solution of linear algebraic equation in MAPLE environment. The features of the calculations with different length of the mantissa, the dimension of the constraint matrix, with the "bad" structural properties and verified the reliability of some "non-strict" evaluation of errors of calculation formulas. Dependency analysis of the relative errors solutions on the dimension of the constraint matrix in an ill-conditioned LM middle dimension shown in Figure 3, with the length of the mantissa m = 5, 10, 16 etc.



...

Figure 3.

One of the empirical formulas accuracy of calculations given in [Khimich, 2007] as follows: if Masheps = 2.2 * 10 ^ (-16), M-condition number (order), the calculation error may be filed as d = Masheps - M (order).

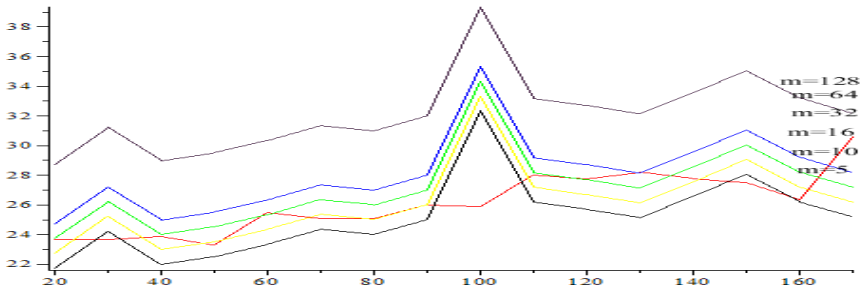


Figure 4.

Other empirical formula estimating the order of accuracy of the calculations [Alfeld, 1987]:

if the number is written in the form of floating-point, where m - mantissa, b - base power. e - the procedure is usually base 2. If there is an algorithm that runs in machine interval arithmetic with a mantissa length t1 and mantissa length t2, where t2 = t1 + ll> 0, then the limit of absolute and relative error will decrease in time bl.

Computer experiment set: 1) a significant influence of the structural properties of the constraint matrix, the dimension of the model, the length of the mantissa conditionality computational properties of algorithms; 2) The computational complexity of the condition number; 3) Loss of credibility evaluation formulas with "growth" ill-conditioned, the dimension of the model; 4) The special influence of the leading elements and standards leading columns on the quality of evaluations.

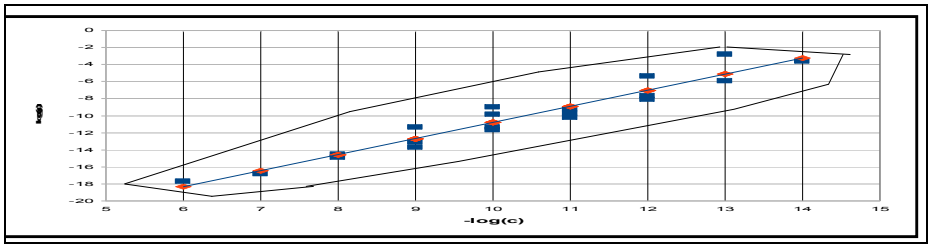


Figure 5. Option piecewise linear approximation (localization) of the range of the order of accuracy in the coordinate system of values and the order of accuracy of the order of the condition number.

Computer experiment also found close to a linear dependence [Bogatenko, 2012]:

1) The accuracy of the solution (and matrix inversion) of the condition number (for fixed data type and algorithm) and the dimension of the model; 2) The accuracy of the solution on the data type (for a fixed number of conditionality and algorithm); 3) close to the polynomial speed depending on the dimension of the problem (for fixed data types, including conditionality and dimension model); 4) the need to develop new algorithms for analysis of linear systems , in which there are " evaluators " condition numbers during calculations.

Conclusion

This work continues (and in some sense generalizes) study the influence of small perturbations of course linear models [Voloshyn, Kudin, 2010 - 2012] and emphasizes the need to develop "analysis certainly small" analysis tools like computer models, in the first place, for the localization of solutions systems of linear algebraic equations. Subsequent publications will address the challenges of "computer convergence" of iterative processes and ways of solving them by using proposed in this paper (and previous publications) approaches based on "analysis certainly small".

Acknowledgements

The work is published with the financial support of the project ITHEA XXI Institute of Information Theory and Applications FOI ITHEA Bulgaria www.ithea.org Association and the creators and users of intellectual systems ADUIS Ukraine www.aduis.com.ua.

Bibliography

[Ahlefeld, 1987] Ahlefeld G., Herzberger J. Introduction to interval computation. – Moscow: Mir, 1987, 360p. (in Rus.)

[Bogaenko, 2012] Bogaenko V.A., Kudin V.I., Skopetskiy V.V. On peculiarities of the organization on the basis of calculating the basis matrices // Cybernetics and Systems Analysis, N 4, 2012. P.146 -154 (in Rus.)

[Clark, 1988] Clark F. Optimization and nonsmooth analysis. Moscow: Nauka, 1988. 280p. (in Rus.)

[Demmel, 2001] Demmel J. Numerical Linear Algebra : Theory and Applications. Moscow : Mir, 2001. – 430p. (in Rus.)

[Khimich, 2007] Khimich A.N. Numerical mining software MIMD - computer INPARKOM. – Kiev: Naukova Dumka, 2007. – 220p. (in Rus.)

[Kudin, 2002] Kudin V.I. Application of basis matrices in the study of linear systems // Bulletin of Kiev University. Series Sci. science, 2002. № 2. - P.56-61 (in Ukr)

[Matthews, 2001] Matthews D.G., Fink K.D. Numerical methods. - Moscow- Kiev: Williams, 2001. – 703p. (In Rus.)

[Mihalevitch, 1965] Mihalevitch V.S. Sequential optimization algorithms and their application // Cybernetics, 1965, № 1, (P.45 -55), № 2 (P.85 -89) (in Rus.)

[Mihalevitch, 1961] Mihalevitch V.S., Shore N.Z. Method of sequential analysis for solving variational problems in management, planning and design // Report on the IV All-Union Mathematical Congress. - L.: 1961. - P.91. (in Rus.)

[Mihalevitch, Voloshyn, 1981] Mihalevitch V.S., Volkovich V.L., Voloshyn A.F. The method of sequential analysis in large demantion linear programming // Cybernetics, № 4, 1981. - P.114 - 120 (In Rus.)

[Moiseev, 1981] Moiseev N.N. Mathematical problems of system analysis. - Moscow: Nauka, 1981. – 488p. (in Rus.)

[Samarskiy, 1989] Samarskiy A.A., Gulin A.G. Numerical methods . - Moscow: Nauka, 1989. – 432p. (in Rus.)

[Volkovich, Voloshyn, 1978] Volkovich V.L., Voloshyn A.F. On a scheme of sequential analysis and screening options (Algorithm W) // Cybernetics, № 4, 1978. - P.98 – 105 (In Rus.)

[Voloshyn, 1987] Voloshyn A.F. Optimal area localization method in mathematical programming // Reports of the AS of USSR, 1987, Vol. 293, № 3. - P.549 -553 (in Rus.)

[Voloshyn, 2006] Voloshyn O. About Problems of Decision Making in Social and Economic Systems // International Journal "Information Theories & Applications", Vol.13, N1, 2006. - P.31- 38.

[Voloshyn, 2010] Voloshyn A.,Kudin V. Methods of analysis of small perturbations of linear models // Natural and Artificial Intelligence, ITNEA , Sofia, 2010. - P.41-47 (in Rus.)

[Voloshyn, 2011] Voloshyn O., Kudin V. Course small perturbations of linear models // International Books Series "Information Models of Knowledge", Vol.5, N3, 2011. - P.226-231 (in Rus.)

[Voloshyn, 2012] Voloshyn O., Kudin V. Sequential analysis of certainly small course for linear models in computer simulations // Intern. Journ. "Information Technologies & Knowledge", Vol. 6, N3, 2012. - P. 240-249 (in Rus.).

[Voloshyn, 2013] Voloshyn O.F., Gnatienko G.M., Kudin V.I. Sequential analysis: Technology and Applications. - Kiev: Stylos, 2013. – 304p. (in Ukr.)

[Yushkevich, 1970-1972] Yushkevich A.P. and other. History of mathematics from ancient times to the beginning of the XIX century: In 3 vol. - Moscow : Nauka, Vol. №1 , 1970 (352p.), Vol. №2 , 1971 (300p.), Vol. №3 , 1972 (496p.) (in Rus.)

Authors' Information

Oleksii Voloshyn *- Professor, Kyiv National Taras Shevchenko University, Faculty of Cybernetics; e-mail: olvoloshyn@ukr.net*

Research interests: decision-making theory, decision support systems, information technology, expert systems, mathematical economics, fuzzy analysis, e-learning

Vladimir Kudin *- Dr. Sc., Senior Researcher, Kyiv National Taras Shevchenko University; e-mail: V_I_Kudin@mail.ru*

Research interests: optimization techniques, systems analysis, computational method , decision-making, mathematical economics

A STUDY OF APPLICATION OF NEURAL NETWORK TECHNIQUE ON SOFTWARE REPOSITORIES

Ana Maria Bautista, Tomas San Feliu

Abstract: *Context: The use of software repositories is recent in software engineering. Different techniques have been applied to software engineering problems. We want to know the degree of application of neural networks as data mining technique. The results have allowed the preliminary state of the art of the application of neural network techniques on software repositories.*

Keywords: *neural networks, repository, data mining.*

ACM Classification Keywords: *I.2.6 Artificial Intelligence –Connectionism and neural nets, H.2.7 Database Administration – Data warehouse and repository.*

Introduction

The main goal of a software repository is to maintain the project data for later use. The data in mature engineering are used to check and compare different techniques.

In software engineering there are now a wide variety of software repositories. Software engineering as an emerging discipline is beginning to use the software repositories to compare different techniques.

The systematic review can be considered as a secondary study that reviews articles related to a specific search topic. As a secondary study provides an overview of an area of research to evaluate existing evidence [Kitchenham, 2009] and can provide identifying gaps in primary studies, where they may require new or enhanced studies.

These studies require a rigorous search and inclusion criteria and exclusion that are clearly defined in the research protocol and are presented in the report results.

The purpose of this study is to determine which use is made of the repositories and techniques have been tested and compared with them.

A software engineering (SE) data repository is defined as a set of well-defined, useful, and pertinent real-world data related to software projects, called datasets, which include quantitative and descriptive information about resources, products, processes, techniques, management, etc. Such data are being collected for various purposes by recognized organizations, as well as by individual software organizations and researchers. In most scientific and engineering disciplines, these data are useful for conducting benchmarking, experimental, and empirical studies. While highly varied and widely available in mature disciplines, data repositories are much less frequently found in emerging disciplines, including software engineering, as illustrated by the Guide to the Software Engineering Body of Knowledge [SWEBOK, 2004].

Mining software repositories (MSR) has become a fundamental area of research for the Software Engineering community, and of vital importance in the case of empirical studies. Software repositories contain a large amount of valuable information that includes source control systems storing all the history of the source code, defect tracking systems that host defects, enhancements and other issues, and other communication means such as mailing lists or forums.

To extract information from the Software Repositories different techniques are used. Mohanty et al. classify intelligent techniques in the following [Mohanty, 2010]:

1. Different neural network (NN) architecture including multilayer perception (MLP) and cascade correlation NN;
2. Fuzzy logic;
3. Genetic algorithm (GA);
4. Decision tree;
5. Case-based reasoning (CBR);
6. Soft computing (hybrid intelligent systems).

The other techniques:

1. Analogy based;
2. Support vector machine;
3. Self organizing maps (SOM).

Specifically, this work will focus on studying the application of neural networks in existing repositories.

Neural networks are used broadly in the studies we have selected. Therefore consider your extension, we will set the yields obtained in the literature and discuss the tendency in recent years.

This paper will be organized as follow. Research methodology section will describe the systematic review. Next section title Data Collection, reports the most relevant information gathered. Results section will report the review results analyzing collected date ordered by research questions. Discussion section summarizes the main findings. Study limitation section will discuss the assumptions and considerations of the study. Finally, conclusion and future work will outline the main conclusions obtained in future research works

Research methodology

This section provides an overview of the steps involved in the process systematic review, including the formulation of research question, the search strategy, the inclusion and exclusion criteria, and finally the data collection process.

Systematic mapping studies are a type of systematic literature review that aims to collect and classify research papers related to a specific topic [Peticrew, 2006; Kitchenham, 2007; Petersen, 2008].

This section provides an overview of the steps involved in the process of mapping review following Petersen et al. [Petersen, 2008] including the formulation of the research questions, the search strategy for primary studies, the inclusion and exclusion criteria, and the data collection process.

- **Research questions**

The main goal addressed by this study is to analyze the use of repositories and techniques by the research community and to consider its weakness to undertake the appropriate scientific research.

In this study the following research questions were considered:

- ✓ *Research question number 1 (RQ1):* Which and how many journals and conferences include techniques for mining software repositories research papers? To identify what are the main literature sources where the software repository analysis are published.
- ✓ *Research question number 2 (RQ2):* How comprehensive is the use of neural networks in the analysis of software repositories? Is necessary to know the use of neural networks for the analysis of repositories and is required to characterize the types of networks are used.

- ✓ *Research question number 3 (RQ3):* In which years have conducted studies with neural networks? How have been evolving the use of neural networks over time. The aim is to establish whether there is a trend over the last years.

- **Search for primary studies strategy**

The following search engine Google Scholar was used to make a general search for relevant papers in journals and conference proceedings. This search engine was selected because it is major search engine and it has a good usability.

Search is based in the two most used open software repositories: ISBSG and PROMISE. Search terms Data Mining ISBSG Repository and "Data Mining" "PROMISE Repository" were used. Only have been considered papers published from year 2010. Search was completed in January 2014. There were 304 papers, 94 corresponding to the first search term and 210 to the second.

Some studies use both repositories and therefore they are being selected for both search terms. Also it was noted that some papers have been published in different journal and conference proceedings or in different years. Duplicate references have been eliminated, overall 46 papers.

Finally, 258 references remained. The overall primary study selection is summarized in Figure 1.

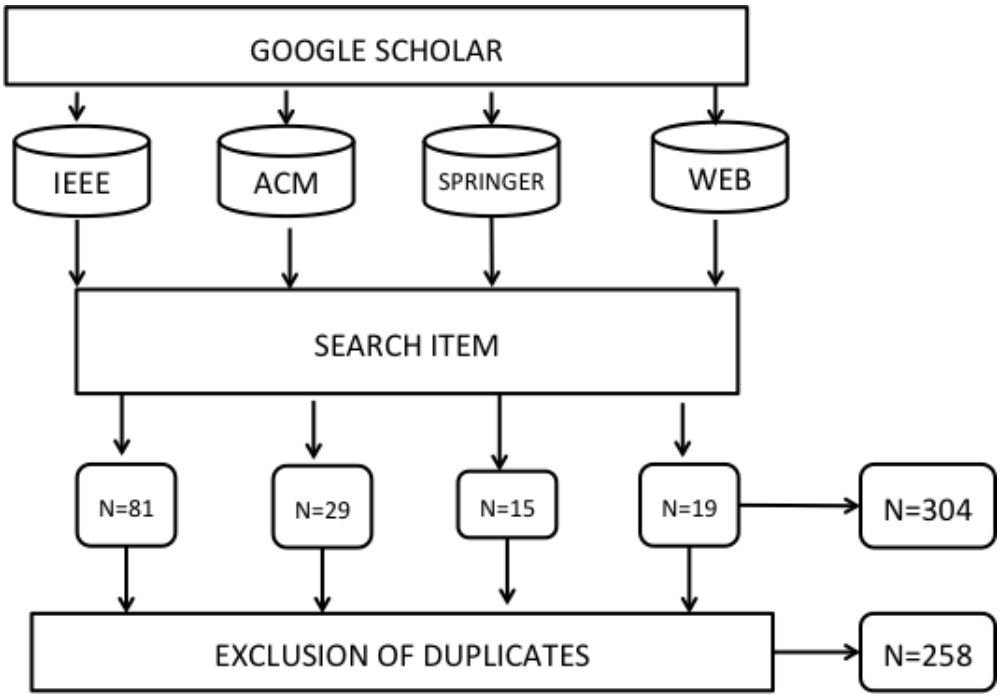


Figure 1. Search process for the selection of studies

- **Inclusion and exclusion criteria**

Inclusion and exclusion criteria are required to evaluate each primary study. In order to improve its reliability, only one author performed the screening process.

Inclusion and exclusion criteria are required to assess each potential primary study. In order to improve its reliability, the filtering process based on inclusion and exclusion criteria. All conflicts were resolved via discussion.

The following list describes the filtering process:

- ✓ *F1:* The first filter (F1) was used to identify papers that didn´t speak about data mining techniques and for this reason they didn´t answer our search questions, 82 papers were filtered;
- ✓ *F2:* The second filter (F2) was used to identify reviews that didn´t anything new to the rest of papers, 5 papers were filtered.
- ✓ *F3:* The third filter (F3) was used to locate the papers that were not written in English, 2 papers were filtered.

169 papers satisfied the logical condition (F1 AND F2 AND F3). Filtering process is summarized in Figure 2.

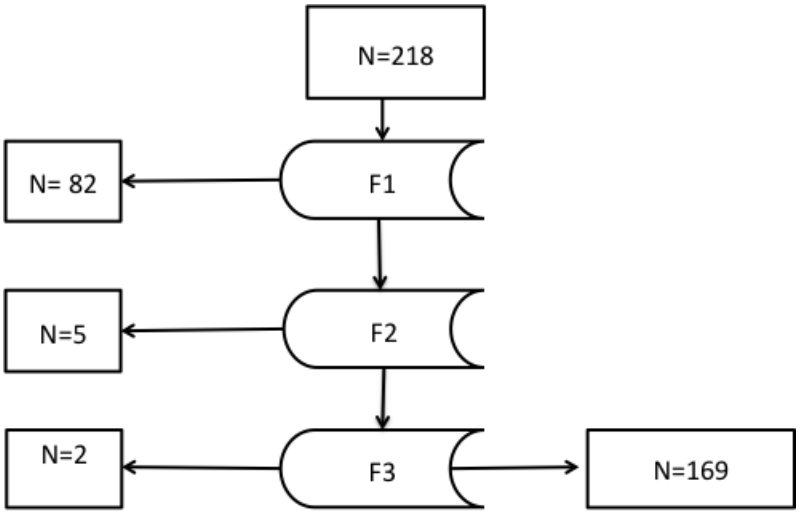


Figure 2. Filtering process

Data Collection

After the filtering process, the most relevant information was obtained from remained studies.

After the filtering process, the most relevant information was obtained from each of the 169 remained studies. This includes both general information and data addressing the five research questions. After reading each paper, the data was extracted and stored in a spreadsheet.

Papers classified by sources are shown in Table 1. Table 1 shows the evolution along the filtering process.

Table 1. Papers by sources

	Initial dataset	Duplicates	Filtered	Final dataset
IEEE	81	12	20	49
ACM	29	9	8	12
Springer	15	2	5	8
Others	179	23	56	100
TOTAL	304	46	89	169

It was analyzed the filtering process from the point of view of publication year. This point of view is shown in Table 2.

Table 2. Papers by published year

	Initial dataset	Duplicates	Filtered	Final dataset
2010	60	5	24	31
2011	51	6	13	32
2012	70	9	24	37
2013	108	25	27	56
2014	15	1	1	13
TOTAL	304	46	89	169

Additionally, the different sources was analyzed the distribution of type of papers. The most important source is the journals. The conferences are the second source of information. The Table 3 and Table 4 present the distribution of studies across different journals and conferences

Table 3. Papers by Journal

Journal	Total
Information and Software Technology	11
IEEE Transaction on Software Engineering	9
International Journal of Software Engineering and Knowledge Engineering	7
Empirical Software Engineering	6
Journal of Systems and Software	4
Information Sciences	4
Software Quality Journal	3
IET Software	2
IEEE Transaction on Reliability	2
International Journal of Software Engineering and Computing	2
International Journal of Reliability, Quality and Safety Engineering	2
Others	36
TOTAL	88

Table 4. Papers by Conference

Conference	Items
International Conference on Predictive Models in Software Engineering	8
International Conference on Software Process and Product Measurement	6
CSI International Conference on Software Engineering	3
International Joint Conference on Neural Networks	3
IEEE/ACM International Conference on Automated Software Engineering	2
International Conference on Computer, Information and Telecommunication Systems	2

International Symposium on Empirical Software Engineering and Measurement	2
IEEE International Conference on Tools with Artificial Intelligence	2
International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering	2
Others	35
TOTAL	65

One import question is related with the occurrence of neural networks in data mining software repository. It was gathered information about the techniques used in selected studies. Sometimes one study contains several techniques. The distribution of techniques used is shown in Table 5.

Table 5. Distribution of techniques

Technique	**Percentage**
Neural Networks	14,29
Fuzzy Logic	2,77
Genetic algorithm	0,85
Decision trees	17,48
Case-base Reasoning	2,13
Hybrid intelligent systems	5,97
Analogy based	5,12
Support vector machine	2,77
Self organizing maps	2,77
Statistical techniques	44,78
Others	1,07

Results

The main results of systematic review are presented following the researched questions

- RQ1: Which and how many journals and conferences include techniques for mining software repositories research papers?

The publications of data mining techniques in software repositories have been captured of IEEE and ACM. The publications are concentrated in the domains of Software Engineering and Knowledge Engineering. In Table 4 is shown the ranking of journals. The three main journals are Information and Software Technology, IEEE Transaction on Software Engineering and International Journal of Software Engineering and Knowledge Engineering. The most important conference in this field of study are International Conference on Predictive Models in Software Engineering, International Conference on Software Process and Product Measurement, CSI International Conference on Software Engineering and International Joint Conference on Neural Networks. The tendency established by the journals is confirmed by the conference. It is denote IJCNN is specialized in neural networks and listed more conference publications.

- RQ2: How comprehensive is the use of neural networks in the analysis of software repositories?

This question is required to identify the use of neural networks in the studio of repositories and it requires characterize the types of networks are used. In the 169 papers selected in our study all kinds of data mining techniques are used. In many cases hybrid techniques are used with the original techniques or algorithms that improve their performance complement. In all studies the performance of at least two techniques are compared and in many numbers much higher. Most frequently statistical techniques are used. Neural Networks are ranked in third position below of decision trees as shown in Table 5. The multilayer perceptron (MLP) and Radial Basis function Network (RBFN) are some of the popular Neural Network architecture.

- RQ3: What years there have been conducted studies with neural networks?

How have been evolving the use of neural networks over time? It intends to establish whether there is a trend over the last years. The figure 3 shows that there is a significant increase over 2013.

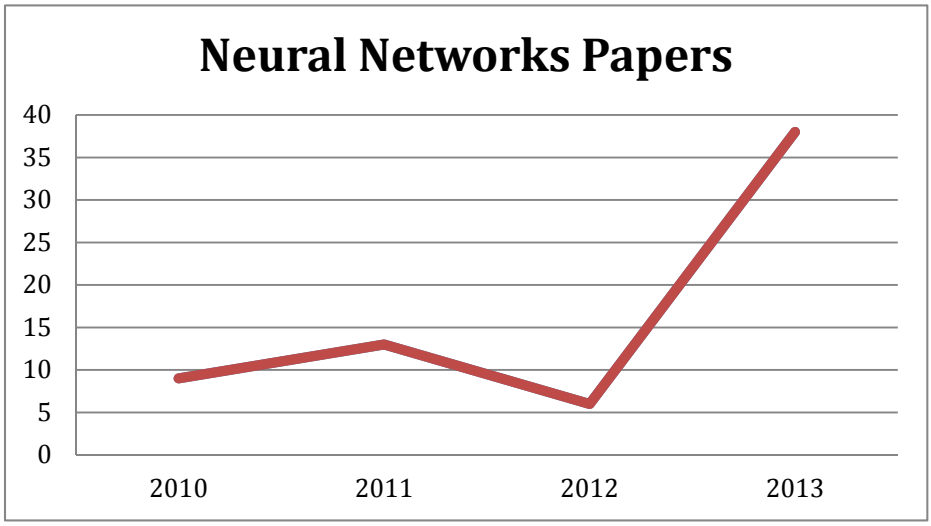


Figure 3. Evolution of Neural Networks publications

Discussion

This section summarizes the main findings of the systematic review. It also includes the limitations of the study and discusses the implications for researchers.

This study shows the extent to which and how software engineering researchers have used ISBSG and PROMISE until January of 2014. Thereby, the papers that have worked with these datasets have been identified and classified by answering a set of research questions. This systematic mapping review conveys a picture of application of neural networks on software repositories.

The search terms Data Mining ISBSG Repository and “Data Mining” “PROMISE Repository” were the input for Google Scholar. This resulted 94 y 210 results respectively. The search was completed in January 2014. After the elimination of duplicates and the filtering process, the most relevant information was obtained from each of the 169 remaining studies.

The first research has been a consolidation of journals and conferences where articles are published on this subject. It is noteworthy that the conference that has emerged in recent years the subject of analysis software repository is maintained.

A second research question shows that the neural networks maintain their presence as data mining technique. Furthermore, from the viewpoint of performance which is a consolidated sample technique. In fact neural networks are consolidated as reference technique in comparisons to characterize its performance.

Regarding the third question, we see that during the years 2010 to 2012 remained stable presence. In the years 2013-2014 a remarkable growth over the previous three years is provided.

Study limitations

It is important to consider that the results obtained form a systematic review could be affected by researchers conducting the review, by the selected search terms, and by the chosen time frame [Elberzhaer, 2012].

The first limitation concerns the search strategy employed, to use ISBSG and PROMISE in search terms, results can be affected.

It is important to indicate that some more recent studies may be missing because the search engines may not have indexed them.

Finally, exclusion of papers written in a language other than English may have lead to bias in the selection process. This could not be avoided due to impossibility of the revision team to address these languages.

The second limitation concerns bias in the data collection. The first information collected was about the identification and general details of the paper such as title, its authors, source where it was published, abstract, year of publication. Full texts have been read when abstract of papers not provide enough information.

Conclusion and future work

This paper presents the results of a systematic review about the usage of techniques on software repositories until January of 2014. After the searching and filtering process, 169 papers were analyzed.

They have cataloged the papers selected based repository using the techniques of data mining that use storing spreadsheet data needed to answer questions research.

Analyzing results of this study note that interest in software repositories usage and trend to employ neural networks techniques to data mining is increasing.

In summary, this paper presents a comprehensive snapshot of actual use of neural network to analyze software repositories.

In the future, the authors intend to explain the period of gathering of papers to keep the work up-to-date and the answering other interesting questions concerning the trend of usage of neural networks. Other future work will delve into the use of neural networks as a tool for mining software repositories to facilitate the work of developers and project managers.

Bibliography

[Cheikhi, 2013] Cheikhi, L., & Abran, A. Promise and ISBSG Software Engineering Data Repositories: A Survey. In Software Measurement and the 2013 Eighth International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2013, Joint Conference of the 23rd International Workshop on IEEE, pp. 17-24

[Elberzhaer, 2012] Elberzhager, F., Münch, J., & Nha, V. T. N. A systematic mapping study on the combination of static and dynamic quality assurance techniques.Information and Software Technology, 2012, 54(1), 1-15

[Fernández-Diego, 2014] Fernández-Diego, M., & González-Ladrón-de-Guevara, F. Potential and Limitations of the ISBSG Dataset in Enhancing Software Engineering Research: A Mapping Review. Information and Software Technology, 2014

[Kitchenham, 2007] B. Kitchenham, S. Charters, Guidelines for Performing Systematic Literature Reviews in Software Engineering, Software Engineering Group, School of Computer Science and Mathematics, Keele University, 2007.

[Kitchenham, 2009] Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., & Linkman, S. Systematic literature reviews in software engineering–A systematic literature review. Information and software technology, 2009, 51(1), 7-15.

[Mohanty, 2010] Mohanty, R., Ravi, V., & Patra, M. R. The application of intelligent and soft-computing techniques to software engineering problems: a review. International Journal of Information and Decision Sciences, 2010, 2(3), pp. 233-272.

[Petersen, 2008] Petersen K., Feldt R., Mujtaba S., Mattsson M., Systematic mapping studies in software engineering, in: 12th Int. Conf. Eval. Assess. Softw. Eng., 2008, p. 1.

[Petticrew, 2006] M. Petticrew, H. Roberts, Systematic Reviews in the Social Sciences: A Practical Guide, John Wiley & Sons, Limited, 2006.

[Robles, 2010] Robles, G. Replicating MSR: A study of the potential replicability of papers published in the Mining Software Repositories proceedings. In Mining Software Repositories (MSR), 2010 7th IEEE Working Conference on pp. 171-180, May 2010

[SWEBOK, 2004] IEEE. 2004. "Guide to the Software Engineering Body Of knowledge- SWEBOK." Los Alamitos, California: IEEE Computer Society, 204 p. (last accessed on 30/01/2013). http://www.computer.org/portal/web/swebok

Appendix A

This section provides the primary studies selected from the systematic review

Abaei, G., & Selamat, A. (2013). A survey on software fault detection based on different prediction approaches. Vietnam Journal of Computer Science, 1-17.

Abaei, G., Rezaei, Z., & Selamat, A. (2013, November). Fault prediction by utilizing self-organizing Map and Threshold. In Control System, Computing and Engineering (ICCSCE), 2013 IEEE International Conference on (pp. 465-470). IEEE.

Afzal, W., Torkar, R., & Feldt, R. (2012). Resampling Methods in Software Quality Classification. International Journal of Software Engineering and Knowledge Engineering, 22(02), 203-223.

Al Khalidi, N., Saifan, A. A., & Alsmadi, I. M. (2012, May). Selecting a standard set of attributes for cost estimation of software projects. In Computer, Information and Telecommunication Systems (CITS), 2012 International Conference on (pp. 1-5). IEEE.

Al Khalidi, N., Saifan, A. A., & Alsmadi, I. M. (2012, May). Selecting a standard set of attributes for cost estimation of software projects. In Computer, Information and Telecommunication Systems (CITS), 2012 International Conference on (pp. 1-5). IEEE.

Alan, O., & Catal, C. (2011). Thresholds based outlier detection approach for mining class outliers: An empirical case study on software measurement datasets. Expert Systems with Applications, 38(4), 3440-3445.

Al-Jamimi, H. A., & Ghouti, L. (2011, December). Efficient prediction of software fault proneness modules using support vector machines and probabilistic neural networks. In Software Engineering (MySEC), 2011 5th Malaysian Conference in (pp. 251-256). IEEE.

Alsmadi, I., & Najadat, H. (2011). Evaluating the change of software fault behavior with dataset attributes based on categorical correlation. Advances in Engineering Software, 42(8), 535-546.

ALTIDOR, W., KHOSHGOFTAAR, T. M., & GAO, K. (2010). Wrapper-based feature ranking techniques for determining relevance of software engineering metrics. International Journal of Reliability, Quality and Safety Engineering, 17(05), 425-464.

Amasaki, S., & Yokogawa, T. (2013, October). The Effects of Variable Selection Methods on Linear Regression-Based Effort Estimation Models. In Software Measurement and the 2013 Eighth International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2013 Joint Conference of the 23rd International Workshop on (pp. 98-103). IEEE.

Amasaki, S., & Yokogawa, T. (2013, October). The Effects of Variable Selection Methods on Linear Regression-Based Effort Estimation Models. In Software Measurement and the 2013 Eighth International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2013 Joint Conference of the 23rd International Workshop on (pp. 98-103). IEEE.

Amasaki, S., Takahara, Y., & Yokogawa, T. (2011, November). Performance Evaluation of Windowing Approach on Effort Estimation by Analogy. In Software Measurement, 2011 Joint Conference of the 21st Int'l Workshop on and 6th Int'l Conference on Software Process and Product Measurement (IWSM-MENSURA) (pp. 188-195). IEEE.

Amasaki, S., Takahara, Y., & Yokogawa, T. (2011, November). Performance Evaluation of Windowing Approach on Effort Estimation by Analogy. In Software Measurement, 2011 Joint Conference of the 21st Int'l Workshop on and 6th Int'l Conference on Software Process and Product Measurement (IWSM-MENSURA) (pp. 188-195). IEEE.

Anwar, S., Rana, Z. A., Shamail, S., & Awais, M. M. (2012, September). Using association rules to identify similarities between software datasets. In Quality of Information and Communications Technology (QUATIC), 2012 Eighth International Conference on the (pp. 114-119). IEEE.

Armah, G. K., Luo, G., & Qin, K. (2013, November). Multi_level data pre_processing for software defect prediction. In Information Management, Innovation Management and Industrial Engineering (ICIII), 2013 6th International Conference on (Vol. 2, pp. 170-174). IEEE.

Azzeh, M. (2011, September). Software effort estimation based on optimized model tree. In Proceedings of the 7th International Conference on Predictive Models in Software Engineering (p. 6). ACM.

Azzeh, M., & Alseid, M. (2013). Value of ranked voting methods for estimation by analogy. Software, IET, 7(4).

Azzeh, M., Cowling, P. I., & Neagu, D. (2010, June). Software Stage-Effort Estimation Based on Association Rule Mining and Fuzzy Set Theory. In Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on (pp. 249-256). IEEE.

Bakır, A., Turhan, B., & Bener, A. (2011). A comparative study for estimating software development effort intervals. Software Quality Journal, 19(3), 537-552.

Bakır, A., Turhan, B., & Bener, A. B. (2010). A new perspective on data homogeneity in software cost estimation: a study in the embedded systems domain. Software Quality Journal, 18(1), 57-80.

Balsera, J. V., Montequin, V. R., Fernandez, F. O., & González-Fanjul, C. A. (2012). Data Mining Applied to the Improvement of Project Management.

Baojun, M., Dejaeger, K., Vanthienen, J., & Baesens, B. (2011). Software defect prediction based on association rule classification. Available at SSRN 1785381.

Bardsiri, V. K., Jawawi, D. N. A., Bardsiri, A. K., & Khatibi, E. (2013). LMES: A localized multi-estimator model to estimate software development effort. Engineering Applications of Artificial Intelligence, 26(10), 2624-2640.

Bardsiri, V. K., Jawawi, D. N. A., Hashim, S. Z. M., & Khatibi, E. (2013). A flexible method to estimate the software development effort based on the classification of projects and localization of comparisons. Empirical Software Engineering, 1-28.

Barros, R. C., Basgalupp, M. P., Cerri, R., da Silva, T. S., & de Carvalho, A. C. (2013, July). A grammatical evolution approach for software effort estimation. In Proceeding of the fifteenth annual conference on Genetic and evolutionary computation conference (pp. 1413-1420). ACM.

Benala, T. R., Mall, R., Srikavya, P., & HariPriya, M. V. (2014, January). Software Effort Estimation Using Data Mining Techniques. In ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India-Vol I (pp. 85-92). Springer International Publishing.

Bettenburg, N., Nagappan, M., & Hassan, A. E. (2014). Towards improving statistical modeling of software engineering data: think locally, act globally!. Empirical Software Engineering, 1-42.

Borges, K. C. A. D., de Barcelos Tronto, I. F., de Aquino Lopes, R., & da Silva, J. D. S. (2013). A Data Pre-Processing Method for Software Effort Estimation Using Case-Based Reasoning. CLEI ELECTRONIC JOURNAL, 16(3).

Borges, R., & Menzies, T. (2012, September). Learning to change projects. In Proceedings of the 8th International Conference on Predictive Models in Software Engineering (pp. 11-18). ACM.

Brady, A., & Menzies, T. (2010, September). Case-based reasoning vs parametric models for software quality optimization. In Proceedings of the 6th International Conference on Predictive Models in Software Engineering (p. 3). ACM.

Bruntink, M. (2013). Towards base rates in software analytics: Early results and challenges from studying Ohloh. Science of Computer Programming.

Castejón-Limas, M., Ordieres-Meré, J., González-Marcos, A., & González-Castro, V. (2011). Effort estimates through project complexity. Annals of Operations Research, 186(1), 395-406.

Catal, C. (2014). A Comparison of Semi-Supervised Classification Approaches for Software Defect Prediction. Journal of Intelligent Systems, 23(1), 75-82.

Catal, C., & Diri, B. (2013). A fault detection strategy for software projects. Tehnički vjesnik, 20(1), 1-7.

Catal, C., Alan, O., & Balkan, K. (2011). Class noise detection based on software metrics and ROC curves. Information Sciences, 181(21), 4867-4877.

Catal, C., Sevim, U., & Diri, B. (2010). Metrics-driven software quality prediction without prior fault data. In Electronic Engineering and Computing Technology (pp. 189-199). Springer Netherlands.

Chang, R., Mu, X., & Zhang, L. (2011). Software Defect Prediction Using Non-Negative Matrix Factorization. Journal of Software (1796217X), 6(11).

Chaturvedi, K. K., & Singh, V. B. (2012, September). Determining bug severity using machine learning techniques. In Software Engineering (CONSEG), 2012 CSI Sixth International Conference on (pp. 1-6). IEEE.

Chatzikonstantinou, G., Kontogiannis, K., & Attarian, I. M. (2013, January). A goal driven framework for software project data analytics. In Advanced Information Systems Engineering (pp. 546-561). Springer Berlin Heidelberg.

Cheikhi, L., & Abran, A. (2013, October). Promise and ISBSG Software Engineering Data Repositories: A Survey. In Software Measurement and the 2013 Eighth International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2013 Joint Conference of the 23rd International Workshop on (pp. 17-24). IEEE.

Cheikhi, L., & Abran, A. (2013, October). Promise and ISBSG Software Engineering Data Repositories: A Survey. In Software Measurement and the 2013 Eighth International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2013 Joint Conference of the 23rd International Workshop on (pp. 17-24). IEEE.

Corazza, A., Di Martino, S., Ferrucci, F., Gravino, C., Sarro, F., & Mendes, E. (2013). Using tabu search to configure support vector regression for effort estimation. Empirical Software Engineering, 18(3), 506-546.

Crespo, D., & Ruiz, M. (2012, December). Decision making support in CMMI process areas using multiparadigm simulation modeling. In Simulation Conference (WSC), Proceedings of the 2012 Winter (pp. 1-12). IEEE.

Czibula, G., Marian, Z., & Czibula, I. G. (2014). Software defect prediction using relational association rule mining. Information Sciences.

Dejaeger, K., Verbraken, T., & Baesens, B. (2013). Toward Comprehensible Software Fault Prediction Models Using Bayesian Network Classifiers. Software Engineering, IEEE Transactions on, 39(2), 237-257.

del Águila, I. M., & Del Sagrado, J. (2011). Requirement risk level forecast using bayesian networks classifiers. International Journal of Software Engineering and Knowledge Engineering, 21(02), 167-190.

Di Martino, S., Ferrucci, F., Gravino, C., & Sarro, F. (2011). A genetic algorithm to configure support vector machines for predicting fault-prone components. In Product-Focused Software Process Improvement (pp. 247-261). Springer Berlin Heidelberg.

Elyassami, S., & Idri, A. (2013). Evaluating software cost estimation models using fuzzy decision trees. Recent Advances in Knowledge Engineering and Systems Science, WSEAS Press, 243-248.

Falessi, D., Cantone, G., & Canfora, G. (2013). Empirical principles and an industrial case study in retrieving equivalent requirements via natural language processing techniques. Software Engineering, IEEE Transactions on, 39(1), 18-44.

Fernandes, P., Lopes, L., Normey, S., & Ruiz, D. (2013, May). Stochastic aware random forests-a variation less impacted by randomness. In Proc. of the Twenty Sixth Int. FLAIRS Conf.(FLAIRS 2013), C. Boonthum-Denecke and GM Youngblood, Eds. AAAI Press (pp. 146-149).

Fernández-Diego, M., & González-Ladrón-de-Guevara, F. (2014). Potential and Limitations of the ISBSG Dataset in Enhancing Software Engineering Research: A Mapping Review. Information and Software Technology.

Gao, K., & Khoshgoftaar, T. M. (2011). Software Defect Prediction for High-Dimensional and Class-Imbalanced Data. In SEKE (pp. 89-94).

GAO, K., KHOSHGOFTAAR, T. M., & WALD, R. (2014). THE USE OF UNDER-AND OVERSAMPLING WITHIN ENSEMBLE FEATURE SELECTION AND CLASSIFICATION FOR SOFTWARE QUALITY PREDICTION. International Journal of Reliability, Quality and Safety Engineering.

Garcia, A., Gonzalez, I., Colomo-Palacios, R., Lopez, J. L., & Ruiz, B. (2011). Methodology for software development estimation optimization based on neural networks. Latin America Transactions, IEEE (Revista IEEE America Latina), 9(3), 384-398.

Gashler, M. S., Smith, M. R., Morris, R., & Martinez, T. (2013). Missing Value Imputation With Unsupervised Backpropagation. arXiv preprint arXiv:1312.5394.

Gay, G. (2010, September). A baseline method for search-based software engineering. In Proceedings of the 6th International Conference on Predictive Models in Software Engineering (p. 2). ACM.

Gayatri, N., Nickolas, S., & Reddy, A. V. (2010). Feature selection using decision tree induction in class level metrics dataset for software defect predictions. In Proceedings of the World Congress on Engineering and Computer Science (Vol. 1, pp. 124-129).

Gayatri, N., Nickolas, S., & Reddy, A. V. (2012). ANOVA Discriminant Analysis for Features Selected through Decision Tree Induction Method. In Global Trends in Computing and Communication Systems (pp. 61-70). Springer Berlin Heidelberg.

Gray, D. P. H. (2013). Software Defect Prediction Using Static Code Metrics: Formulating a Methodology. PhD

Gray, D., Bowes, D., Davey, N., Sun, Y., & Christianson, B. (2011). The misuse of the nasa metrics data program data sets for automated software defect prediction. In Proceeding of EASE 2011

Gray, D., Bowes, D., Davey, N., Sun, Y., & Christianson, B. (2012). Reflections on the NASA MDP data sets. Software, IET, 6(6), 549-558.

Hazrati, N. (2011). Geometric Approaches to Statistical Defect Prediction and Learning (Doctoral dissertation, Concordia University).

He, P., Li, B., Liu, X., Chen, J., & Ma, Y. (2014). An Empirical Study on Software Defect Prediction with Simplified Metric Set. arXiv preprint arXiv:1402.3873.

He, Z., Shu, F., Yang, Y., Li, M., & Wang, Q. (2012). An investigation on the feasibility of cross-project defect prediction. Automated Software Engineering, 19(2), 167-199.

Herbold, S. (2013, October). Training data selection for cross-project defect prediction. In Proceedings of the 9th International Conference on Predictive Models in Software Engineering (p. 6). ACM.

Idri, A., & AMAZAL, F. A. (2012, August). Software cost estimation by fuzzy analogy for ISBSG repository. In Proceedings of the 10th International FLINS Conference on Uncertainty Modeling in Knowledge Engineering and Decision Making, Istanbul.

Jin, C., & Jin, S. W. (2014). Applications of fuzzy integrals for predicting software fault-prone. Journal of Intelligent and Fuzzy Systems.

Johansson, U., Konig, R., Lofstrom, T., & Bostrom, H. (2013, June). Evolved decision trees as conformal predictors. In Evolutionary Computation (CEC), 2013 IEEE Congress on (pp. 1794-1801). IEEE.

Johansson, U., Lofstrom, T., & Bostrom, H. (2013, April). Overproduce-and-select: The grim reality. In Computational Intelligence and Ensemble Learning (CIEL), 2013 IEEE Symposium on (pp. 52-59). IEEE.

Keung, J., Kocaguneli, E., & Menzies, T. (2011). A ranking stability indicator for selecting the best effort estimator in software cost estimation. Automated Software Engineering (submitted).

Keung, J., Kocaguneli, E., & Menzies, T. (2013). Finding conclusion stability for selecting the best effort predictor in software effort estimation. Automated Software Engineering, 20(4), 543-567.

Khan, K. (2010). The Evaluation of Well-known Effort Estimation Models based on Predictive Accuracy Indicators.

Khoshgoftaar, T. M., Gao, K., & Napolitano, A. (2012). An empirical study of feature ranking techniques for software quality prediction. International Journal of Software Engineering and Knowledge Engineering, 22(02), 161-183.

Khoshgoftaar, T. M., Gao, K., & Seliya, N. (2010, October). Attribute selection and imbalanced data: Problems in software defect prediction. In Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on (Vol. 1, pp. 137-144). IEEE.

Kocaguneli, E., & Menzies, T. (2013). Software effort models should be assessed via leave-one-out validation. Journal of Systems and Software, 86(7), 1879-1890.

Kocaguneli, E., Cukic, B., & Lu, H. (2013, May). Predicting more from less: synergies of learning. In Realizing Artificial Intelligence Synergies in Software Engineering (RAISE), 2013 2nd International Workshop on (pp. 42-48). IEEE.

Kocaguneli, E., Gay, G., Menzies, T., Yang, Y., & Keung, J. W. (2010, September). When to use data from other projects for effort estimation. In Proceedings of the IEEE/ACM international conference on Automated software engineering (pp. 321-324). ACM.

Kocaguneli, E., Menzies, T., & Keung, J. W. (2012). On the value of ensemble effort estimation. Software Engineering, IEEE Transactions on, 38(6), 1403-1416.

Koru, G., Liu, H., Zhang, D., & El Emam, K. (2010). Testing the theory of relative defect proneness for closed-source software. Empirical Software Engineering, 15(6), 577-598.

Krishnan, S. (2013). Evidence-based defect assessment and prediction for software product lines. PHDThesis

Krishnan, S., Strasburg, C., Lutz, R. R., Goseva-Popstojanova, K., & Dorman, K. S. (2013). Predicting failure-proneness in an evolving software product line. Information and Software Technology, 55(8), 1479-1495.

KUMAR PANDEY, A. J. E. E. T., & Goyal, N. K. (2012). A Fuzzy Model for Early Software Quality Prediction and Module Ranking. International Journal of Performability Engineering, 8(6).

Liparas, D., Angelis, L., & Feldt, R. (2012). Applying the Mahalanobis-Taguchi strategy for software defect diagnosis. Automated Software Engineering, 19(2), 141-165.

Litoriya, R., Sharma, N., & Kothari, A. (2012, September). Incorporating Cost driver substitution to improve the effort using Agile COCOMO II. In Software Engineering (CONSEG), 2012 CSI Sixth International Conference on (pp. 1-7). IEEE.

Liu, Y., Khoshgoftaar, T. M., & Seliya, N. (2010). Evolutionary optimization of software quality modeling with multiple repositories. Software Engineering, IEEE Transactions on, 36(6), 852-864.

Löfström, T., Johansson, U., & Boström, H. (2013). Effective Utilization of Data in Inductive Conformal Prediction. Neural Networks (IJCNN), The 2013 International Joint Conference on, Aug 2013

Lokan, C., & Mendes, E. (2014). Investigating the Use of Duration-based Moving Windows to Improve Software Effort Prediction: a Replicated Study. Information and Software Technology.

Lopez-Martin, C., Isaza, C., & Chavoya, A. (2012). Software development effort prediction of industrial projects applying a general regression neural network. Empirical Software Engineering, 17(6), 738-756.

Lumpe, M., Vasa, R., Menzies, T., Rush, R., & Turhan, B. (2012). Learning better inspection optimization policies. International Journal of Software Engineering and Knowledge Engineering, 22(05), 621-644.

Ma, Y., Luo, G., Zeng, X., & Chen, A. (2012). Transfer learning for cross-company software defect prediction. Information and Software Technology, 54(3), 248-256.

Mausa, G., Grbac, T. G., & Basic, B. D. (2012, May). Multivariate logistic regression prediction of fault-proneness in software modules. In MIPRO, 2012 Proceedings of the 35th International Convention (pp. 698-703). IEEE.

Mende, T., & Koschke, R. (2010, March). Effort-aware defect prediction models. In Software Maintenance and Reengineering (CSMR), 2010 14th European Conference on (pp. 107-116). IEEE.

Menzies, T., & Zimmermann, T. (2013). Software Analytics: So What?. Software, IEEE, 30(4), 31-37.

Menzies, T., Brady, A., Keung, J., Hihn, J., Williams, S., El-Rawas, O., ... & Boehm, B. (2013). Learning Project Management Decisions: A Case Study with Case-Based Reasoning Versus Data Farming. Software Engineering, IEEE Transactions on,

Menzies, T., Butcher, A., Marcus, A., Zimmermann, T., & Cok, D. (2011, November). Local vs. global models for effort estimation and defect prediction. In Proceedings of the 2011 26th IEEE/ACM International Conference on Automated Software Engineering (pp. 343-351). IEEE Computer Society.

Minku, L. L., & Yao, X. (2011, September). A principled evaluation of ensembles of learning machines for software effort estimation. In Proceedings of the 7th International Conference on Predictive Models in Software Engineering (p. 9). ACM.

Minku, L. L., & Yao, X. (2012, June). Using unreliable data for creating more reliable online learners. In Neural Networks (IJCNN), The 2012 International Joint Conference on (pp. 1-8). IEEE.

Minku, L. L., & Yao, X. (2013). Ensembles and locality: Insight on improving software effort estimation. Information and Software Technology, 55(8), 1512-1528.

Mısırlı, A. T., Bener, A. B., & Turhan, B. (2011). An industrial case study of classifier ensembles for locating software defects. Software Quality Journal, 19(3), 515-536.

Mittas, N., & Angelis, L. (2010). Visual comparison of software cost estimation models by regression error characteristic analysis. Journal of Systems and Software, 83(4), 621-637.

Mittas, N., & Angelis, L. (2013). Ranking and clustering software cost estimation models through a multiple comparisons algorithm. Software Engineering, IEEE Transactions on, 39(4), 537-551.

Mizuno, O., & Hata, H. (2010). An integrated approach to detect fault-prone modules using complexity and text feature metrics. In Advances in Computer Science and Information Technology (pp. 457-468). Springer Berlin Heidelberg.

Mizuno, O., & Hata, H. (2013). A metric to detect fault–prone software modules using text filtering. International Journal of Reliability and Safety, 7(1), 17-31.

Mizuno, O., & Hirata, Y. (2010). Fault-prone module prediction using contents of comment lines. In International Workshop on Empirical Software Engineering in Practice 2010 (IWESEP 2010) (p. 39).

Mohanty, R., Ravi, V., & Patra, M. R. (2010). The application of intelligent and soft-computing techniques to software engineering problems: a review. International Journal of Information and Decision Sciences, 2(3), 233-272.

Nagpal, G., Uddin, M., & Kaur, A. (2012). A Comparative Study of Estimation by Analogy using Data Mining Techniques. Journal of Information Processing Systems, 8(4).

Nagpal, G., Uddin, M., & Kaur, A. (2012). A Hybrid Technique using Grey Relational Analysis and Regression for Software Effort Estimation using Feature Selection. International Journal of Soft Computing and Engineering (IJSCE), 1(6), 20-27.

Nagpal, G., Uddin, M., & Kaur, A. (2014). Grey relational effort analysis technique using robust regression methods for individual projects. International Journal of Computational Intelligence Studies, 3(1), 40-73.

Nassif, A. B., Ho, D., & Capretz, L. F. (2013). Towards an early software estimation using log-linear regression and a multilayer perceptron model. Journal of Systems and Software, 86(1), 144-160.

Nguyen, V. H., & Tran, L. M. S. (2010, September). Predicting vulnerable software components with dependency graphs. In Proceedings of the 6th International Workshop on Security Measurements and Metrics (p. 3). ACM.

Okutan, A., & Yıldız, O. T. (2014). Software defect prediction using Bayesian networks. Empirical Software Engineering, 19(1), 154-181.

Paikari, E., Richter, M. M., & Ruhe, G. (2012). Defect prediction using case-based reasoning: An attribute weighting technique based upon sensitivity analysis in neural networks. International Journal of Software Engineering and Knowledge Engineering, 22(06), 747-768.

Paikari, E., Sun, B., Ruhe, G., & Livani, E. (2011, September). Customization support for CBR-based defect prediction. In Proceedings of the 7th International Conference on Predictive Models in Software Engineering (p. 16). ACM.

Pandey, A. K., & Goyal, N. K. (2010). Predicting Fault-prone Software Module Using Data Mining Technique and Fuzzy Logic. International Journal of Computer and Communication Technology (Special Issue), 2(2-4), 56-63.

Papakroni, V. (2013). Data Carving: Identifying and Removing Irrelevancies in the Data. West Virginia University.

Park, B. J., Oh, S. K., & Pedrycz, W. (2013). The design of polynomial function-based neural network predictors for detection of software defects. Information Sciences, 229, 40-57.

Pelayo, L., & Dick, S. (2012). Evaluating stratification alternatives to improve software defect prediction. Reliability, IEEE Transactions on, 61(2), 516-525.

Peters, F., & Menzies, T. (2012, June). Privacy and utility for defect prediction: Experiments with morph. In Proceedings of the 2012 International Conference on Software Engineering (pp. 189-199). IEEE Press.

Peters, F., Menzies, T., & Marcus, A. (2013, May). Better cross company defect prediction. In Mining Software Repositories (MSR), 2013 10th IEEE Working Conference on (pp. 409-418). IEEE.

Peters, F., Menzies, T., Gong, L., & Zhang, H. (2013). Balancing privacy and utility in cross-company defect prediction. Software Engineering, IEEE Transactions on,

Port, D., Nikora, A., Hayes, J. H., & Huang, L. (2011, January). Text mining support for software requirements: Traceability assurance. In System Sciences (HICSS), 2011 44th Hawaii International Conference on (pp. 1-11). IEEE.

Prakash, B. A., Ashoka, D. V., & Aradhya, V. M. (2013). An Evaluation of Neural Networks Approaches used for Software Effort Estimation. Proc of Int. Conference on Multimedia Processing

Premraj, R., & Herzig, K. (2011, September). Network versus code metrics to predict defects: A replication study. In Empirical Software Engineering and Measurement (ESEM), 2011 International Symposium on (pp. 215-224). IEEE.

Radlinski, L., & Hoffmann, W. (2010). On predicting software development effort using machine learning techniques and local data. International Journal of Software Engineering and Computing, 2(2), 123-136.

Radlinski, L., & Hoffmann, W. (2010). On predicting software development effort using machine learning techniques and local data. International Journal of Software Engineering and Computing, 2(2), 123-136.

Rana, Z. A., Malik, S. A., Shamail, S., & Awais, M. M. (2013, January). Identifying Association between Longer Itemsets and Software Defects. In Neural Information Processing (pp. 133-140). Springer Berlin Heidelberg.

Rodriguez, D., Herraiz Tabernero, I., & Harrison, R. (2012). On software engineering repositories and their open problems.

Rodríguez, D., Ruiz, R., Riquelme, J. C., & Aguilar–Ruiz, J. S. (2012). Searching for rules to detect defective modules: A subgroup discovery approach. Information Sciences, 191, 14-30.

Rodriguez, D., Ruiz, R., Riquelme, J. C., & Harrison, R. (2013). A study of subgroup discovery approaches for defect prediction. Information and Software Technology, 55(10), 1810-1822.

Rodríguez, D., Sicilia, M. A., García, E., & Harrison, R. (2012). Empirical findings on team size and productivity in software development. Journal of Systems and Software, 85(3), 562-570.

Sami, A., & Fakhrahmad, S. M. (2010, March). Design-level metrics estimation based on code metrics. In Proceedings of the 2010 ACM Symposium on Applied Computing (pp. 2531-2535). ACM.

Sankar, K., Kannan, S., & Jennifer, P. (2014). Prediction of Code Fault Using Naive Bayes and SVM Classifiers. Middle-East Journal of Scientific Research, 20(1), 108-113.

Sasidharan, R., & Sriram, P. (2014). Hyper-Quadtree-Based K-Means Algorithm for Software Fault Prediction. In Computational Intelligence, Cyber Security and Computational Models (pp. 107-118). Springer India.

Seliya, N., & Khoshgoftaar, T. M. (2011). The use of decision trees for cost□sensitive classification: an empirical study in software quality prediction. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(5), 448-459.

Seo, Y. S., Bae, D. H., & Jeffery, R. (2013). AREION: Software effort estimation based on multiple regressions with adaptive recursive data partitioning. Information and Software Technology, 55(10), 1710-1725.

Setiono, R., Dejaeger, K., Verbeke, W., Martens, D., & Baesens, B. (2010, October). Software effort prediction using regression rule extraction from neural networks. In Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on (Vol. 2, pp. 45-52). IEEE.

Sharma, N., Bajpai, A., & Litoriya, M. R. (2012). Comparison the various clustering algorithms of weka tools. facilities, International Journal of Emerging Technology and Advanced Engineering, vol 2, Iss5, May 2012,4, 7.

Shepperd, M., Song, Q., Sun, Z., & Mair, C. (2013). Data quality: Some comments on the nasa software defect data sets. Software Engineering, IEEE Transactions on, vol 39, Iss 9

Shihab, E., Kamei, Y., Adams, B., & Hassan, A. E. (2013). Is lines of code a good measure of effort in effort-aware models?. Information and Software Technology, 55(11), 1981-1993.

Slankas, J., & Williams, L. (2013, May). Automated extraction of non-functional requirements in available documentation. In Natural Language Analysis in Software Engineering (NaturaLiSE), 2013 1st International Workshop on (pp. 9-16). IEEE.

Smith, M. R., & Martinez, T. (2011, July). Improving classification accuracy by identifying and removing instances that should be misclassified. In Neural Networks (IJCNN), The 2011 International Joint Conference on (pp. 2690-2697). IEEE.

Smith, M. R., & Martinez, T. (2013). An Extensive Evaluation of Filtering Misclassified Instances in Supervised Classification Tasks. arXiv preprint arXiv:1312.3970.

Smith, M. R., Martinez, T., & Giraud-Carrier, C. (2010). An empirical study of instance hardness (Doctoral dissertation, Brigham Young University. Department of Computer Science).

Smith, M. R., Martinez, T., & Giraud-Carrier, C. (2013). An instance level analysis of data complexity. Machine Learning, 1-32.

Song, Q., Jia, Z., Shepperd, M., Ying, S., & Liu, J. (2011). A general software defect-proneness prediction framework. Software Engineering, IEEE Transactions on, 37(3), 356-370.

Steff, M., & Russo, B. (2011, September). Measuring architectural change for defect estimation and localization. In Empirical Software Engineering and Measurement (ESEM), 2011 International Symposium on (pp. 225-234). IEEE.

Tierno, I. A., & Nunes, D. J. (2012, April). Assessment of Automatically Built Bayesian Networks in Software Effort Prediction. In CIbSE (pp. 196-209).

Tosun Misirli, A., Murphy, B., Zimmermann, T., & Basar Bener, A. (2011, September). An explanatory analysis on eclipse beta-release bugs through in-process metrics. In Proceedings of the 8th international workshop on Software quality (pp. 26-33). ACM.

Tosun, A., Bener, A. B., & Kale, R. (2010, July). AI-Based Software Defect Predictors: Applications and Benefits in a Case Study. In IAAI.

Tosun, A., Bener, A., Turhan, B., & Menzies, T. (2010). Practical considerations in deploying statistical methods for defect prediction: A case study within the Turkish telecommunications industry. Information and Software Technology, 52(11), 1242-1257.

Turhan, B., Tosun Mısırlı, A., & Bener, A. (2013). Empirical evaluation of the effects of mixed project data on learning defect predictors. Information and Software Technology, 55(6), 1101-1118.

Turhan, B., Tosun, A., & Bener, A. (2011, August). Empirical evaluation of mixed-project defect prediction models. In Software Engineering and Advanced Applications (SEAA), 2011 37th EUROMICRO Conference on (pp. 396-403). IEEE.

Twala, B., & Cartwright, M. (2010). Ensemble missing data techniques for software effort prediction. Intelligent Data Analysis, 14(3), 299-331.

Verma, R., & Gupta, A. (2012, September). An approach of attribute selection for reducing false alarms. In Software Engineering (CONSEG), 2012 CSI Sixth International Conference on (pp. 1-7). IEEE.

Vivanco, R., Kamei, Y., Monden, A., Matsumoto, K. I., & Jin, D. (2010, May). Using search-based metric selection and oversampling to predict fault prone modules. In Electrical and Computer Engineering (CCECE), 2010 23rd Canadian Conference on (pp. 1-6). IEEE.

Wang, H., Khoshgoftaar, T. M., & Liang, Q. (2013). A STUDY OF SOFTWARE METRIC SELECTION TECHNIQUES: STABILITY ANALYSIS AND DEFECT PREDICTION MODEL PERFORMANCE. International Journal on Artificial Intelligence Tools, 22(05).

Wang, H., Khoshgoftaar, T. M., Van Hulse, J., & Gao, K. (2011). Metric selection for software defect prediction. International Journal of Software Engineering and Knowledge Engineering, 21(02), 237-257.

Wang, H., Khoshgoftaar, T. M., Wald, R., & Napolitano, A. (2012, August). A novel dataset-similarity-aware approach for evaluating stability of software metric selection techniques. In Information Reuse and Integration (IRI), 2012 IEEE 13th International Conference on (pp. 1-8). IEEE.

Wang, S., & Yao, X. (2013). Using class imbalance learning for software defect prediction. Reliability, IEEE Transactions on,

WANG, S., MINKU, L. L., & YAO, X. (2013). ONLINE CLASS IMBALANCE LEARNING AND ITS APPLICATIONS IN FAULT DETECTION. International Journal of Computational Intelligence and Applications, 12(04).

Wang, T., Li, W., Shi, H., & Liu, Z. (2011). Software Defect Prediction Based on Classifiers Ensemble. Journal of Information & Computational Science, 8(16), 4241-4254.

Wang, X., Zhang, L., & Shi, Y. (2010, August). A Knowledge Discovery Case Study of Software Quality Prediction: ISBSG Database. In Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on (Vol. 3, pp. 219-222). IEEE.

Wen, J., Li, S., Lin, Z., Hu, Y., & Huang, C. (2012). Systematic literature review of machine learning based software development effort estimation models. Information and Software Technology, 54(1), 41-59.

Wieloch, M., Amornborvornwong, S., & Cleland-Huang, J. (2013, May). Trace-by-classification: A machine learning approach to generate trace links for frequently occurring software artifacts. In Traceability in Emerging Forms of Software Engineering (TEFSE), 2013 International Workshop on (pp. 110-114). IEEE.

Xu, J., Ho, D., & Capretz, L. F. (2010). AN EMPIRICAL STUDY ON THE PROCEDURE TO DERIVE SOFTWARE QUALITY ESTIMATION MODELS. International Journal of Computer Science & Information Technology, 2(4).

Yu, L. (2012). Using Negative Binomial Regression Analysis to Predict Software Faults: A Study of Apache Ant. International Journal of Information Technology & Computer Science, 4(8).

Yu, L., & Mishra, A. (2012). Experience in predicting fault-prone software modules using complexity metrics. Quality Technology & Quantitative Management, 9(4), 421-433.

Zafar, H., Rana, Z., Shamail, S., & Awais, M. M. (2012, December). Finding focused itemsets from software defect data. In Multitopic Conference (INMIC), 2012 15th International (pp. 418-423). IEEE.

Zhang, H., Nelson, A., & Menzies, T. (2010, September). On the value of learning from defect dense components for software defect prediction. In Proceedings of the 6th International Conference on Predictive Models in Software Engineering (p. 14). ACM.

Authors' Information

Ana Maria Bautista *– E.T.S. Ingenieros Informáticos. Universidad Politécnica de Madrid, Madrid; e-mail: am.bautista@alumnos.upm.es*

Major Fields of Scientific Research: Artificial Intelligence

Tomás San Feliu *– E.T.S. Ingenieros Informáticos. Universidad Politécnica de Madrid, Madrid; e-mail: tomas.sanfeliu@upm.es*

Major Fields of Scientific Research: Software Engineering, Computer Science

METHOD OF DOMAIN MODELS DESIGNING

Elena Chebanyuk

***Abstract**: Method of domain models designing is proposed in this paper. Domain models are represented as class diagrams. Initial information for domain models designing contains both analytical representation of domain data streams and domain entities.*

Domain data streams are defined using collaboration diagram. Analytical representation of domain data streams as a subset of Cartesian product of the following sets: objects and messages are proposed. Also an approach of decomposition collaboration diagram into fragments for defining data streams is represented.

Rules of defining relations between classes of domain model by means of analysis both analytical and graphical representation of domain data streams are formulated. Both description of domain entities and interconnections between them are represented in terms of algebra describing software static models.

Proposed method involves an iterative approach for domain model designing. Firstly it is necessary to obtain information about class diagram constituents. Then information about methods of classes is complimented using analytical representation of domain data streams. Using both analytical representation of domain data streams and analytical description of entities relations between class diagram constituents are defined and clarified.

An example of designing domain model for domain "designing cutting schemas for leather goods details" is represented. Also information about interconnections of domain entities and domain processes is represented.

***Keywords**: class diagram; collaboration diagram; model transformation; model driven architecture; set-theory tool; transformation rules; cutting schemas designing.*

***ACM Classification Keywords**: D.2.2 Design Tools and Techniques; D.2.11 Software Architectures*

Introduction

Using models in software development processes increases productivity of various development activities, such as domain analysis, automated code generation, designing domain specific languages, representation of a software system with necessary details, testing, requirement analysis, software documentation, code reuse and other tasks. It is a background for development of special technics and approaches for software models transformation.

Often software models are represented as UML diagrams. Most of the models that are used in software development process can be divided into static and dynamic (behavioral).

Static software models emphasize a structure of a software system using objects, attributes, operations, and relationships. Examples of static software models are class diagrams, component diagrams, packages diagrams and composite structure diagrams [Gupta 2012].

Dynamic of behavioral software models emphasize the dynamic behavior of a software system by showing collaborations among objects, processes and data flows and changes to the internal states of objects. It includes collaboration diagrams, sequence diagrams, activity diagrams and state machine diagrams [Gupta, 2012].

Important task of Model Driven Architecture (MDA) is designing of languages, technics, rules and other tools for transformation of behavioral (dynamic) models into static software models. Using of such technics allows applying an MDA approach for raising effectiveness of processes in software development life cycle according agile methodology.

It is a background of appearing series of papers that are devoted to different aspects of software development processes, creating of analytical tools, generating new artifacts from behavioral software models [Gupta, 2012], estimation of code reuse effectiveness, tools for an analytical description of software static models [Chebanyuk, 2013] and other aspects that are based on software models represented in a form of UML diagrams [Acretoaie, 2013; Whittle, 2009; Kappel, 2012].

This paper is a continuation of researches that are presented in paper [Chebanyuk, 2013].

Related works

Paper [Gupta, 2012] represents an approach of generating test cases based on use case models that are refined by state diagrams. State diagrams are transformed into usage graphs and then to usage models from which test cases are generated. Also an automated approach of using Adaptive Agent to automatically generate test scenarios from the UML Activity Diagrams is represented. Also the features of PETA tool for the solving of automatic generation of test cases are presented. PETA tool is Java/eclipse based platform for automated software testing.

But an operation [Gupta, 2012] of representation UML activity diagram as state table and writing it into in to some file is rather consuming when activity diagram is large. Also mechanical errors are possible when test cases are generated.

Recently, several approaches adopting the Model Transformation technics to software development processes have been proposed [Whittle, 2009]. These approaches use the concrete syntax of the source and target models to define transformation rules, and thus propose a change to the overall model transformation mechanism. Namely, the transformation definition directly references the source and target models.

Paper [Acretoaie, 2013] is devoted to definition and implementation of a model transformation language focused on usability. This language uses transformation templates and attributes for refactoring models. Before refactoring a model must satisfy to some preconditions. And after refactoring some postconditions must hold true too. But templates that are proposed in paper [Acretoaie, 2013] relate only to class diagram. Use of information from behavioral models allows clarifying patterns, designing new templates, and increasing an effectiveness of models refactoring procedure.

Also an approach originates in the Aspect Oriented Modeling (AOM) field. For example (MATA) [Kappel, 2012] is an aspect composition language based on graph transformation rules expressed in concrete syntax.

Task and challenges

To design transformation rules in order to convert software models of one type (for example static model of one type converts to static model of another type).

But more effective domain models can be designed using a complex approach considering initial information from both types of software models static and dynamic. Also it is necessary to consider information about peculiarities both data streams with and structural components of domain model.

Domain model that is designed according to the proposed method must meet the following requirements of completeness, information content, accuracy and not contradictory. Also the format that is proposed for saving

information about domain models should be compatible with formats of ontology saving (for example RDF or OWL), to allow transformation, refactoring, verification and other operations with models.

Task: to propose the method of domain models designing. Domain model must meet the following requirements: completeness, information content, accuracy and not contradictory. In order to meet this requirements domain model should be designed using information from both behavioral software models and structural components of domain. In order to analyze domain processes and data streams to propose the rules of collaboration diagrams analysis. To design rules for mapping fragments of collaboration and class diagrams.

Analytical representation of domain processes

Represent an analytical description of domain processes, using behavioral software models, namely collaboration diagram. Appointment of a collaboration diagram is the next "Collaboration and sequence diagrams are used to capture dynamic interactions between objects and system. A collaboration diagram consists of objects and associations that describe how the objects communicate. An interaction occurs when two or more objects are used together to accomplish one complete task [Gupta, 2012].

Introduce the following notation.

A set of collaboration diagram objects is denoted as *Object*.

A set of collaboration diagram messages is denoted as *Messages*.

Consider a process of some domain entity creation.

Data streams that are occurring before some domain entity creation are denoted as input data streams.

A subset of the set *Object* that contains objects that are used in input data streams is denoted as $Object^{entity}$.

$Messages^{entity}$ - is a subset of the set *Messages that* are used in input streams of some domain entity creation. This entity should be matched with some object in collaboration diagram, namely having the same name.

Subsets $Object^{entity}$ and $Messages^{entity}$ contain only those objects and messages that are directly interconnected with considering collaboration diagram object (or domain entity). This interconnection is represented by arrows in graphical notation of collaboration diagram.

Consider any collaboration diagram element $object \in Object$. Describe a process of creating domain entity as a subset of Cartesian product of the following sets: $Object^{entity}$ and $Messages^{entity}$.

$$\begin{cases} Entity \subseteq Object^{entity} \times Messages^{entity} \\ Object^{entity} = \{object_0,...,object_n\} \\ Messages^{entity} = \{message_0,...,message_m\} \\ Object^{entity} \times Messages^{entity} = < object_0 message_0 >, < object_1 message_0 > .. < .object_n message_m > \end{cases} \quad (1)$$

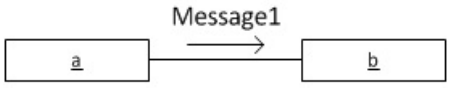
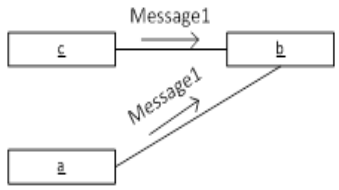
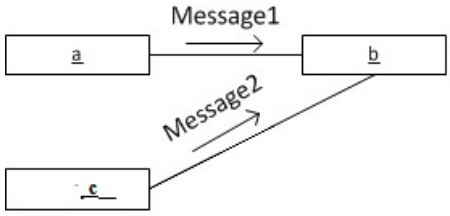
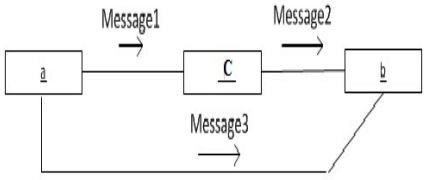
Expression (1) shows that for creation of some domain entity objects and messages can interoperate in any order. It allows usage of alternative data streams for domain entities creation.

Templates for input data streams analysis

Proving the condition of reliability collaboration diagram interconnections shows actual domain data streams. Collaboration diagram analysis shows, that graphical notation proposes several types of input data streams. We

analyze them in order to represent templates for transformation of collaboration diagram fragments into their analytical representation and class diagrams fragments.

Table 1. Templates for input data streams analysis

Graphical representation of input streams	Interconnections between objects
1	2
 **Figure 1.** The first type of input data stream	Template one (Figure 1) shows the first type of input data stream. Analytical description of this template is denoted as $$a \to (Message1) \to b \quad (2)$$ This template shows that in order to create object B it is necessary to use an object A or any of its properties. This template corresponds to composition relation between objects A and B.
 **Figure 2.** The second type of input data stream	Template two (Figure 2) shows the second type of input data stream. Analytical description of this template is denoted as $$a,c \to (Message1) \to b \quad (3)$$ This template shows that in order to create object B it is necessary to use objects A and C or any of theirs properties. This template corresponds to composition relations between pairs of objects A, B and B, C respectively. *The note If the number of input data streams will be more than two an input data stream will be considered as template two.*
 **Figure 3.1.** The third type of input data stream  **Figure 3.2.** The fourth type of input data stream	Template three (Figure 3.1) shows the third type of input data stream. Analytical description of this template is denoted as $$a,c \to (Message1, Message2) \to b \quad (4)$$ Template three shows that there are alternative input streams for creation of object B. This template corresponds to aggregation relations between pairs of objects A, B and B, C respectively. *The note If the number of input objects data streams are more than two an input data stream will be considered as template three.* Template four (Figure 3.2) shows the fourth type of input data stream. Analytical description of this template is denoted as $$a,c \to (Message2, Message3) \to b \quad (5)$$ Template four also shows that there are alternative input streams for creation of object B. As analytical description of template four matches with analytical description of template three we consider that this template also corresponds to aggregation relations between pairs of objects A, B and B, C respectively.

Prove that streams of the first and second types correspond to composition relations.

Consider a stream of the first type. A set from n operations that are executed by object B is denoted as $\mathrm{E}^b = \{op_1, op_2, ..., op_n\}$. Number of operations in the set E^b is denoted as N^b; number of operations that are depended upon object A or its properties is denoted as $N^b(a)$.

In order to prove composition relations it is necessary to prove that $N^b(a) = N^b$.

Consider that $N^b(a) < N^b$. Then it is necessary to find at least one input interconnection between object B and other collaboration diagram objects. Then an analytical representation of input stream will meet to (4) or (5).But input stream corresponds to the condition (2). That is to create an object B we need to involve an object A or some of its properties. This implies that the object B depends upon the object A or some of its properties and equality is proved.

Proof that input stream of the second type is considering to composition relations between all object pairs of objects A, B and A, C is similar. Data streams are considered pairwise.

Thus, analytical representation of the second type of data streams is denoted as follows:

$$\begin{cases} Entity = b\, b \subseteq Object^b \times Messages^b \\ Object^b = \{a, c\} \\ Messages^b = \{Message_1\} \\ Object^b \times Messages^b =< a, c, Message_1 > \end{cases} \quad (6)$$

Prove that streams of the third and fourth types correspond to aggregation relations.

In order to define aggregation relations it is necessary to prove that $N^b(a) < N^b$. It shows, that there are some operations, that are made by object B and these operations do not depend upon object A.

Consider a case with two input streams, then

$$N^b(a) + N^b(c) = N^b \quad (7)$$

As a collaboration diagram meet the following requirements: completeness, information content, accuracy and not contradictory an object B can be created with the help of two operations, namely

$$a \to (Message1) \to b, or \quad (8)$$

$$c \to (Message2) \to b \quad (9)$$

Let us assume that the object B was created by means of operation (8). Then not all operations depend upon the object C or its properties, otherwise the equality (7) is not proved.

Thus, analytical representation of the third and the forth types of data streams are denoted as follows:

$$\begin{cases} Entity = b\ b \subseteq Object^b \times Messages^b \\ Object^b = \{a, c\} \\ Messages^b = \{Message_1, Message_2\} \\ Object^b \times Messages^b =< a, Message_1 >< c, Message_2 > \end{cases} \quad (10)$$

The note in order to denote a message its number on collaboration diagram is used.

Method of domain models designing

1. Obtaining information about domain entities in terms of algebra describing software static models.

Domain entities are described as classes of this algebra.

Define a class C as a subset of Cartesian product of the following sets: properties – A, fields – X and methods -B

$$\begin{cases} C \subseteq A \times X \times B \\ A \times X \times B = \{< \alpha_1, \chi_1, \beta_1 >, < \alpha_1, \chi_1, \beta_2 >, < \alpha_1, \chi_2, \beta_1 > \ldots < \alpha_n, \chi_m, \beta_k >\} \end{cases} \quad (11)$$

where n – is a number of class C properties, m – is a number of its fields, k – is a number of methods. Every triple can contain one empty set. All properties and method of a class C with modifier private are denoted as follows $C^{private}$, public - C^{public} and protected - $C^{protected}$ respectively. Class C is denoted as follows:

$$C = C^{public} \cup C^{private} \cup C^{protected} \quad (12)$$

At least one set from the expression (12) can't be empty.

All elements of a set X (fields of class C) are related to $C^{private}$, that is

$$C^{private} = C^{private} \cup X \quad (13)$$

If when the description starts the name of class is known class is denoted as follows $C(name)$ [Chebanyuk, 2013].

A set of problem domain entities is denoted as follows:

$$\Theta = \{C(name)_1, C(name)_{2,\ldots,} C(name)_p\}$$

where p – is the number of domain entities.

2. Designing a collaboration diagram showing processes and data streams of domain model

3. Forming a set P from those collaboration diagram objects that match with domain entities names.

$$P = \Theta \cap Object, P \neq \varnothing, Object \neq \varnothing, \Theta \neq \varnothing \quad (14)$$

The power of the set P shows how many domain entities match with names of the collaboration diagram objects.

4. For every $object \in Object$ analytical description of input data streams according to (1) is formed.

5. Clarification of domain entities description

Consider a class C(Name) satisfying the following conditions $C(Name) \in P, C(Name) = Entity$.

A set of methods $B^{C(Name)}$ of this class is supplemented by elements of a set $Messages^{entity}$, namely

$$B^{C(name)|} = B^{C(name)} \cup Messages^{entity} \quad (15)$$

6. Defining association relations between domain entities.

Consider a pair of objects. A class C(Name) satisfying the following conditions

$$C(Name) \in P, C(Name) \neq Entity, C(Name) \in Object^{entity} \quad (16)$$

and collaboration diagram object named Entity. Association relations are set between all classes, that are satisfied to the condition (16) and this object.

According to algebra, describing software static models association relationships are denoted as follows:

$$F(C_0^{|})^{acc} = F(C_0^{|}) \cup \Omega, \Omega \neq \varnothing \quad (17)$$

where $F(C_0^{|})^{acc}$ - is a functionality of class $C_0^{|}$ when it is interconnected by relationship of association with class $C_0^{||}$. Classes $C_0^{|}$ and $C_0^{||}$, that are not interconnected by relationship of inheritance. Define a set of $C_0^{|}$

methods in class that are called from class C_0^{II} as Ω. If classes C_0^{I} and C_0^{II} are interconnected by relationship of association, then the functionality of C_0^{I} spreads on methods from the set Ω.

7. Clarifying association types between domain entities

Consider a class $C(name) \in P$. Denote a set of its properties as $A^{C(name)}$. Consider an element $\alpha^{C(name)} \in \mathrm{A}$. Denote a collaboration diagram object with the same name as property $\alpha^{C(name)} \in \mathrm{A}$ as OB.

Both graphical and analytical representations of object OB are analyzed. Then we compare input stream of object OB with analytical (6), (10) and graphical representation (Figure 1-3) of input streams. If object OB input stream matches to (6) or (10) representation we set proper type of association relation between classes OB and $C(name) \in P$.

This operation is made for every element of the set P.

Analytical description of aggregation and composition relations was also proposed by algebra describing software static models [Chebanyuk, 2013].

Functionality of a class C_0^{I} when it is interconnected by relationship of aggregation with class C_0^{II} is denoted as follows:

$$F(C_0^{I})^{aggr} = F(C_0^{I}) \cup (F(C_0^{II}) \setminus F(C_0^{II\,private})\ ) \quad (18)$$

where $F(C_0^{I})^{aggr}$ - is a functionality of class C_0^{I} when it is interconnected by relationship of aggregation with the class C_0^{II}. When object of type C_0^{II} is created in a method of class C_0^{I}, it is possible to call all methods of this object C_0^{II}, excluding private.

Considering, that aggregation and composition relationships are differ by its content, not by structure when classes C_0^{I} and C_0^{II} are interconnected by relationship of aggregation, the functionality of class C_0^{I} is spreads similar to (14) and is denoted as follows:

$$F(C_0^{I})^{comp} = F(C_0^{I}) \cup (F(C_0^{II}) \setminus F(C_0^{II\ private})) \quad (19)$$

where $F(C_0^{I})^{comp}$ - is a functionality of class C_0^{I} when it is interconnected by relationship of composition with the class C_0^{II}.

The note if between some domain entities association relation was set and after that relation of composition or aggregation was defined more strong relation remains, namely association changes to composition or aggregation. Aggregation can be changed into composition.

Designing domain model for the domain "designing leather goods for cutting schemas"

1. Obtaining information about domain entities in terms of algebra describing software static models.

Consider main domain entities: details, linear effect – L, double grid – W, layout – Li, section – Sec and cutting schema -Sh. Math apparatus for designing of single and double grid was described in paper [Chebanyuk, 2013]. Analytical description of classes $C(Laying)$, $C(Leather\_laying)$ and $C(Leather\_layng\_two)$ also was proposed in this paper.

Consider domain entity – linear effect.

According to [Чупринка, 2011] defining of grid vectors is based on dense combinations of rectangles that have been described around details. When two details are combined together there are six types of linear effects (Figure 4).

The note: if detail is rotated on angle 180 by axis X it's considered as detail of the second type.

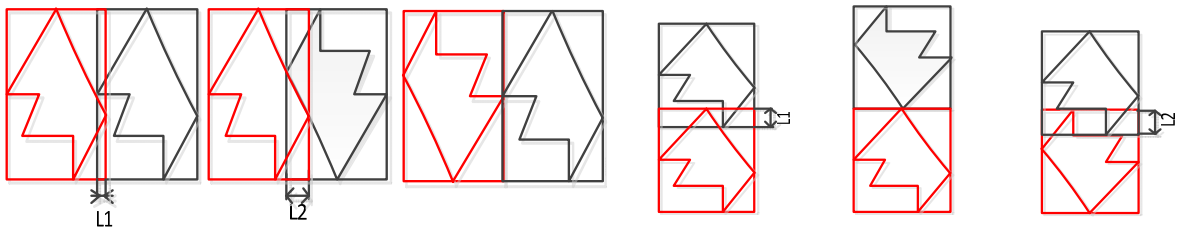


Figure 4. Six types of linear effects when different details are combined together

Class Linear, namely $C(Linear)$, has the next properties – details that are used for defining a linear effect, namely $C(\text{detail}1)$, $C(\text{detail}2)$; grid for constructing of a laying, namely $C(grid)$, vectors for designing of a grid, namely a_1, a_2 and g [Chebanyuk, 2013] and a percent of material usage, namely P.

Methods of class $C(Linear)$ – an estimation of a material usage percent, namely estimate(); creation of a laying, namely - create(); saving of laying parameters, namely - save(); defining maximum linear effect for one detail combination, namely – L1_Max();defining maximum linear effect for two details combination, namely – L2_Max().

Analytical representation of a class $C(Linear)$:

$$\begin{cases} \text{C(Linear)} = \text{AxB} \\ \text{A} = \{\alpha_0,\ldots,\alpha_6 \mid \alpha_0 = \text{C(detail1)}, \alpha_1 = \text{C(detail2)},\ \alpha_2 = \text{C(grid)}, \alpha_3 = \text{P}, \alpha_4 = a_1, \alpha_5 = a_2, \alpha_6 = \\ \text{B} = \{\beta_0,\ldots,\beta_4 \mid \beta_0 = \text{create}, \beta_1 = \text{estimate}, \beta_2 = \text{save}, \beta_3 = \text{L1\_Max()}, \beta_4 = \text{L2\_Max()}\} \end{cases} \quad (20)$$

Using grids layings are built. Domain entity $C(Laying)$ was considered in paper [Chebanyuk, 2013]. Layings are building blocks for constructing layouts.

Consider a domain entity – layout. Denote the height of material as – Height_m, length - Length_m. Layout – it is rectangular area of a material, length Length_L (0< Length_L < Length_m) and height - Height_L (0< Height_L < Height_m) with details that are systematically placed.

Class Layout, namely $C(Layout)$, has the next properties –length and height of layout, namely Height_L and Length_L; details that are used in layout, namely $C(\text{detail}1)$ and $C(\text{detail}2)$; square of this details, namely Sq1 and Sq2; double grid for designing of a laying , namely $C(W)$, density of layout den_L - and a layout period, namely Period.

Methods of class $C(Layout)$ – an estimation of a material usage percent, namely estimate(); creation of a layout, namely - create(); saving of layout parameters, namely - save(); defining layout period, namely – Define_Period(); defining density of layout – Define_density().

Analytical representation of a class $C(Layout)$:

$$\begin{cases} \text{C(Layout)} = \text{AxB} \\ \text{A} = \{\alpha_0,...,\alpha_8 \mid \alpha_0 = \text{Height\_L}, \alpha_1 = \text{Length\_L}, \alpha_2 = \text{C(detail1)}, \alpha_3 = \text{C(detail2)},\ \alpha_4 = \text{C(W)} \\ \alpha_5 = \text{Period}, \alpha_6 = den\_L, \alpha_7 = \text{Sq}_1, \alpha_8 = Sq_2\} \\ \text{B} = \{\beta_0,...,\beta_4 \mid \beta_0 = \text{create}, \beta_1 = \text{estimate}, \beta_2 = \text{save}, \beta_3 = \text{Define\_Period}, \beta_4 = \text{Define\_density}\} \end{cases} \quad (21)$$

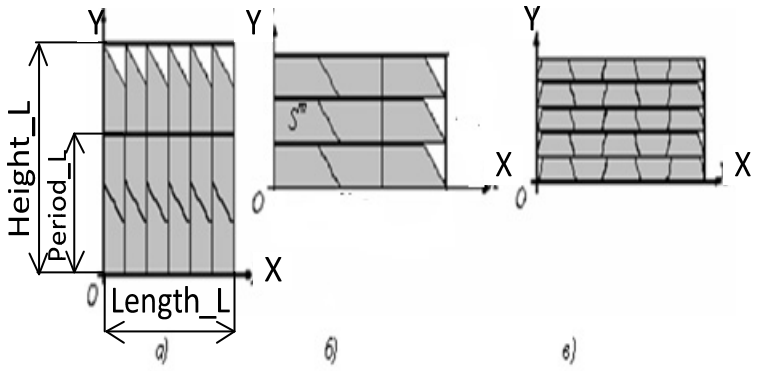


Figure 5. Examples and parameters of layouts

A set of layouts is saved in data structure – List. Operations with the list of layouts are represented in class $C(Lay)$.

Consider a domain entity – section. According to definition section (Figure 6) combines details from some layouts [Чупринка, 2011].

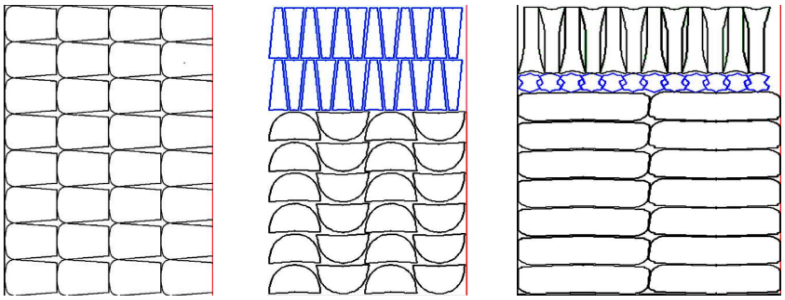


Figure 6. Examples of sections from one, two or three layings

Class Section, namely $C(Section)$, has the next properties – length and height of section, namely Height_S and Length_S; layouts that are used in section, namely list of $C(Lay)$ class; density of section den_S.

Methods of class $C(Section)$ – an estimation of a material usage percent, namely estimate(); creation of a section, namely - create(); saving of section parameters, namely - save(); visualization of section, namely – visualize().

Analytical representation of a class $C(Section)$:

$$\begin{cases} \text{C(Section)} = \text{AxB} \\ \text{A} = \{\alpha_0,...,\alpha_3 \mid \alpha_0 = \text{Height\_S}, \alpha_1 = \text{Length\_S}, \alpha_2 = \text{C(Lay)},\ \alpha_3 = \text{den\_S}\} \\ \text{B} = \{\beta_0,...,\beta_4 \mid \beta_0 = \text{create}, \beta_1 = \text{estimate}, \beta_2 = \text{save}, \beta_3 = \text{Visualize}, \beta_4 = \text{Swap\_Layings}\} \end{cases} \quad (22)$$

Cutting schema consist from sections [Чупринка, 2011]. Screenshot with an example of cutting schema is represented on Figure 7.

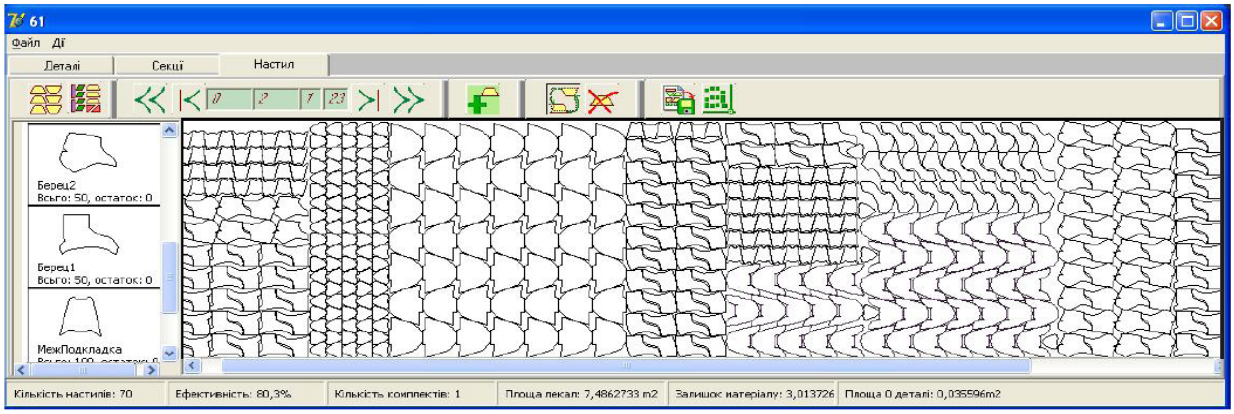


Figure 7. Screenshot with an example of cutting schema

2. Designing a collaboration diagram showing processes and data streams of domain model. Collaboration diagram must meet the following requirements of completeness, information content, accuracy and not contradictory. As collaboration diagram was constructed after detailed analysis of domain entities (20)-(22) it meets all this requirements.

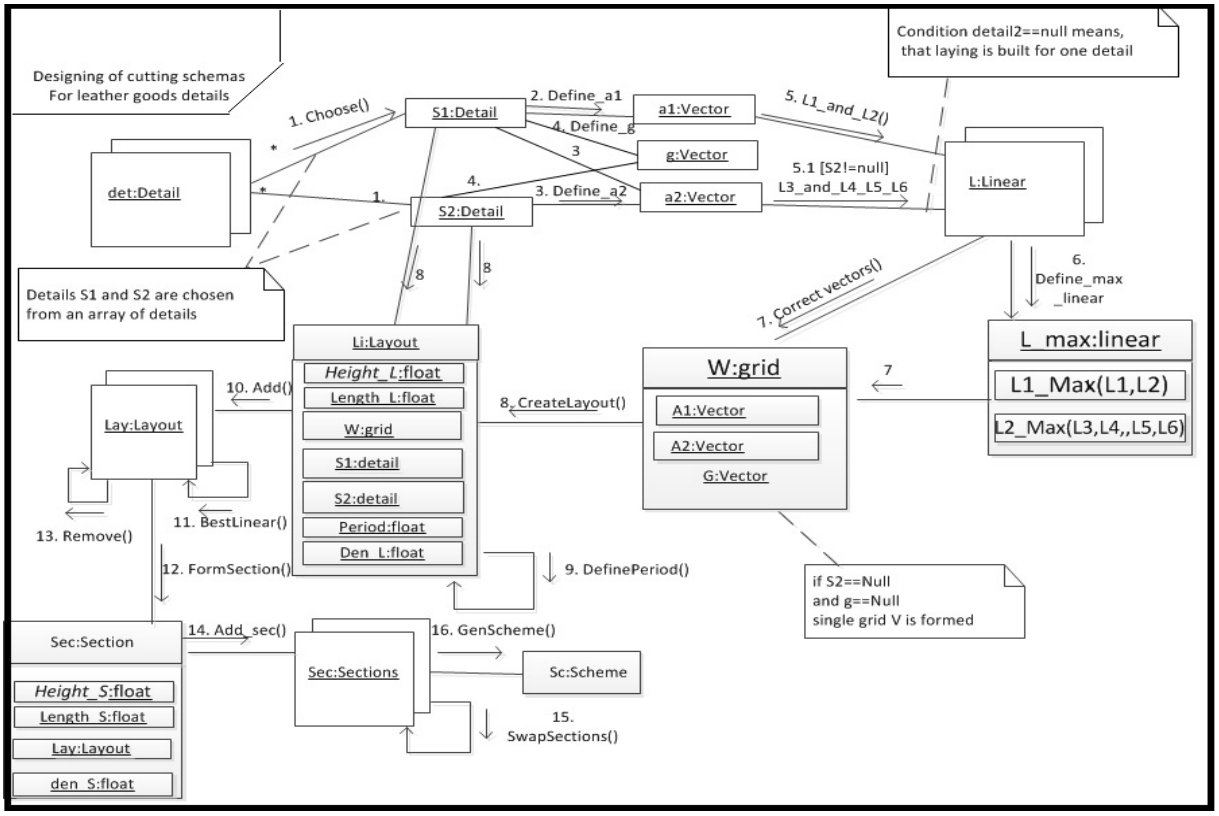


Figure 8. Collaboration diagram, showing streams of domain “designing leather goods for cutting schemas”

3. Forming a set P (14) from those collaboration diagram objects that are matched with domain entities names.

$$P = \{p_0,...,p_5 \mid p_0 = \det, p_1 = L, p_2 = W, p_3 = Li, p_4 = Sec, p_5 = Scheme\} \quad (23)$$

All collaboration diagram objects match with domain classes.

4. For every $object \in Object$ analytical description of input data streams according to (1) is formed.

Describe a process of domain entities creation (1).

$$\begin{cases} Entity = L \; L \subseteq Object^L \times Messages^L \\ Object_L = \{a_1, a_2, g\} \\ Messages_L = \{5, 5.1,\} \\ Object^L \times Messages^L = < a_1, 5 >, < a_1, a_2, g, 5.1 > \end{cases} \quad (24)$$

Consider the expression (24). $Object^L$ - is a set of objects that are used in input data stream when entity L is created, respectively $Messages^L$ is a set of such messages. Elements of the set $Messages^L$ are the numbers of proper messages. Other domain entities are created similarly.

$$\begin{cases} Entity = W \; W \subseteq Object^W \times Messages^w \\ Object^W = \{a_1, a_2, g, L, L\_\max\} \\ Messages^W = \{7, 7.1\} \\ Object^W \times Messages^w = < a_1, a_2, g, 7 >, < L\_\max, 7.1 > \end{cases} \quad (25)$$

$$\begin{cases} Entity = Li \; Li \subseteq Object^{Li} \times Messages^{Li} \\ Object^{Li} = \{W, S_1, S_2\} \\ Messages^{Li} = \{8\} \\ Object^{Li} \times Messages^{Li} = < W, S_1, S_2 8 > \end{cases} \quad (26)$$

$$\begin{cases} Entity = Sec \; S \subseteq Object^{Sec} \times Messages^{Sec} \\ Object^{Sec} = \{Lay\} \\ Messages^{Sec} = \{12\} \\ Object^{Sec} \times Messages^{Sec} = < Lay, 12 > \end{cases} \quad (27)$$

5. Clarification of domain entities description.

5.1. Consider object L. Type of class L is Linear (Figure 8). According to (15) a set of methods B^{Linear} is complimented by elements and class Linear is complimented by methods:

$$(B^{Linear})^* = B^{Linear} \cup L1\_and\_L2 \cup L3\_and\_L4\_L5\_L6$$

5.2 Consider object Li. Type of class L is Layout (Figure 8). Consider its input message CreateLayout (26). A set of methods B^{Layout} contains a constructor. In this case the set B^{Layout} is not complimented.

5.3. Consider object sec. Type of class L is Section (Figure 8). Consider its input message FormSection (Figure 8). A set of methods $B^{section}$ contains a constructor. In this case the set $B^{section}$ is not complimented.

6. Defining association relations between domain entities.

6.1. Define relations between object L (object of type Linear) and other objects that are elements of the set P. Classes a_1, a_2 and g are properties of class L (21) and elements of the set $Object^{L`}$. Considering pairs of classes Linear and a_1, Linear and a_2, Linear and g .Between classes in this pairs association relations are set. According to representation of association relation (17) we can write:

$$\begin{cases} F(L)^{acc} = F(L) \cup \Omega \\ \Omega = \{\beta_0, \beta_1 \mid \beta_0 = L1\_and\_L2, \beta_1 = L3\_and\_L4\_L5\_L6\} \end{cases} \quad (28)$$

6.2. Define relations between object W and other objects that are elements of the set P.

All properties of class W match with elements of the set $Object^W$ (26). Considering pairs of classes W and a_1, W and a_2, W and g, W and L, W and L_max. Between classes in this pairs association relations are set. According to representation of association relation (17) we can write:

$$\begin{cases} F(C(W))^{acc} = F(W) \cup \Omega \\ \Omega = \{\beta_0 = Create\_W\} \end{cases} \quad (29)$$

Analysis of other collaboration diagram objects is made similar.

7. Clarifying association types between domain entities

7.1. Consider properties of class Linear that are matched with elements of the set Piths properties are S1 and S2. We analyze analytical and graphical representation of input streams for these objects. Input streams of objects S1 and S2 corresponds to the first data stream (Figure 1). According to this analysis composition relation between classes Detail and Linear is set.

Check this information using knowledge about domain. According to definition of linear effect it is not defined without information about details (Figure 4) this fact proves composition relations between classes W and detail (S1 and S2 are instances of this class).

$$\{F(C(L))^{comp} = F(C(L)) \cup (F(C(\det)) \setminus F(C(\det^{private}))) \quad (30)$$

7.2 Consider properties of class W that are matched with elements of the set P. These properties are a_1, a_2, g. We analyze analytical and graphical representation of input streams for these objects. Input streams correspond to the second data stream (Figure 2). According to this analysis composition relation between pairs of classes W and a_1, W and a_2, W and g is set. Substitute weak association type to more strong. And composition relations between all these three pairs are remained.

Check this information using knowledge about domain “designing cutting schemas for leather goods details”. According to definition of double grid it is consisted from vectors. This fact proves composition relations between following pairs of classes W and a_1, W and a_2, W and g.

But it is not necessary to use a procedure of calculating linear effects according to some algorithms of cutting schemas designing. This fact proves association relations between classes W and L.

As a result represent an analytical description of relations is denoted as follows

$$\begin{cases} F(C(W))^{acc} = F(W) \cup \Omega \\ \Omega = \{\beta_0 = Create\_W\} \\ F(C(W))^{comp} = C(W) \cup (F(C(a_1)) \setminus F(a_1^{private}) \cup (F(C(a_2)) \setminus F(a_2^{private}) \cup (F(C(g)) \setminus F(g^{private}) \end{cases} \quad (31)$$

Analysis of other collaboration diagram objects is made similar. Represent the analytical description of relations other collaboration object diagrams.

Analytical representation of class Linear relations.

$$\begin{cases} F(C(Li))^{acc} = F(Li) \cup \Omega \\ \Omega = \{\beta_0 = CreateLayout\} \\ F(C(Li))^{comp} = C(W) \cup (F(C(Detail_1)) \setminus F(Detail_1^{private}) \cup (F(C(Detail_2)) \setminus F(Detail_2^{private}) \end{cases} \quad (32)$$

Analytical representation of class Section relations.

$$\begin{cases} F(C(Sec))^{acc} = F(Sec) \cup \Omega \\ \Omega = \{\beta_0 = FormSection\} \\ F(C(Sec))^{comp} = F(C(Sec)) \cup (F(C(Lay)) \setminus F(Lay^{private}) \end{cases} \quad (33)$$

Model of domain "designing cutting schemas for leather goods details" designed according to the proposed method is represented on Figure 9.

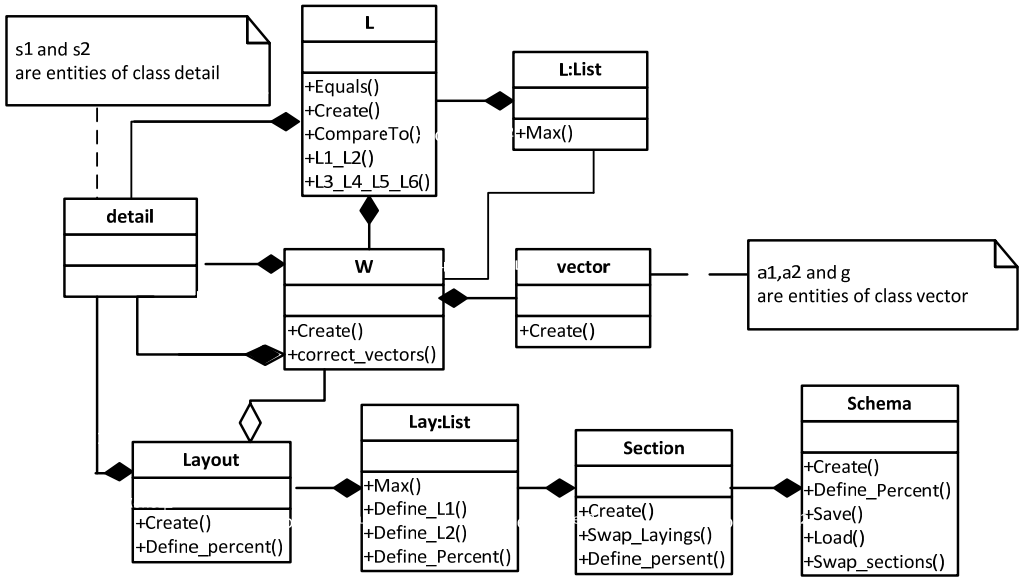


Figure 9. Model of domain "designing cutting schemas for leather goods details"

Conclusion

Review of papers that are devoted to model reuse shows that researchers mostly pay attention to design transformation rules in order to convert software models of one type (for example static model of one type converts to static model of another type).

The peculiarity of the proposed method of domain models designing is that it is necessary to use information both from static and dynamic models. It allows obtaining more initial data that can be considered while domain model designing. And also it influences to the quality of result.

Format of saving information about domain models is compatible with languages for ontology description (for example RDF or OWL). It simplifies the procedure of model refactoring, describing new ontologies, splitting and dividing models and other.

Rules of defining relations between entities that are based on domain streams analysis can be basic for designing transformation technics when other dynamic models are transformed into static. Using of this technics allow to help solving many actual task in MDA area.

Further exploration

Using the method of designing domain models and algebra describing software static models allows to represent frameworks analytically by means of grouping class diagrams constituents according to design patterns and

propose a method of matching problem domain processes to constituents of a class diagram while designing frameworks.

For this it is necessary to propose a concept of mapping processes characteristics and design patterns. This concept also allows defining necessary components from a framework can be reused while analyzing applications functional requirements.

Bibliography

[Acretoaie, 2013] V. Acretoaie. Delivering the Next Generation of Model Transformation Languages and Tools, Europian conference of object oriented programming, pp. 2-10, 2013.

[Chebanyuk, 2013] E. Chebanyuk Algebra describing software static models, INFOS 2013 - "Intelligent Information and Engineering Systems", in press, 2013.

[Gupta, 2012] S. Gupta, J. Singla. A component-based approach for test case generation, International Journal of Information Technology 5/2, pp. 239-243, 2012

[Kappel, 2012] Kappel, G., Langer, P., Retschitzegger, W., Schwinger, W., Wimmer, M., Model transformation by-example: a survey of the _rst wave, in: Conceptual Modeling and Its Theoretical Foundations, pp. 197-215, 2012.

[Whittle, 2009] Whittle, J., Jayaraman, P., Elkhodary, A., Moreira, A., Ara_ujo, J., MATA: A United Approach for Composing UML Aspect Models Based on Graph Transformation, In: Transactions on Aspect-Oriented Software Development VI - Special Issue on Aspects and Model-Driven Engineering, Springer, 2009, pp. 191-237.

[Чупринка, 2011] В.І. Чупринка, О.В. Чебанюк, Автоматизоване проектування раціональних схем розкрою рулонних матеріалів на деталі виробів шкіргалантереї, Вісник Східноукраїнського національного університету ім; Даля №7(2), 2011,с. 46-50.

Authors' Information

Elena Chebanyuk *– lecturer in National aviation university, associate professor of software engineering department, Ukraine; e-mail: chebanyuk.elena@gmail.com*

Major Fields of Scientific Research: Domain analysis, Domain engineering, Code reuse.

METHOD FOR VIDEO SHOT DETECTION AND SEPARATION

David Asatryan, Manuk Zakaryan

Abstract*: Shot boundary detection is main step in video management systems like browsing and indexing. In this paper, we shortly describe an earlier proposed shot detection algorithm based on the structural properties of video frames. Two mathematical models for decision making, i.e. for similarity threshold determination are proposed and compared. The first model allows determination of threshold in case of using mean-square deviation for frames similarity and when the image pixel is assumed to be normal distributed random variable. The second model is based on using the structural similarity measure based on Weibull model for image gradient magnitude distribution. Global and adaptive approaches are considered for frames similarity threshold determination. Results of experiments to detect real video shots are given. It is shown that adaptive threshold determination method generally gives more acceptable results than global threshold determination method. At the same time the* W^2 *based threshold determination approach gives more accurate results than PSNR based approach, therewith corresponds to HVS perception.*

Keywords*: video segmentation, shot detection, similarity measure, threshold determination*

ACM Classification Keywords*: Image Processing and Computer Vision*

Introduction

In recent years content-based indexing and retrieval of digital video becomes an active research area. Why it is so important to have reliable transition detection algorithm and techniques? Firstly, shots are generally considered as the elementary units constituting a video. Detecting shot boundaries means recovering those elementary video units, which can be a ground for existing video abstraction and high-level video segmentation algorithms. Secondly, during video production each transition type is chosen carefully in order to support the content and context of video sequences. So in future it may be easy to automatically detect transitions according to their type and content. Also shot detection can be used in security purposes, so that if a usual scene is changed abruptly, it can be easily detected by controlling tools.

There is several video segmentation algorithms proposed in the literature currently which are based on various statistical properties of the shots. The most of existing methods for video cut detection use some inter-frame difference metric. In frame pair where this difference is greater than a predefined threshold is deemed to be a shot boundary or cut location. They can be classified as: intensity/color based, edge/contour based and motion based [Lienhart, 2001]. In the class of intensity/color– based methods, the histogram–based methods achieve better results comparing with other methods [Gargi, 2000]. Various approaches to histogram–based cut detection were proposed using color based histogram metrics or histogram differences in several color spaces.

In this paper, we will make short overview of our proposed method and algorithm, which already showed the effectiveness in cut detection problems [Zakaryan, 2012; Asatryan, 2014]. The main idea of this method is based to well-known thesis that the human visual system makes an imaginary segmentation mainly using the notional and structural features of the frame.

After having good shot boundary detection algorithms which is mainly based on similarity or dissimilarity measures between consequent frames the next important step of segmentation is to have a threshold which

describes a comparison value. We can divide threshold setting methods into two groups: the fixed threshold method and the adaptive threshold method [Zhi, 2005].

The fixed threshold method determines optimal thresholds from repeated experiments. However, they require much experimental iterations and must find different optimal threshold for various video sequences. Most of them iterate adjustment of thresholds until they get the best results. These methods may have long processing time. In general, variation of thresholds is relatively large to use a fixed threshold for all video sequences. Thus, some algorithms for shot detection were improved by analyzing the whole video sequences for setting multiple thresholds instead of fixed threshold [Cheng, 2002]. These methods may also have long processing time. Thus, it is difficult to apply them to actual applications requiring real-time operations.

Meanwhile, the adaptive threshold based segmentation algorithms get sub-optimal threshold according to [Kim, 2005]. There are several ways for adaptive threshold detection in existing literature. Average calculation in which supposed that when there is a shot boundary, the frame difference is more than the average frame difference before the shot boundary. And if there is a hard cut or shot boundary, the frame difference is also more than the average frame difference after the shot boundary. Another widely used method is sliding window method, when a window of specific length is chosen and similarity comparison between frames is done in rang of this window, and after each step the window is moving forward. Our model of adaptive threshold detection is based on statistical model of image similarity measure described in [Asatryan, 2009].

Proposed Algorithm for Shot Detection

Shot detection algorithm is usually based on consecutive determination of similarity of neighboring frames and detecting abrupt similarities between them. Let consider the frames $f_1, f_2, ..., f_k, f_{k+1}$ and the sequence of corresponding values $\mu_{1,2}, \mu_{2,3}, ..., \mu_{k,k+1}$ of predefined similarity measure μ. When $\mu_{i,i+1} \le t_c$ for $i = 1,2,...,k-1$ and $\mu_{k,k+1} > t_c$ then point k is assumed as a cut point [Asatryan, 2014]. The sequence of frames between neighboring cut points is considered as a shot. The quality of shot detection algorithm is generally depends on properties of chosen similarity measure.

Measure for structural similarity estimation of images. We consider a model of image structure based on the set of edges which are determined by the gradient field of the image. For simplicity we consider the Gray Scale (8 bit) format image $I = \{I(m,n)\}$ with $M \times N$ sizes, $m = 0,1,...,M$; $n = 0,1,...,N$. Let $\|G_H(m,n)\|$ and $\|G_V(m,n)\|$ at a point (m,n) of an image be the horizontal and vertical gradients, determined by one of known gradient methods, and the matrix of gradient magnitude $\|M(m,n)\|$, where

$$M(m,n) = \sqrt{G_H^2(m,n) + G_V^2(m,n)} \tag{1}$$

We follow [Asatryan, 2009] and suppose that the gradient magnitude (1) is a random variable with Weibull distribution density

$$f(x;\eta,\sigma) = \frac{\eta}{\sigma}\left(\frac{x}{\sigma}\right)^{\eta-1} \exp\left[-\left(\frac{x}{\sigma}\right)^{\eta}\right], x \ge 0, \eta > 0, \sigma > 0 \tag{2}$$

As a measure of structural similarity of two images with probability distribution density of gradient magnitude $f_1(x;\eta_1,\sigma_1)$ and $f_2(x;\eta_2,\sigma_2)$ accordingly, we accept

$$W^2 = \frac{\min(\eta_1,\eta_2)\min(\sigma_1,\sigma_2)}{\max(\eta_1,\eta_2)\max(\sigma_1,\sigma_2)},\ 0 \le W^2 \le 1, \tag{3}$$

where the corresponding parameters are represented as statistical assessments gotten from the corresponding samples of gradient magnitude.

Two Mathematical Models for Threshold Determination

As already has been mentioned the fixed threshold which is also called *global threshold* is a technique based on time series of feature values of a similarity measure μ, which in the ideal case is supposed to show a single large peak at hard cut locations. A hard cut is declared each time the feature value $f(t)$ surpasses a globally fixed threshold.

In general, similarity threshold determination is a challenging problem due to large variety of types of video sequences. Therefore this problem is mostly solved in experimental way. Any theoretical approach to this problem requires some general limitations on formal features of video frames. So it is necessary to take into account that the results obtained under these limitations can also have limited reliability [Yi, 2012].

We propose two mathematical models for determination of acceptable threshold when the similarity measures PSNR (peak-to-noise-ratio) and W^2 defined by formula (3) are used.

Global threshold by PSNR. It is well known that when PSNR>30 dB, the human visual system (HVS) practically does not notice the difference between two compared images. But there arises a question: when else HVS is not notice the similarity of images? It can be accepted the following answer: when the structures of images are quite different or images don't have any interpretable structure. In the capacity of such images we can consider random images, i.e. when the intensities of pixels of images are independent random variables. Thus we can set a statistical model which based on assumption that the adjacent frames consists of pixels with independently distributed random intensities. To concretize the model we assume that the intensities have normal distribution $X \sim N(m_j, \sigma_j)$, $j = 1,2$. Then, of course,

$$P\left\{-3\sigma_j \le \left|X - m_j\right| \le 3\sigma_j\right\} \approx 0.997 \tag{4}$$

Let PSNR is calculated by formula as follows

$$PSNR = 10 \log_{10} \frac{\Delta^2}{MSE^2}, \quad MSE^2 = \frac{1}{MN} \sum_m \sum_n \left[I_1(m,n) - I_2(m,n)\right]^2, \tag{5}$$

where $\Delta = \max_{m,n} \left|I_1(m,n) - I_2(m,n)\right|$ is dynamic range of difference between image pixels intensities.

Let, for simplicity, $\sigma_1 = \sigma_2 = \sigma$, $m_1 = m_2$, $M,N >> 1$. Then under (4) we can calculate $\Delta \approx 6\sigma$, $MSE^2 \approx 2\sigma^2$, and

$$PSNR = 10 \log \frac{\Delta^2}{2\sigma^2} = 10 \log \frac{36\sigma^2}{2\sigma^2} \approx 12.5\,dB \tag{6}$$

so we can put $t_c = 12.5\,dB$.

We have shown in [Asatryan, 2014] that the decisions on cut presence using threshold of $t_c = 12.5dB$ conform with visual analysis of PSNR chart for video sequence.

Adaptive threshold by PSNR. In general, determination of adaptive threshold for cut detection can be based on using the samples of similarity measure μ between consecutive frames of the ordinary shot. Decision on cut presence is made as it is described in previous section.

In case of PSNR it is reasonable to use the ratio $\frac{\Delta^2}{\mathrm{MSE}^2}$ instead of PSNR. Threshold t_c is calculated by formula

$$t_c = \bar{X} - lS \text{ for } l = 2 \text{ or } l = 3.$$

Unfortunately, nowadays there is no convenient model for distribution of measure W^2 for random images like considered above. Therefore determination of relevant threshold was performed experimentally. The series of experiments show that acceptable threshold t_c for W^2 varies between 0.6 and 0.8.

Adaptive threshold by measure W^2. For adaptive threshold detection we use a statistical model based on similarity measure of consecutive frames. Unfortunately, the exact distribution of measure W^2 is not yet known, but because of large number of frames in each shot we can use the normal model for asymptotic distribution of W^2 and write

$$P\{\bar{x} - 2s < W^2 \le \bar{x} + 2s\} \approx 0.95 \quad (7)$$

where $\bar{x}$ is the average value and s is the mean square deviation of corresponding samples $W_1^2, W_2^2, ..., W_K^2$, within the current shot, and K is the number of current frame.

If W^2 between K-th and $K+1$-th frames satisfies equation $W^2 < \bar{x} - 2s$, then we consider that there is a cut in $K+1$-th frame.

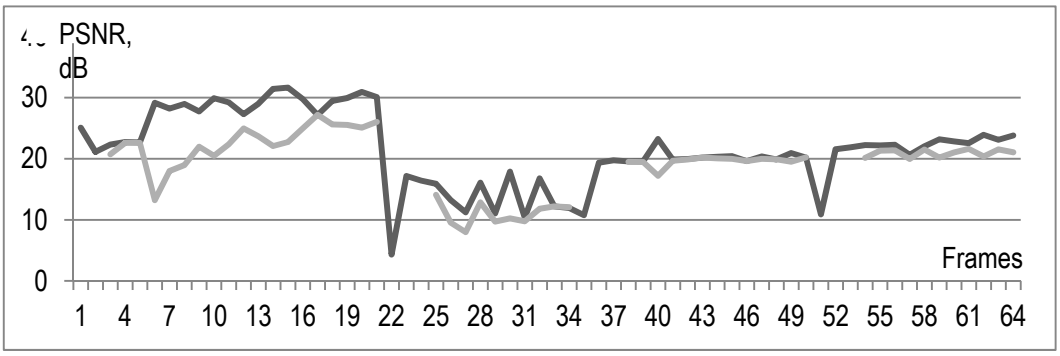


Figure 1. Dependence of PSNR (black curve) and PSNR adaptive threshold (Grey curve) on frame number

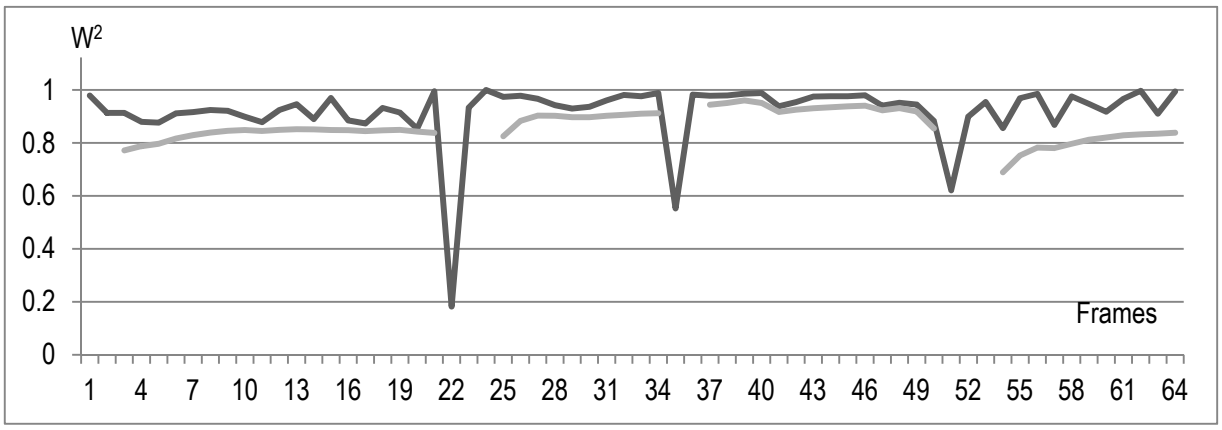


Figure 2. Dependence of W^2 (black curve) and W^2 adaptive threshold (Gray curve) on frame number

Results of Experiments

Described method of shot boundary detection was tested for various video sequences and some results have been given in previous articles [Zakaryan, 2012; Asatryan, 2014]. Here we illustrate the effectiveness of method when adaptive threshold is used.

The results of experiments for both PSNR and W^2 methods are shown in Figure 1 and Figure 2 with black it is shown the similarity graph and the gray one is the adaptive threshold determined above.

More clear result have been reached when we use $l = 3$. As it can be seen from these two figures, for both methods our proposed adaptive algorithm works acceptable, but better and accurate result we gone for W^2 method.

Conclusion

In this paper, we consider the problem of video shots detection and propose two mathematical models for similarity threshold determination. Methods for global and adaptive threshold determination for each model are proposed. Results of experiments to detect real video shots are given. It is shown that adaptive threshold determination method generally gives more acceptable results than global threshold determination method. At the same time the W^2 based threshold determination approach gives more accurate results than PSNR based approach, therewith corresponds to HVS perception.

Bibliography

[Asatryan, 2009] D. Asatryan, K. Egiazarian. Quality Assessment Measure Based on Image Structural Properties. Proc. of the International Workshop on Local and Non-Local Approximation in Image Processing, Finland, Helsinki, pp. 70-73, 2009

[Asatryan, 2014] D.G. Asatryan, M.K. Zakaryan. Improved Algorithm for Video Shot Detection. International Journal "Information Content and Processing", vol. 1, pp. 66-72, Number 1, 2014

[Cheng, 2002] Y. Cheng, X. Yang, and D. Xu, A method for shot boundary detection with automatic threshold, Proceedings of IEEE TENCON, vol. 1, pp. 582-585, 2002.

[Gargi, 2000] U. Gargi, R. Kasturi, S. H. Strayer, Performance Characterization and Comparison of Video - Shot – Change Detection Methods, IEEE Transactions on Circuits and Systems for Video Technology, vol. 10, pp. 1-13, 2000.

[Kim, 2005] W. Kim, K. Moon and J.Kim. An automatic shot change detection algorithm using weighting variance and histogram variation. 11th International Conference on Advanced Communication Technology, vol. 2, pp.1282-1285, 2009.

[Lienhart, 2001] R. Lienhart, Reliable transition detection in videos, A survey and practitioners guide. International Journal of Image and Graphics, vol. 1, pp. 469-486, 2001

[Yi, 2012] Huo Yi, Zhang Pengzhou , Wang Yanfeng, Adaptive threshold based video shot boundary detection framework, Image Analysis and Signal Processing (IASP), 2012 International Conference on Digital Object Identifier, pp. 1- 5, 2012.

[Zakaryan, 2012] M.K. Zakaryan. Novel Approach to Video Cut Detection and Segmentation. В сборнике научных трудов 7-й годичной научной конференции Российско-Армянского (Славянского) университета, vol. 1, pp. 1-5, 2012.

[Zhi, 2005] M. Zhi and A. N. Cai, Shot change detection with adaptive thresholds, IEEE International Workshop on VLSI Design and Video Technology, pp. 147-149, 2005.

Authors' Information

David Asatryan *– Professor, Head of group of the Institute for Informatics and Automation Problems of NAS Armenia, 1, P.Sevaki Str., 0014, Yerevan, Armenia; e-mail: dasat@ipia.sci.am*

Major Fields of Scientific Research: Digital signal and image processing.

Manuk Zakaryan *– Ph.D. student at Russian – Armenian (Slavonic) University, Software Developer at EGS Armenia; e-mail: zakaryanmanuk@yahoo.com*

Major Fields of Scientific Research: Digital signal and image processing, Software developing.

MARKET-REQUIRED COMPETENCE TOPIC DYNAMICS

Tigran Topchyan

Abstract: *The Job market is an ever moving and evolving entity. So are the competencies and qualifications it demands of prospective employees. In our previous work we modelled these competencies using topic models, but in order to have a more effective understanding of the market, we have to take into account the dynamics of the system as well. Here we propose using dynamic topic modelling as a means of analysing the job market for competencies and qualifications stemming from our previous research.*

Keywords: *dynamic topic modelling, qualification, competence, dynamics*

ACM Classification Keywords: *H.0 Information Systems-General*

Introduction

In our previous work we researched the issue of inferring qualifications and analysing them from unstructured data. To this end, we experimented with probabilistic topic models for knowledge extraction [Topchyan, 2014] to extract the underlying topical structure of a dataset of job descriptions. The trained model was able to effectively extract meaningful topics from the dataset, which could be mapped to requirements and qualifications job employers were looking for in potential employees. We outlined the power and flexibility of the extracted model and proposed a number of ways to analyse job description datasets, which could be useful in the context of qualification analysis.

We also extended upon these results and researched the connection between our extracted representations and established competency and qualification ontologies accepted as an industry standard [Topchyan, 2014].

While such an analysis is useful, the job market and as a result competency and qualification requirements, are in constant motion and evolve over time. Some skills rise in importance, some decline. This constant shift is very important to build a comprehensive understanding of the industry requirements. To build an informative model of the competencies and qualifications we would require up-to-date information. For this job description are again an ideal means of analysis, as they are updated daily and shadow the dynamics and evolution of the competencies and qualification requirements. And if we consider the stream of job descriptions as time slices of document corpora we can again apply topic modelling to not only find the topics of the document corpora, but the topic dynamics as well, the topic change over time.

Dynamic Topic Models

In our previous work we only considered static topic models, which models a particular snapshot of a document collection without taking time into account. On the other hand, dynamic topic models view the document collection as slices of documents separated based on time-stamp. Several dynamic topic models have been proposed in the literature [Wang, 2006],[Zhang, 2006]. Here we propose one of the first approaches due to its similarity to the method used in our previous work [Topchyan, 2014] and its proven performance. We could also use standard static LDA [Blei, 2003], but retrained for each slice, but [Blei, 2006] shows that this approach is more effective.

The dynamic topic model (DTM) , as described in [Blei, 2006] models the document corpus as being separated into time slices, which contain a certain set amount of documents. The particular time slices can be any equally spaced out time denominations - hours, days, months, years. Contrary to static LDA [Blei, 2003], where the K topics were

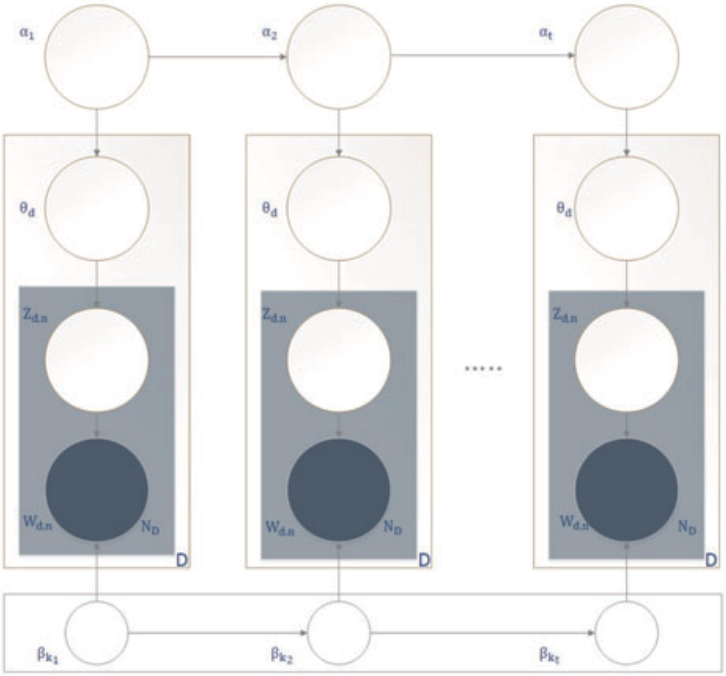


Figure 1: Graphical model for DTM.

described as Dirichlet distribution over a fixed vocabulary V of words β_k, DTM chains the natural parameters of each topic $\beta_{t,k}$ in a state space model that evolves with Gaussian noise;

$$\beta_{t,k}|\beta_{t-1,k} \sim \mathcal{N}(\beta_{t-1,k}, \sigma^2 I) \tag{1}$$

The document-specific topic proportions θ are modelled in a similar fashion by using a logistic normal with mean α to express uncertainty over proportions;

$$\alpha_{t,k}|\alpha_{t-1,k} \sim \mathcal{N}(\alpha_{t-1,k},\ \delta^2 I) \tag{2}$$

The rest of the model can be interpreted as local instances of LDA with the parameters α and β propagating trough time. The generative process itself for a document collection of D document and K topics with t slices. It can be summarised as follows:

1. For topic k in K draw $\beta_{t,k}|\beta_{t-1,k} \sim \mathcal{N}(\beta_{t-1,k}, \sigma^2 I)$
2. Draw $\alpha_{t,k}|\alpha_{t-1,k} \sim \mathcal{N}(\alpha_{t-1,k},\ \delta^2 I)$
3. For document d in D
   (a) Draw $\eta \sim \mathcal{N}(\alpha_t, \alpha^2 I)$
   (b) For each word:
      i. $Z \sim Mult(\pi(\eta))$
      ii. $W_{t,d,n} \sim Mult(\pi(\beta_{t,z}))$

The generative process is also summarised as a graphical model in Figure. 1.

The model inference is an extension of the ideas from the static models. In this case variational Kalman filtering was used. The time slices were all initially initialized as local LDA models and then the filtering was applied for accurate posterior inference [Blei, 2006] .

Analysis

We applied the Dynamic Topic Model on our corpus of ten thousand gathered documents from the IT industry. The documents were gathered across a span six weeks. We trained two models one with 20 topics and one with 100 topics. The training time took 5 and 15 hours respectively. In this section we present the results for the 20 topic model.

Our system supports gathering document from a multitude of sources, but for this experiment we limited it to one in order to gauge the performance and viability of the model for our purposes. Generally the purposed timespan will not contain drastic change, but nonetheless we want to present our findings on the applicability of such a model in the context of competence and qualification analysis.

The DTM is a very robust model and gives us not only the topics in time, but also allows us to infer the document influence model. Meaning we can find how influential each document was to the change in the topics across time. This can potentially have very interesting applications in our research, in particular in the discovery of turning points of competence. But in this work we are more interested in the per-topic dynamics.

Per-Topic: In the context of dynamics analysis we are particularly interested in how each topic changes in time. We are interested in how the topics change in time, meaning how the probabilities of each word change with each time-step. We analysed our corpus and using concepts presented in our previous work [Topchyan, 2014] found the more coherent and popular topics in the dataset and analysed them.

We present the dynamics as a plot of the change of probabilities of four of the most salient words of the topic as well as an embedded wordcloud timeline representation. The size of the words in the wordcloud corresponds to their probability in the given topic. An example of a topic related to practical competences related to databases and problem solving skills can be seen in Figure 2.

From Figure 2 we can see the dynamics of the topic. As we mentioned the time-slices are too close together for any drastic changes to occur, but nonetheless we can notice the steady increase in probability for most of the terms, signifying the fact that the separate time slices contained job descriptions with requirements along the lines of the ones already seen previously. This is fairly logical as the these competencies are fairly popular and sought after in the modern industry and are fairly practical so they occur more often.

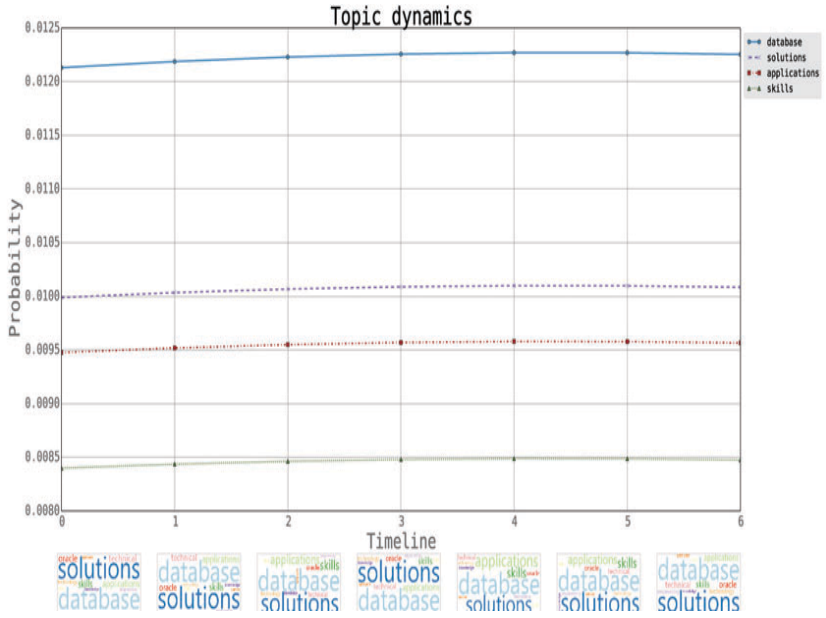


Figure 2: Topic dynamics for database and problem solving competences.

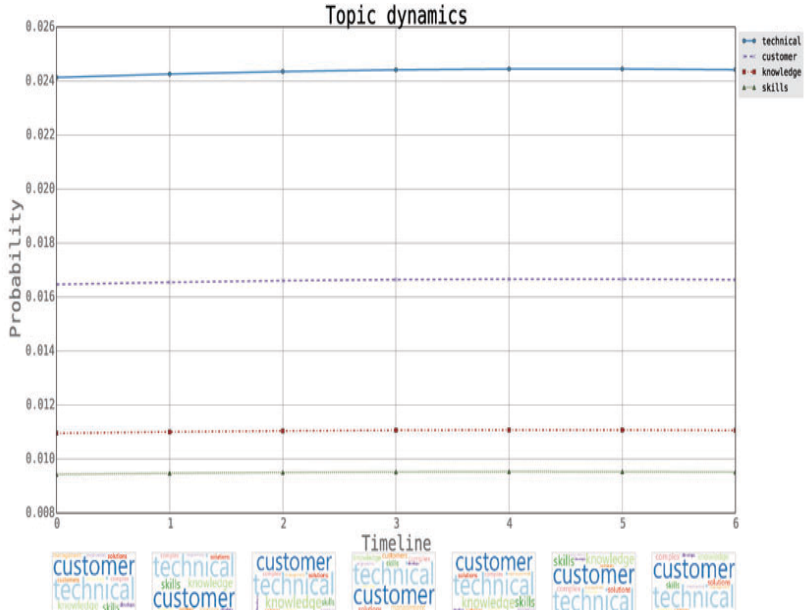


Figure 3: Topic dynamics peronsal and customer relation competences.

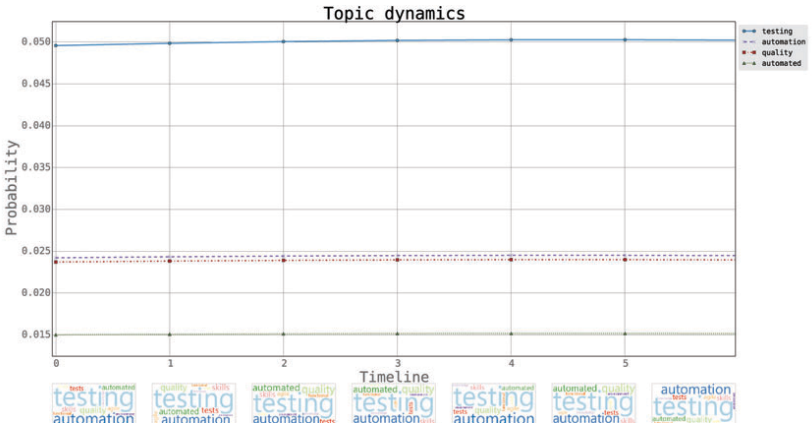


Figure 4: Topic dynamics for topic, that fits perfectly in the CDIO 'IMPLEMENTING' competency group [Crawley, 20], as discussed in our previous work .

On the other hand the dynamics in Figure 3 show that the more probable term start to decrease in probability after a number of time-steps. This can interpreted as the topic contains competences related to more abstract knowledge related to working with customers, and it stands to reason, that in the very practical IT industry there might be a certain drop off in such competence requirements.

Figure 4 present a very interesting result concerning a topic, that is directly related to a CDIO competency group [Crawley, 20], specifically the "Test, Verification, Validation and Certification" competency. We can see that the probabilities rise and do not change thought the timestep. We explain this by the fact, that this is a very sought after competency in the industry and virtually every job description usually contains something concerning testing and verification. This is an encouraging result and proves, that the model is useful for inferring insight from the data.

Conclusion

We researched the viability of using dynamic topic models for analysing competencies and qualification requirements from job market data. The encouraging preliminary results showed that topics extracted using topic modelling can be interpreted to reflect the dynamics of the market and the sought after competencies. In conjunction with

out previous work on competency topic modelling this can lead to invaluable contributions to competency and qualification research in the future.

Bibliography

[Blei, 2003] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. In: Latent dirichlet allocation. J. Mach. Learn. Res, 3:993–1022, mar 2003.

[Blei, 2006] Blei, David M and Lafferty, John D. In: Dynamic topic models. Proceedings of the 23rd international conference on Machine learning, 113–12, 2006.

[Wang, 2006] Wang, Chong and Blei, David and Heckerman, David. In: Continuous time dynamic topic models. arXiv preprint arXiv:1206.3298, 2012.

[Zhang, 2006] Zhang, Jianwen and Song, Yangqiu and Zhang, Changshui and Liu, Shixia. In: Evolutionary hierarchical dirichlet processes for multiple correlated time-varying corpora. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, 1079–1088, 2010.

[Topchyan, 2014] Tigran Topchyan. In: Engineering competence frameworks and topic modelling. Mathematical Problems of Computer Science, 2014.

[Topchyan, 2014] Tigran Topchyan. In: Job Market requirements and qualities extraction for qualification analysis. Mathematical Problems of Computer Science, 2014.

[Crawley, 20] Edward F Crawley, Johan Malmqvist, William A Lucas, and Doris R Brodeur. In: The cdio syllabus v2. 0 an updated statement of goals for engineering education. Proceedings of 7th International CDIO Conference, Copenhagen, Denmark, 2011.

Authors' Information

Tigran Topchyan *Yerevan State University, Yerevan, Armenia; e-mail: tigran@topsoft.am*
Major Fields of Scientific Research: Machine Learning
Education organization

HUMAN AND TELECOMMUNICATION TECHNOLOGIES LIFE CYCLES COMPARISON

Galina Gayvoronska, Illia Ganytskyi, Petr Yatsuk

***Annotation**: Main stages of human life cycle are described. The human and telecommunication technologies life cycle's comparison is done. The general approach to the consideration of telecommunication technologies life cycle is shown. Also the paper shows that the overall picture of telecommunication technology development can be obtained only with an integrated approach to the issue. Only a combination of methodological, statistical approach and development of the standardization process will provide effective assessment, required by companies for development the industry.*

***Keywords**: life cycle, live cycle stages, telecommunication technologies*

***Keywords classification of ACM**: C. Computer Systems Organization D.2.9 Management (K.6.3) Life cycle.*

Introduction

Nowadays, a lot of attention is paid to study the behavior of processes and objects throughout their life cycle. The term life cycle is defined as a set of processes and stages of living nature organism's development, of technical systems, products from the birth, prior the termination, or the end of their use. Graphic representation of the life cycle [Konovalov, 2011] is dome-shaped, consisting of two more or less symmetrical monotonic curves. One shows ascent or becoming the system. Another - the decline and degradation, old age and death. The middle part of the cycle - a relatively gentle transition from one development area to another associated with the maturity of the system when it has a maximum capacity, viability and sustainability.

Big efforts aimed for studying the living organisms and humans life cycle. There are different approaches to the definition of the human life cycle, but most of them can be described as follows: birth, infancy, childhood, adolescence, youth, maturity, wisdom, old age and death. However, mankind has not stopped on the study of living organisms only, and brings his knowledge to other areas of science and technology. In each of the areas appears its lifecycle. The sphere of telecommunications is not an exception. Telecommunication operators need to adequately perceive the new technologies expectations within their industry, choose moderate or aggressive approach to their implementation, combined with investment risk, the need to understand the costs and benefits of getting the introduction of new technologies for the successful operation. To do this, operators need an understanding of the technology life cycle and assess their development stage.

The general approach to the consideration of telecommunication technologies life cycle

Telecommunication operators need to pay attention to four aspects to take the necessary decisions on the technology development. The first aspect - understanding general life cycle of technology, the second aspect - the definition of the current stage of technology development, in which there is an interest, a third aspect - is the collection and processing of statistical data for the technology introduction, and the fourth aspect - forecasting the number of technology users, based on the collected statistics. Consider these aspects in detail.

One of the steps of understanding technology life cycle is the collection various kinds statistical data of specific technology and further analysis of the processing data in different sections. For example, in 1995, a consulting

company Gartner Inc (NYSE: IT), a world leader in the study of information technology, has introduced the concept of Hype cycle (cycle technology maturity). The essence of the concept is that each technology passes the same stages during its existence. These stages are as follows (Figure 1): "Technology Trigger" (the appearance of innovation and articles in the press), "peak of inflated expectations" (waiting for the new revolutionary properties), "trough of disillusionment "(identification of gaps in technology, disappointment), the "slope of enlightenment " (implementation in large companies), and the" plateau of productivity "(maturity of technology, the use with the knowledge of all the positive and negative characteristics).

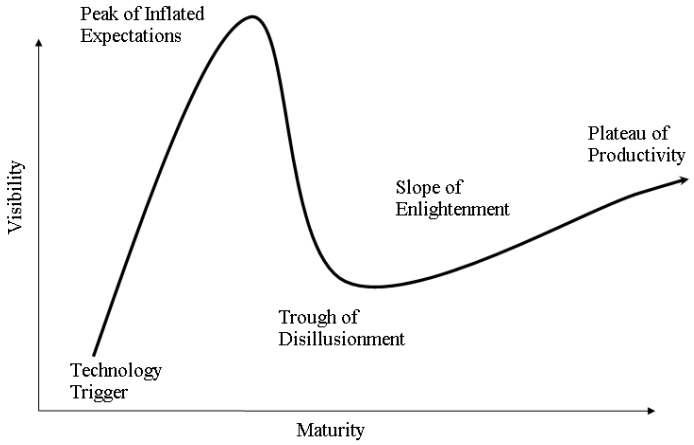


Figure 1. Hype cycle graph

At the time of analysis each specific technology is at some point of its development. Gartner has been working on the collection and aggregation of data from the time of its creation, and gained a considerable amount of statistical data for the analytical predictions. Hype Cycle of new technologies from one of the reports of the company, which was published in Forbes magazine in September 2012 [Hung LeHong, Jackie Fenn, 2012] is presented in Figure 2.

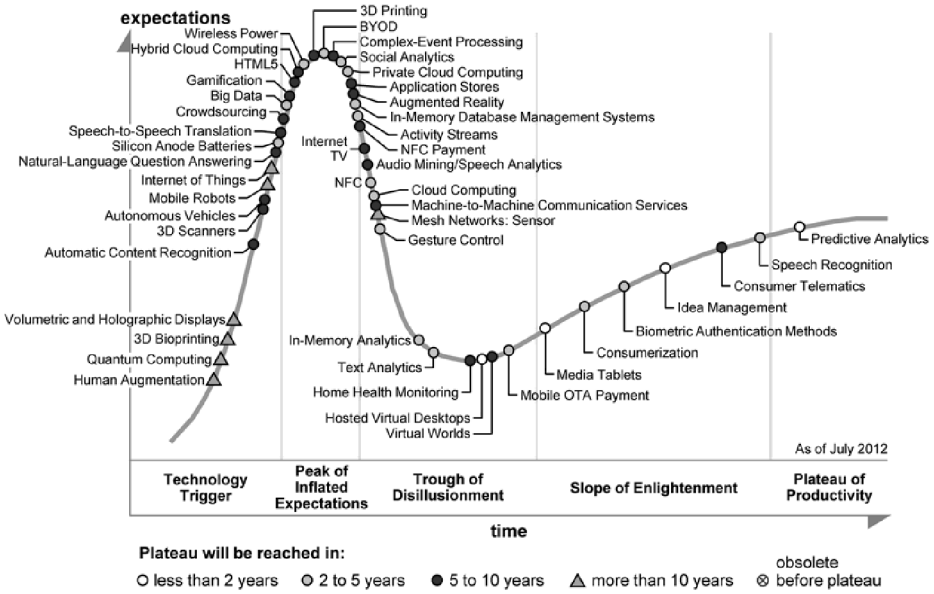


Figure 2. Gartner's 2012 Hype cycle for emerging technologies

This graph - an attempt to show the current state companies of the industry and make a forecast for the latest information technology. Comparison of different technologies - a rather difficult task. Because it makes sense to compare technologies for identical solutions or very similar tasks. On Figure 3, which is taken from the website of the company "Eurasia Telecom" [Sokolov, 2004] the points of some telecommunications technologies for the real conditions are marked. The results of this comparative evaluation will have practical interest only if technology includes all the features of the network in which they operate, so the comparison of technologies appropriate to carry out the specific conditions of each country.

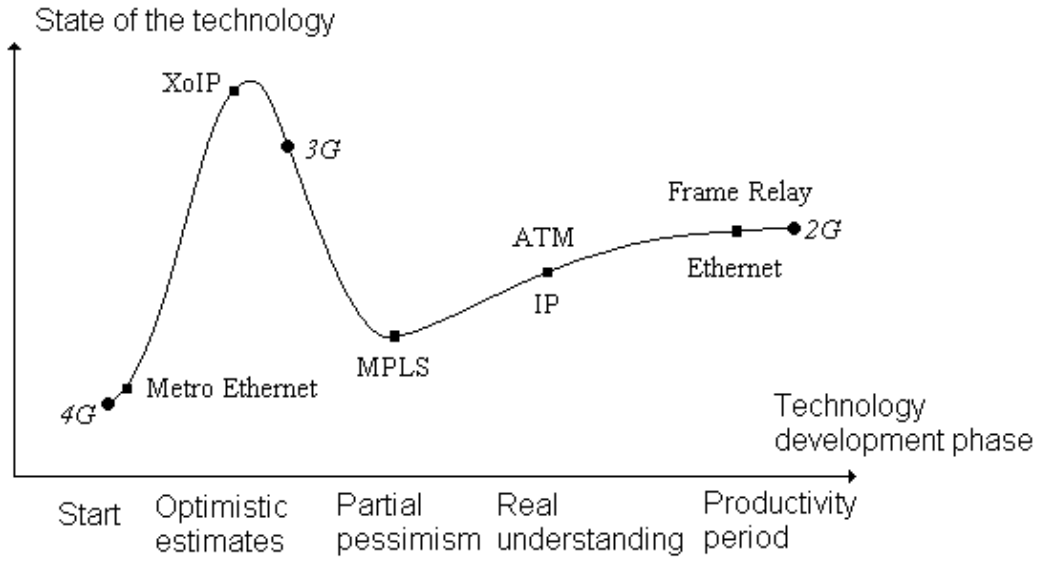


Figure 3. Hype cycle for emerging technologies by "Eurasia Telecom"

Conditions of Ukraine can obtain results that are significantly different from the characteristics of the development of technologies in Western Europe and the U.S., they also need to be interpreted taking about the realities of our country. Completely different conclusions and recommendations on the feasibility of certain technologies can be obtained in result. In this connection there is the problem of studying statistics, determining the life cycles of development of information and communication technologies in Ukraine.

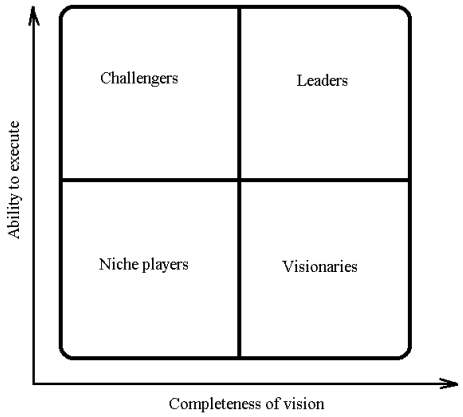


Figure 4. Gather's Magic Quadrant

The concept of Hype cycle gives a necessary vision the condition of the particular technology in specific conditions, especially at the beginning of its development. But for decision to implement this technology, you must also understand the ability the equipment suppliers to provide the necessary services and to have the information about the major players in the market. Gather Inc. developed the approach (Figure 4) called the "Magic Quadrant" for this. This approach is based on an assessment of 2 options: "ability to execute" and "completeness of vision". The first parameter is a generalized assessment of such factors as vendor's financial viability, market responsiveness, product development, sales channels and customer base. The second parameter consists of the evaluation of innovative solutions the company, its views on the current development of the market. This approach, based on a common set of criteria for all technology providers, helps to choose the most successful and competitive company. For clarity, results are displayed using two linear scales that correspond to the parameters. Each provider is applied to the scale where "leaders" have positive scores, and "niche players" are negative. Using these two approaches allows telecommunication operators to evaluate the cycle maturity of the technology in part, its current status, and decides the need and timing of its introduction. If the technology is already at the stage of the "plateau of productivity" according to the Hype cycle graph then there is the problem of estimating the timing of life technology. For this it is necessary to collect statistical information about its use. An example of statistical data needed for use of this approach is shown in Figure 5.

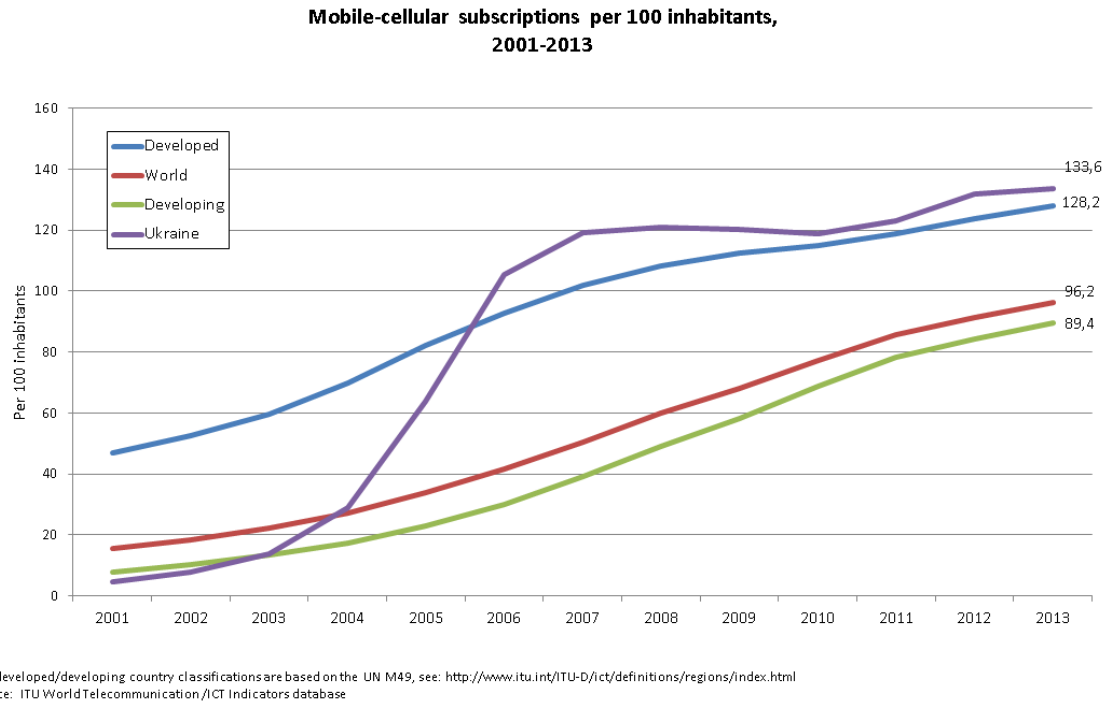


Figure 5. The density of mobile users (2G)

This figure reflects the density of mobile technology GSM users for the period from 2001 to 2013. The figure contains four graphs describing the density of mobile technology GSM standard users for developing countries, for developed countries, for all countries of the world and for Ukraine. As you can see the development of GSM technology in Ukraine matches with the global values at this moment. If we compare this graph with the Hype cycle graph, you can determine that the development of GSM technology for Ukraine is at the upper limit of the "plateau of productivity", which is also confirmed by Figure 3, and see that technology gradual extinction will

begin soon. Such analysis enables network operators to estimate the trends in technology for the country and compare them with the world and determine the stage of their development.

The next aspect in understanding the life cycle of technology is to predict the density of users. The predicted resulting values will evaluate the need to continue the use of a particular technology or its out from service.

Another step towards the understanding lifecycle process, and in particular telecommunications technology, is an attempt to standardize their life cycle. Since the implementation of technology can be hardware, software or hardware and software, the existing standards, dedicated to telecommunication technologies, can be dividing according to this principle. Those standards became high development that governs the life cycle of software (more than 20 standards). The most significant of these are the standards ISO / IEC 12207:2008 System and software engineering - Software life cycle processes, ISO / IEC 15288:2008 System and software engineering - System life cycle processes, IEEE 830-1998 Recommended practice for software requirements specifications. These standards describe the life cycle of software development, and can be used to confirm the stage of development of telecommunication technology, if its implementation involves the development of software.

The main focus is on international standardizing organizations while making description of life cycle for telecommunication technology and its hardware implementation. So European Telecommunications Standards Institute (ETSI) devoted a special technical report ETSI TS 103 199 V1.1.1 (2011-11) for this issue. The methodology of life cycle assessment is described in details in this report for various telecommunication technologies. The standards governing environmental aspect in the life cycle technology received the most development. This is standard ISO 14040:2006 (Environmental management, life cycle assessment, the principles and framework), this is the ISO 14044:2006 (Environmental management, life cycle assessment, requirements and guidelines). In 2011 ETSI published the standard ETSI TS 103 199 Life Cycle Assessment of ICT (Life cycle assessment of information and communication technologies). While these standards have an environmental focus, they are fairly well describe the entire process from creating materials for the production of telecommunication equipment to the withdrawal of equipment from use and the consequences for the environment. These documents have a complete description of the process of creating equipment for existing technologies, and can be used for new ones.

With the approaches proposed in analyzed documents the team of authors is working on the development the method of life cycle assessment of telecommunications technology for the conditions of Ukraine and its application to the research process of these processes and providing prognostic estimates for our country. Some results of these studies will be presented in the report.

Conclusion

The mankind's aspiration for cognition the living organisms and itself, for definition the life cycle passed on other areas of science. We need the understanding what stage of development we are, and what stage of development is for technologies that we create. This understanding is necessary for our success existence. If we talk about the technical field, in particular for telecommunications technology, the understanding of the life cycle and stages of their development is necessary for successful functioning of all the companies working in this field. The paper shows that the overall picture of technology development can be obtained only with an integrated approach to the issue. Only a combination of methodological, statistical approach and development of the standardization process will provide effective assessment, required by companies for development the industry.

Bibliography

[Hung LeHong, Jackie Fenn, 2012] "Key Trends to Watch in Gartner 2012 Emerging Technologies Hype Cycle" / Hung LeHong, Jackie Fenn // Gartner. Access mode: http://www.forbes.com/sites/gartnergroup/2012/09/18/key-trends-to-watch-in-gartner-2012-emerging-technologies-hype-cycle-2/

[Konovalov, 2011] Entropy, deformation, heat and life cycle / Konovalov A.A. // Academy of trinitarism. M., el. № 77-6567, publ. 16543, 03.06.2011.

[Sokolov, 2004] Telecommunication networks / Sokolov N.A. - M. Alvarez Publishing, 2004. - 640 p.

ITU-T. Phoneline networking transceivers – Payload format and link layer requirements. Recommendation G.989.2. – Geneva, 2001.

Authors information

Galina Gayvoronska *– School of information technologies and cybersecurity of V.S. Martinovsky Institute of Refrigeration Cryotechnologies and Ecoenergetics ONAFT, technical science's doctor, professor, chief of the information-communication technologies' department; Dvoryanskaya str., 1/3, Odessa-26, 65026, Ukraine; tel. (048)-720-91-81, e-mail: GSGayvoronska@gmail.com*

Major fields of scientific research: optimization of transient periods at telecommunication networks' evolution. Calls' streams, load and internodal inclination in nets. Problems of perspective access networks' and fully optical switching systems' development.

Illia Gannytskyi *– School of information technologies and cybersecurity of V.S. Martinovsky Institute of Refrigeration Cryotechnologies and Ecoenergetics ONAFT, PhD Telecommunication technologies, lecturer in the information-communication technologies' department, Dvoryanskaya str., 1/3, Odessa–26, 65026, Ukraine; tel. (048)-720-91-48;*

e-mail: igannytskyi@ikt-osar.od.ua

Major fields of scientific research: streams of calls on telecommunication network

Petro Yatsuk *– National Commission for the State Regulation of Communications and Informatization, Chairman, Khreshchatyk Str., 22, Kyiv – 01001, Ukraine, +380 44 2789793;*

e-mail: office@nkrzi.gov.ua , vdz@nkrzi.gov.ua

Major fields of scientific research: telecommunication networks

FEATURES OF PLANNING TECHNOLOGIES IN COMPUTER-AIDED MANAGEMENT SYSTEMS OF AN AGRICULTURAL ENTERPRISE

Iryna Balchenko, Vitaliy Lytvynov, Maria Shulga

***Abstract**: The design principles of the functional subsystem of planning for automatizing management of a crop cultivating agricultural enterprise are shown. Main directions and tasks for the effective work planning of an agricultural enterprise are described.*

***Keywords**: information technologies, planning, automatizing of management, agriculture enterprise.*

***ACM Classification Keywords**: H.4 Information system applications, I.2.1 Applications and Expert Systems, G.1.6 Optimization*

Introduction

The practice of use of information technologies by agricultural producers shows that the attention in the process of designing of a computer-aided management system is paid to implementing accounting functions, while the function of work process management is fulfilled by using traditional methods. This approach to functionality of the system has resulted in following: designing of long-term plans has almost fallen out of the planning system, both current and operational planning problems are mainly implemented in the sole variant, the implementation is carried out in the context of low reliability of the original information [Улезько, 2008].

However, a computer-aided management system of an agricultural enterprise is first of all a management system. It is designed to manage and control any technological processes. When the system was being designed, its specific features affecting automatizing of management were identified. Taking these features into account, a list of functional subsystems was selected. The subsystem of planning plays a special role among these subsystems. The general approach to its construction is described below.

Description of the subsystem of planning of an agricultural enterprise

The purpose of the subsystem of planning:

- Fixing the use of fields under crops;
- Calculating plans of agricultural work and operations;
- Planning the supply of resources necessary to perform agricultural work.

The classification of types of plans

It is supposed to use the following types of plans in the system:

1) Plans of seasonal work, which could potentially be divided into:
   - Plans of work (tillage, sowing, cultivating, harvesting);
   - Plans of operations. They are detailed plans of work, supported by typical technological maps.
2) Executive plans of operations. They are schedules of work for a short period (up to five days). Schedules are supported by technological maps, tied to the state of the fields and forecasted weather conditions.
3) Plans of supply of resources. They determine purchasing and involving the following resources: fuel and lubricants, fertilizers and plant protection products, seeds, tractor operated machinery, staff.

General technology of planning

General technology of planning is:

1) Fixing the fields use for a season based on market needs and crop rotation.
2) Estimating preliminary sequence of agricultural work in accordance with the life cycle of crops.
3) Detailing of operational representation of work and supporting their technological maps (on the basis of regulatory information).
4) Specifying the operations included in the work. A proper adjusting of the plan fragment is made on a five-day period. The adjusting is carried out depending on the state of the field, crops, weather, and provision of the work with resources.

The use of average predictive information concerning the factors that influence cultivating and harvesting is assumed.

Another approach to solving the problem is to consider as much as possible high quality prognostic information and bind the technological maps to field conditions and conditions of acceptable use of machinery (e.g. maintenance and repair). Here the planning process is oriented to the use of advanced machine technology, for example, the use of aggregates, mechanizing performance of groups of agricultural operations in a complex way.

The demand for machine resources, maintenance staff, fuel and lubricants, fertilizers, plant protection products and seeds is easily calculated from the plans of operations supported by refined technological maps with specified types of resources required to perform the operations.

Planning should take into account the specifics of the seasonal work, including time-frames, duration and parallelism of operations within the season. This planning should take into consideration the above aspects especially in its criteria basis. For example, the optimization criteria for cultivating will differ significantly from the criteria for harvesting. This is explained by the fact that the latter requires reducing the duration as much as possible because of the possible loss of yield.

As a central object of planning and management it is advisable to allocate a field with a single crop rotation. The whole subsystem of planning is carried out for it (Figure 1).

The complexity of planning problems provides several types of processes:

- Season;
- Campaign;
- Work;
- Operation.

The automating of planning problems is fulfilled taking them into consideration. The season means the whole cycle of agricultural work from tillage to harvesting.

Each agricultural campaign contains several types of work. The work is understood as a complex of interrelated operations. Work from the basic plan may be included into the executive plan, taking into account the state of the field. The executive plan is a set of operations that must be performed in the nearest future.

The operation – is a single continuous process with common technological features. For example, "early spring tillage" can be executed by means of different operations using different tillage technologies.

On the basis of the executive plan suitable daily schedules of staff, use of equipment, and supply with necessary materials are formed.

On the one hand, such a construction of the plan makes it possible to monitor effectively the implementation of each elementary operation. On the other hand it allows you to "see the whole picture," that is, to ensure control over implementation of the entire plan.

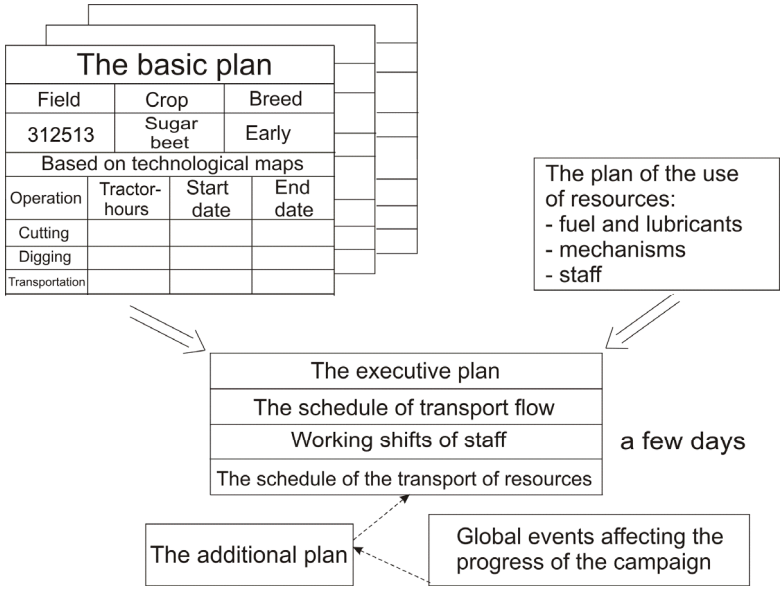


Figure 1. Functional subsystem of planning

Brief characteristics of tasks

To discuss technological features of the subsystem of planning more substantively, its tasks will be briefly reviewed:

1) Planning of annual cultivating states the breeds and amount of crops planned to be cultivated within a season.
2) Planning of fields use fixes the use of each field of the agricultural enterprise within a season for cultivating this or that agricultural crops, taking into account crop rotation.
3) Planning of annual agricultural work in the context of each field and all stages of agrocampaign (sowing, cultivating, harvesting). The structure of the life cycle of a crop determines the list of agricultural work, their optimal terms for each field and each work of agrocampaign.
4) Planning of operations to perform agricultural work for each field and all stages of agrocampaign. It determines further detailing of the plan of agricultural work to the level of agricultural operations. It takes into account weather conditions, soil conditions, weediness, presence of disease clusters and outbreaks of pests.
5) Planning of machine resources required to carry out agricultural works by type of machinery and terms of their use.
6) Planning of use of machine resources, taking into account the process of maintenance and repair, relocation to the place of use, transportation etc.
7) Planning of fertilizers, plant protection products, fuel and materials needed, terms of their purchase and transportation to storage places.
8) Evaluating of required volumes of use of own machine resources, lease and purchase of external resources within an agrocampaign.
9) Preparing of the operation plan of work in relation with the fields.
10) Maintaining standard technological maps of operations (including transport operations supporting basic operations), linking them to the passport information of a field, crops, soil and weather conditions.
11) Creating orders on the use of equipment and invoices to obtain the necessary resources.
12) Correcting of the executive plan of work (considering the information about the soil and crops condition, the volume of work performed) for a field under conditions of each campaign.

13) Taking care of meteorological information, both short-term and long-term weather forecasting.
14) The complex of tasks to plan transporting of harvest. It specifies the use of vehicles during the harvesting.

The order of the tasks in this sequence partially takes into account the sequence of their solving in the process of planning. However, their functionality is different: the tasks of resource supply depend significantly on the task 4. The accuracy and utility of resource planning will be determined depending on the consideration of external influencing factors. Therefore, first of all, an approach to its solving will be described. This approach takes into account external influencing factors on the planning processes deeper. Also some other tasks will be described.

Planning of annual agricultural work in the context of each field and all stages of agrocampaign

To design a plan of annual agricultural work the following data should be considered:

1) Life cycle of the crop on the phases of development (pre-emergence, vegetative, reproductive, etc.), which are characterized by:
   - Typical starting date, duration, ending of the phase under the annual conditions;
   - Qualitative characteristics (number of leaves, etc.);
   - Moisture reserves in the first meter of soil at the beginning of the phase;
   - Number of daylight hours at the beginning of the phase;
   - Average air temperature and precipitation per phase (based on figures for the previous seasons);
   - Quantitative characteristics (mass).
2) Fields, which are characterized by:
   - Area (size);
   - Type of soil (sand, mold, etc.);
   - Mechanical composition and structure of the soil;
   - Types and quantity of fertilizers;
   - Extent of infestation and weed species;
   - Previous crop in the rotation;
   - Crops and their breeds planned to be cultivated.
3) Standard technological maps of crop cultivation.
4) Standard technological maps of operations.

The problem can be formulated as follows: to determine the scope of work for all basic campaigns of cultivating and harvesting of the crop on each field.

Let us single out a common designing technology of an annual plan of agricultural work:

1) If there is a standard technological map of crop cultivating, the scope of work is determined basing on it.
2) If there is no standard technological map of crop cultivating, the scope of work can be determined by taking into account the life-cycle phases of the crop (for the phases of development of crops, coinciding with periods of mechanized field work).

According to the list of phases of crop development and their conformity to the periods of work, the work itself can be determined. Thus, the compiled list of work can serve as a standard technological map of the crop.

With the aid of standard technological maps of specific work and the results of projected phases of crop development the terms of work can be clarified.

The technique of phenological predictions of terms of individual phases of development can be used. The use of the described technique allows specifying terms of the beginning, end and duration of phenological phases of the life cycle of the crop. This technique is based on empirical equations [Арютов, 2010].

The problem solving enables to obtain the stages of work for each field, characterized by:

- Optimal terms;
- A set of standard (including alternative) operations;
- Conditions.

Designing of the executive plan and its providing with technological maps of operations

An executive plan is based on the following factors:

- Short-term weather forecast;
- Current state of the soil;
- Current state of the crop at each phase of the life cycle of its cultivating;
- Supply of resources.

In the process of designing of an executive plan the following points should be determined:

- The possibility of shifting of a specific work from the basic plan to the executive one. This step is performed basing on the short-term weather forecast, prioritization of the work, the terms of the work and compliance of the current state of the crop to the phase of the life cycle of its cultivating;
- The possibility to include additional work into the executive plan. This step is performed in the case of out-of-management condition of the current state of the soil or crop;
- Operation specification of the work of the executive plan;
- A technological map for each operation. The technological map is clarified on the basis of data concerning the state of the soil, crop and resources supply.
- The resources necessary to fulfill the operation, and schedules of their use. This step is performed on the basis of the technological map of the operation and availability of resources for the operation.

While the work is carried out its quality is evaluated. Evaluation is performed in accordance with established standards. Work can be considered completed only when all the operations are over and the quality of the work is confirmed. After that, a corresponding note is entered into the basic plan. Then the work is removed from the executive plan. Otherwise, the operations may be re-added to the executive plan.

Supporting and binding the technological maps of operations to the passport information of fields, crops and weather conditions

Normative-reference information in the form of a technological map of cultivating can exist for the crop rotation. A technological map is the complete description of all operations, their sequence indicating transitions and technological conditions. Typical zonal technological maps of crop cultivating were designed and published for each crop. On the basis of the zonal technological maps planners can design a basic plan of crop cultivation. This plan takes into consideration any specific field conditions, working conditions and the technical equipment of the agricultural enterprise.

Thus, the technological map is transformed into a basic plan of crop cultivation. In this case the technological foundation of the operation remains unchanged. It should be represented by the technological map of the operation. If the technological map of the crop cultivation answers the question: "what should be done to get products?", the operating one answers the question: "how can agricultural work be performed with high quality?". The main purpose of the operating technology is to prevent the shortcomings in the work and to do it in the best

possible time with high quality and the lowest cost of labor and resources. The map must contain seven components [Маслов, 2011]:

1) Initial data or working conditions (the area of the field, its topography, soil and machine resistivity, expenditure norms of seeds, fertilizers and pesticides, crop-producing power of main and sideline products, the distance of transportation of materials).
2) Agrotechnical requirements.
3) The structure of tractor operated machinery, preparing it for use.
4) Work to prepare the field and the working area.
5) Work of the tractor operated machinery in the field.
6) Control and evaluation of work quality.
7) Safety instructions.

Binding of the technological map to different factors (conditions and state of the field, crops, resources, weather forecasts) is done by selecting the most efficient available alternative. Such choice can be made:

- On the basis of formal dependencies of technological norms of the operation on the complex of factors. The dependences can be obtained experimentally;
- When the formalizing of this kind of dependency is impossible – on the basis of expert decisions (of an agronomist, a planner, a controller).

Thus, each operation will be based on such a technological map that describes the individual characteristics of the field in the most effective way.

Planning of resource needs, terms of their purchase and transportation to storage

When designing the basic plan it is necessary to determine the total resource needs. Then a schedule of use of labor and material resources is to be prepared for different cases:

- When own resources are supplied;
- When the resources from outside organizations are supplied.

Following points are to be fulfilled:

- Identifying and distributing the resources required for the operations according to the time;
- Identifying the resource suppliers;
- Taking into account the factors affecting the availability of resources for operations;
- Identifying the terms and conditions of storage of resource stocks;
- Designing the schedules of supply of resources.

The total demand for some types of resources often depends on the intensity of work. First of all it relates to the engaged equipment and staff. Such multicriterial task should be solved by means of designing the schemes of involving and interacting of these resources. Other types of resources, such as seeds, fertilizers, fuel and lubricants, plant protection products can be calculated before the season basing on data about the crop, area and condition of the field, and consumption norms. Designing of schedules of delivery of these resources to the warehouse directly depends on the terms and conditions of their storage.

To optimize resource allocating and resolve resource conflicts alignment methods can be used. They take into account the limits of resource consumption. These methods allow using resources in the most efficient way. Resource leveling can eliminate peaks in the use of resources and set the level of their use below the maximum limit. This is done by shifting some work to later dates. Moreover, the natural shift of work is possible due to the impact of unforeseen external factors. The influence of these factors cannot be eliminated in a short time. Therefore, to establish the terms of the work the apparatus of fuzzy set theory is supposed to be used. The

apparatus implies some "blurring" of the terms of the work. Due to the possibility of emergencies resources may be needed in somewhat larger volumes. In this case, total resource requirements can be planned on somewhat larger scale. Or supply with resources should be planned when designing additional work.

Correcting of the executive plan of work

The perfect fulfilling of plans is known to be impossible to reach. Therefore, when designing a subsystem of planning the specific character of extraordinary situations must be taken into account. Global extraordinary situations are events associated with unsatisfactory condition of crops or soil, disruption of operations, or even a whole range of works. When planning the additional work it is necessary to correlate the current condition of the soil or crop with the terms, the cost of additional work and the extent of possible improvement of the soil or crop condition. Thus, an expert should participate in the decision making that concerns including the additional work into the executive plan.

The executive plan represents the current set of operations that must be performed taking into account the condition of crops, soil and external factors. Considering the foregoing, the executive plan can be viewed as an important component of the mechanism of adaptation of the system managing an agriculture enterprise.

The role of the cartographic system

Unfortunately, it is rather difficult to obtain measuring information concerning the conditions of agricultural objects. The complexity is determined by the length of large objects of work and other specific factors. Shortage of measurement data leads to considerable uncertainty of the results. Practically most of the data is formed only on the basis of agronomists' observations [Garam, Pashko, 2005]. However, it is better to estimate with the help of special equipment:

- By direct contact;
- Without contact by means of remote sensing.

The obtained data can be analyzed using GIS technologies. Spatial analysis provides real processes of dynamics of objects, events or processes. The real processes of dynamics of spatial analysis are obtained by processing and applying a wide range of well-known techniques and models [Balchenko, 2012]. So, geoinformation spatial data analysis can be represented in the form of thematic cartograms of spread of contamination of crops and soil by pests (Figure 2), weeds, diseases, as well as a variety of maps of the movement control of machinery and motor transport, weather forecast

Thus, in case of emergency of situations of global nature, it is necessary to provide the ability to create additional operations. The additional operations can be based on the basic operations regulating the order of actions in case when certain kinds of events take place. For example, in the case of pests, the action can be scheduled to eliminate them by appropriate means.

Conclusion

The concepts in designing the functional subsystem of planning provided by the authors will enable to rationalize the whole system of management of an agricultural enterprise. This is accomplished through the use of modern methods of information-analytical support along with mathematical methods of planning.

The important factors increasing the scientific validity of the designed plans are information technology and mathematical methods of planning of agricultural works. Using the methods of mathematical programming and multicriterial optimization is the most reasonable way of planning. The presented method allows obtaining plans balanced by various criteria of optimality.

Figure 2. Example of use of cartograms to demonstrate the spread of contamination by pests

Bibliography

[Balchenko, 2012] Balchenko I.V., Kazimir V.V., Klimenko V.P. Modern methods of control over a condition of an agrarian ecosystem. // Journal of Chernihiv State Technological University, 2012. – № 3 (59). – P. 213-220.

[Garam, Pashko, 2005] Garam V. P., Pashko A.O. Modern control of agrotechnical processes in the plant growing. // Science and innovation, 2005. – T1. № 2. – P. 110-116.

[Арютов, 2010] Арютов Б.А., Важенин А.Н., Пасин А.В. Методы повышения эффективности механизированных производственных процессов по условиям их функционирования в растениеводстве: Учебное пособие. – Издательство "Академия Естествознания", 2010.

[Маслов, 2011] Маслов Г.Г., Припоров Е.В., Палапин А.В. Разработка операционных технологий выполнения сельскохозяйственных механизированных работ (методические рекомендации). – Краснодар: КубГАУ, 2011 – 191с.

[Улезько, 2008] Улезько А.В. Информационное обеспечение адаптивного управления в аграрных формированиях /А.В. Улезько, Я.И. Денисов, А.А. Тютюников. – Воронеж: изд-во «Истоки», 2008. – 106 с.

Authors' Information

Iryna Balchenko *– Postgraduate, the assistant lecturer, Chernihiv National Technological University, 95, Shevchenko street, Chernihiv-27, Ukraine, 14027; e-mail: phiona06@yandex.ru*

Major Fields of Scientific Research: modeling of complicated systems, information technology in planning of agricultural work

Vitaliy Lytvynov *– Dr. Sc. Prof. Chernihiv National Technological University, 95, Shevchenko street, Chernihiv-27, Ukraine, 14027; e-mail: vlitvin@ukrsoft.ua*

Major Fields of Scientific Research: modeling of complicated systems, computer-aided management systems, decision support systems

Maria Shulga *– MSc student, Chernihiv National Technological University, 95, Shevchenko street, Chernihiv-27, Ukraine, 14027; e-mail: shulga-marichka@mail.ru*

Major Fields of Scientific Research: expert systems in the planning of agricultural work

ALGORITHMS OF AUTOMATE MODEL CONSTRUCTION FOR BUSINESS GAME EXECUTION SUBSYSTEM

Alexandr Deryabin, Lidia Shestakova, Olga Vikentyeva

Abstract: *Changes of professional environment, caused by introduction of new technologies and techniques, create a necessity in continuous education and development of professional competences. In these conditions, managers and other company employees face the choice of methods and tools of personnel training. A business game is one of the most productive tools of business-education.*

This paper elaborates the ideas embodied previously, which considered the conceptual approach to a toolkit creation for active training techniques in a form of competence-based business game studio. Competence-based business game studio is an ergatic system for development of professional competences, which are required to ensure organization's business processes.

Business game control is performed with the help of automate module, which executes the interpretation of expressions in algorithm logical scheme language.

Unified business process is the input data for automate model construction. Unified business process is acquired from the information about company's real business processes, represented in poorly formalized ways. Unified business process transforms into unified training business process, which includes possible business game trainee actions and information about input, output, administrative data and business process operation execution mechanism. The next step to automate model is the construction of scenario graph, which represents a more formalized description of unified training business process. Next, the administrating algorithm of the business game in algorithm logical scheme language is built. The paper considers the algorithms of transition between models that allow automating the process of acquiring algorithm logical scheme in order to implement the business games scenario.

Keywords*: active learning methods, business-game, business-process, automate model, algorithm logical scheme, algorithm, metamodels*

ACM Classification Keywords*: K.3 Computers and Education: K.3.2 Computer and Information Science Education – Information systems education. K.4 Computers and Society: K.4.3 Organizational Impacts – Employment. I. Computing Methodologies: I.2 Artificial Intelligence: I.2.1 Applications and Expert Systems – Games.*

Introduction

Changes of professional environment, caused by introduction of new technologies and techniques, create a necessity in continuous education and development of professional competences. In these conditions, managers and other company employees face the choice of methods and tools of personnel training. Currently, e-learning systems have gained much popularity. Implementation of such systems enables to cut costs related to training organization in companies and educational institutions through process automation and use of Internet technologies. However, the majority of e-learning systems only support the traditional education model slightly modifying it with contemporary communication technologies. Hence, a certain trend must be noticed, the trend

implies use of active training techniques and business games in particular [Zichermann, 2011], [Wells,1990]. A business game (BG) is one of the most productive tools of business-education. Business games allow trainees to acquire hands-on experience, fosters professional competences and teamwork efficiency in real-like conditions.

This paper elaborates the ideas previously embodied in [Deryabin, 2013], [Викентьева, 2013], [Викентьева, 2014], which considered the conceptual approach to a toolkit creation for active training techniques in a form of competence-based business game studio (CBGS). CBGS is an ergatic system for development of professional competences, which are required to ensure organization's business processes.

Domain formalization for automate model construction

Automate model is constituted by a sequence of operators in algorithm logical schema (ALS) language, which implements the algorithm of business game control. Any expression in ALS language may be represented as following:

L = {H, A, P, ω, ↑, ↓, K}, where,

H – algorithm start operator;

K – algorithm finish operator;

P – conditional transfer;

A – control action;

ω– unconditional transfer;

↑ - transfer beginning;

↓ - transfer end.

BG control is performed with the help of automate module, which executes the interpretation of ALS expressions. Furthermore, the module recognizes each consecutive operator and either conveys subsequent game scene code to the operational module, or carries out the conditional or unconditional transfer to the next operator, depending on the game status code received from the operational module.

Unified business process (UBP) is the input data for automate model construction. UBP is acquired from the information about company's real business processes, represented in poorly formalized ways (graphical models, text description, and regulatory documents). UBP transforms into unified training business process (UTBP), which includes map of operations that contains a tree of possible business process operations (BG trainee actions) and operation model, which contains information about input, output, administrative data and business process operation execution mechanism. The map of operations also includes decision making points, which stand for player (user) interaction during the business game. Figure 1 uses following notation: UBP – unified business process; UTBP – unified training business process; RBP– real business process; MOp – map of operations; OP – operation model; DMP – decision making point; SG – scenario graph; ALS – algorithm logical schema; AM – automate model; OM – operational model.

After the construction of UTBP the process of BG design may be separated into two steps:

- operational model construction, which includes scene model, resource model and screen model;
- automate model construction, which is needed to build ALS expressions control the business game.

This paper considers the process of automate model construction and algorithms used for UTBP's map of operations transformation into a sequence of models. This sequence is required to build ALS expressions.

DSM-platform Metaedit+ [Mazanek, 2008], [Сухов, 2009] was used to develop interrelated metamodels for UTBP description: "Operation", "Map of Operations" and "Decision Making Point".

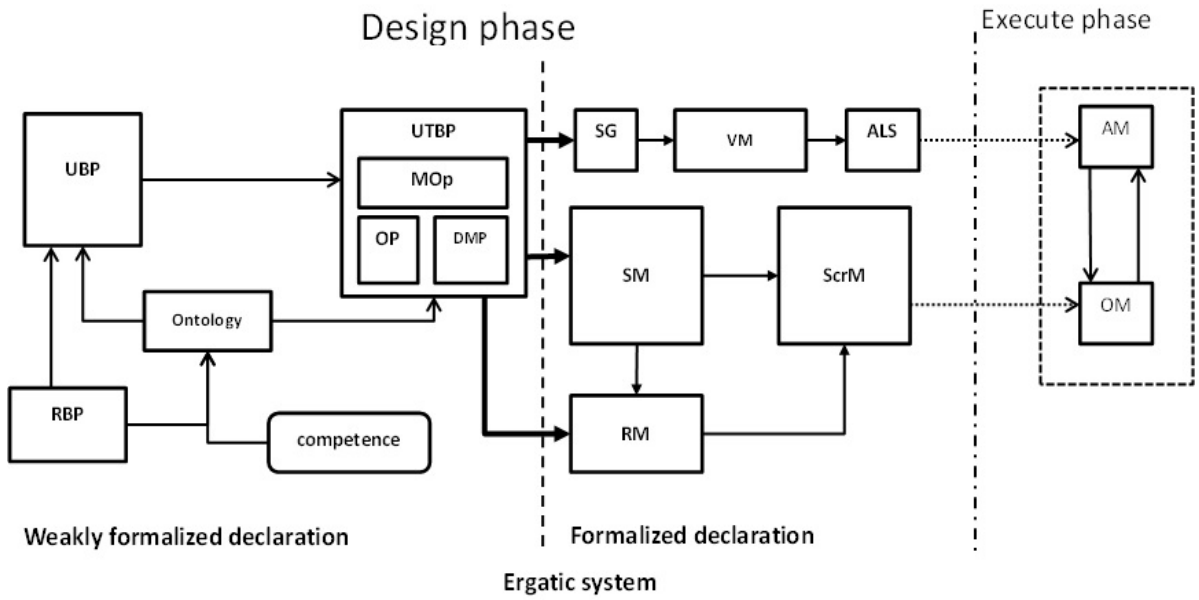


Figure 1. Domain formalization for automate model construction

"Operation" metamodel describes separate operations (tasks), which constitute the business process and include resources (informational, financial, manpower), equipment, actors etc.

"Map of Operations" metamodel enables to describe business process as a multivariant sequence of operations and moments where player has to make decisions.

"Decision Making Point" metamodel allows to describe the process of decision-making, unfolding it through a sequence of reactions.

Let us consider an instance of streamlined business process (Figure 2), which consists of actions D1, D2, D3, D4 and business condition BC1 that determines the rule of business process execution.

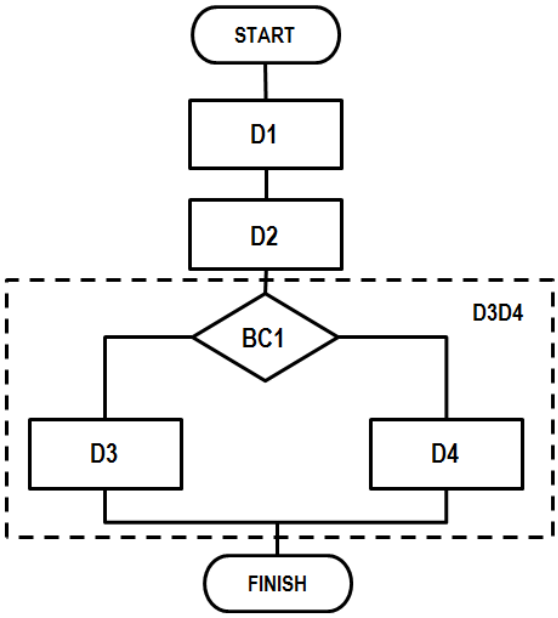


Figure 2. The process of transformations from UBP to unified training business process

Let us build the map of operations for the business process (Figure 2) with the help of “Map of Operations” and “Decision Making Point” metamodels. The map of operations represents a tree, which includes decision making points S and actions D. Decision making point enables transfer from one action to another, with transfer from a single point to multiple actions being possible. Each branch of the tree may contain action Di only once. Thus, the result is a hierarchical structure, where each branch represents a possible business game path.

UBP may be quite difficult and include not only sequential operations, but various business conditions and recurring operations as well. To build the UTBP, it is proposed to replace the conditional and cyclical structures with generalized blocks. This enables to describe the UBTP as a linear sequence of operations. A new map of operations is the built for each generalized block.

Hence, the algorithm of UBP to map of operations transformation may be as following:

1. All blocks with business conditions in the UBP must be replaced with generalized blocks. This also applies to linear block sequences if needed. The replacement must be carried out in such way that the SBP may be executed linearly (convert the UBP to a linear form). In result, the UBP will contain only generalized actions D1, …, Dn.

2. Build start and finish blocks.

3. Build a list of possible actions D = {D1, …, Dn, Dk}, where Dk – transfer to the finish block, Di – execution of action i.

4. Build decision making point (DMP) S0.

5. Build branches for the current DMP regarding the actions from list D.

6. For each branch built, except the branches with transfer to the finish block:

    6.1. Build a list of possible actions, that does not contain Di: D={D1,…, Dn, Dk}\{Di}.

    6.2. If list D={Dk}, build a transfer to the finish block and return to step 6, otherwise go to step 6.3.

    6.3. Build decision making point St.

    6.4. Build branches regarding the actions from list D.

    6.5. Return to step 6.

Figure 3 depicts the result of algorithm application to the UBP (Figure 2). Generalized blocks with business conditions have the same structure. Map of operations of a block must contain a DMP (Si0), where the player would be able to choose different courses of actions with or without condition testing. If the test is present, it is possible to pick one of the two actions: test is positive (BC=1) or test in negative (BC=0). In both cases DMPs are used to choose actions. Execution of a generalized block ends with a DMP to alleviate the introduction of the block into the general map of operations. More complicated cases, which imply execution of several actions in one block, are resolved by contraction of operations into a single generalized block. Figure 4 shows the map of operations for generalized block D3D4 (Figure 3) with business conditions (Figure 2).

The next step to automate model is the construction of scenario graph (SG), which represents a more formalized description of UTBP map of operations.

Scenario graph (Figure 5) is a graph, with BG scenes as graph’s vertexes. These scenes contain conditions of transitions and corresponding DMPs of the UTBP. Edges of the graph comply with the actions, taken during the transitions between scenes. The amount of edges and vertexes is large enough to employ graph representation that includes a shared bus with indexed edges.

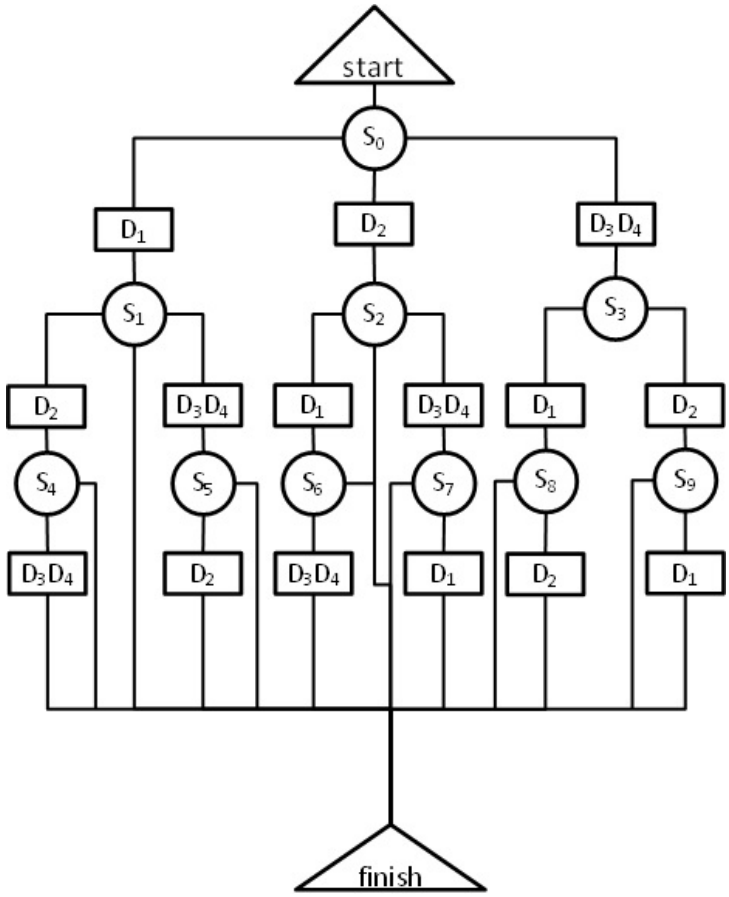


Figure 3. UTBP Map of Operations

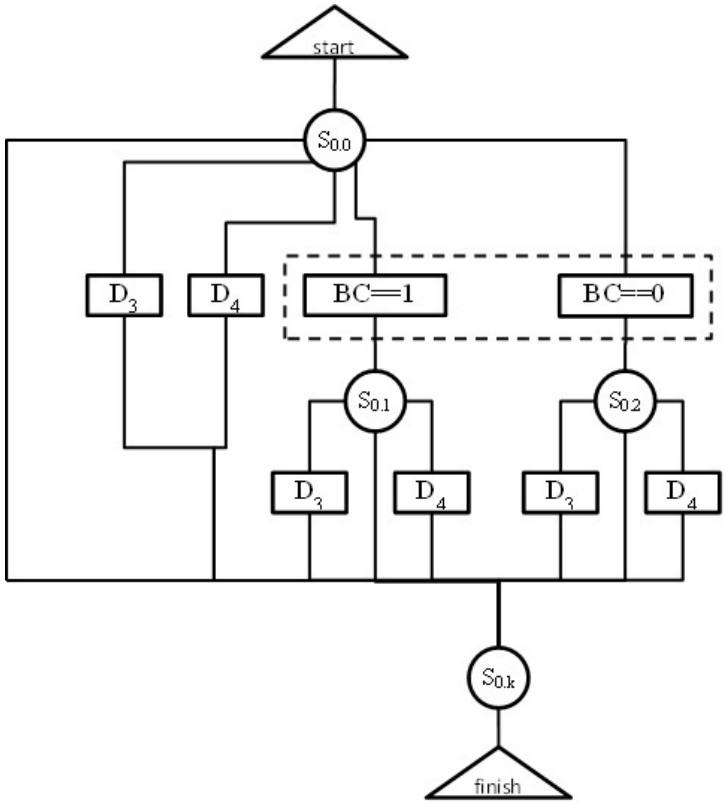


Figure 4. Generalized Block's Map of Operations

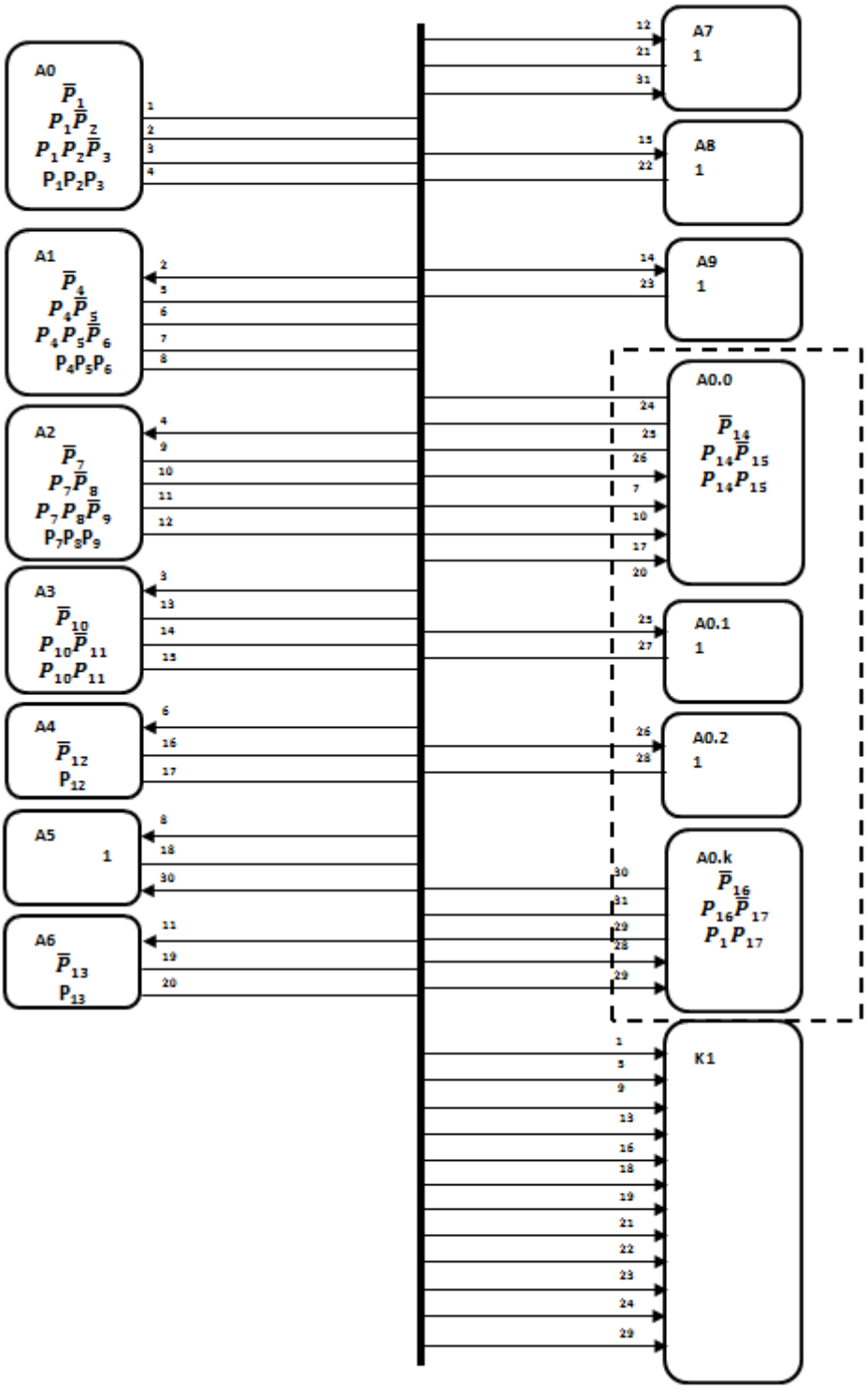


Figure 5. Scenario Graph

Let us consider the algorithm of SG construction:

1. Each decision making point (DMP) of the UTBP map of operations is put at the corresponding SG vertex Ai.

2. For every action from the action list D of each DMP in the map of operations a set of conditions P must be made. The set is written down to the corresponding SG vertex.

2.1. The amount of elements in condition set is defined as $k_P = \log_2(k_D) + 1$, where k_D – amount of elements in list D.

2.2. The set of conditions is formed as following:

2.2.1. Define index of the first condition in the set *as [index of the last condition in previous DMP + 1* and generate indexes for each condition in the set {*P1, P2, …, Pk$_p$+1*}.

2.2.2. Define condition values in the set: first condition – $\overline{P1}$, second condition – $P1\overline{P2}$, third condition $P1P2\overline{P3}$, last condition – $P1P2, \ldots Pk_P + 1$. $\overline{Pi}$ – negation of condition $\overline{Pi}$.

3. For each SG vertex build outgoing edges corresponding to the conditions of the previously formed set. Then assign each edge sequential indexes starting with 1.

4. Each outgoing edge has a corresponding entering edge with the same index, which is built according to the following rule: each SG vertex Ai has a corresponding DMP, represent the connection between DMPs Si and Sj as a connection between SG vertexes Ai and Aj and build an entering edge. Entering edges have direction.

5. If a branch connection DMPs has a generalized block, do as following:

5.1. For the generalized block a SG vertex Bk is built (0<=k<=n, n – amount of generalized blocks). Entering edges correspond to edges outgoing from vertex Ai. Outgoing edge correspond to edges entering vertex Aj.

5.2. Generalized block is represented as a SG with vertexes Aij, where i – index of the block in a sequence of generalized blocks, j – index of a DMP in the generalized block i.

5.3. Edges entering Bk transform into edges entering block Ai0.

5.4. Edges outgoing from Bk transform into edges outgoing from block Aik.

5.5. Дуги, входящие в Bk, преобразуются во входные дуги блока Ai0.

Thus, SG contains conditions for transitions between BG scenes. To be able to transform a SG into an algorithm in ALS language that can executed by the automate module, it is needed to build a vertex-adjacency matrix (Figure 6). Matrix elements are represented by the sets of conditions, which are tested during transitions between SG vertexes.

Vertex-adjacency matrix is built as following:

1. Fill the matrix with zeroes.

2. If the transition between vertexes Ai and Aj exists, then the corresponding matrix cell is filled with the transition condition P

3. If the transition between vertexes Ai and Aj exists and is unconditional, the corresponding matrix cell is filled with number 1.

	A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	K1	V0	V1	V2	Vk
A0	0	$P_1\overline{P}_2$	$P_1P_2\overline{P}_3$	$P_1P_2P_3$	0	0	0	0	0	0	$\overline{P}_1$	0	0	0	0
A1	0	0	0	0	$P_4\overline{P}_5$	$P_4P_5\overline{P}_6$	0	0	0	0	$\acute{P}_4$	$P_4P_5P_6$	0	0	0
A2	0	0	0	0	0	0	$P_7P_8\overline{P}_9$	$P_7P_8P_9$	0	0	$\overline{P}_7$	$P_7\overline{P}_8$	0	0	0
A3	0	0	0	0	0	0	0	0	$P_{10}P_{11}$	$P_{10}\overline{P}_{11}$	$\overline{P}_{10}$	0	0	0	0
A4	0	0	0	0	0	0	0	0	0	0	$\overline{P}_{12}$	P_{12}	0	0	0
A5	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
A6	0	0	0	0	0	0	0	0	0	0	$\overline{P}_{13}$	P_{13}	0	0	0
A7	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
A8	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
A9	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
K1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V0	0	0	0	0	0	0	0	0	0	0	$\overline{P}_{14}$	0	$P_{14}P_{15}$	$P_{14}\overline{P}_{15}$	0
V1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
V2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Vk	0	0	0	0	0	$P_{16}P_{17}$	0	$P_{16}\overline{P}_{17}$	0	0	$\overline{P}_{16}$	0	0	0	0

Figure 6. Vertex-adjacency matrix with logical

The base of conditions for vertex Ai is the condition set P with the maximum length in the vertex-adjacency matrix row. The set P must correspond to vertex Ai. For instance, for vertex A0 (fig. 6) base of conditions is represented by set P1P2P3.

Using the vertex-adjacency matrix it possible to derive the administrating algorithm of the BG in ALS language as following:

1. Write the start and finish operators (H, K).

2. Write SG vertex names (Ai) between operators H and K.

3. For each operator Ai write base of conditions and operator of unconditional transition ω (coincides with 1 in the vertex-adjacency matrix).

4. After each condition and operator of unconditional transition write ↑. For instance, base of conditions and transition operators for operator A0 will be P1↑P2↑P3↑ω↑.

5. In front of every operator Ai put operator ↓.

Indexes of operators ↑ and ↓ (transitions between Ai and Aj) are defined as following:

6. Initial index t=0.

7. From A0 to K:

7.1. Consider every condition Pk in string Aj and assign index t for Pk↑t.

7.2. If the Ai and Aj cells of adjacency matrix have $\overline{Pk}$, then the transition between Ai and Aj is executed according to this condition. In this case it is necessary to assign index t to operator ↓, which stands in front of operator Aj, where Aj is the column of the matrix cell with condition $\overline{Pk}$. Then increase counter t by 1.

If the Ai, Aj cell has no negated conditions, then the transition from Ai to Aj is executed through operator of unconditional transition with the corresponding index. Increase counter t by 1.

If the transition is executed towards already indexed operator ↓, then this index does not change, rather the index of operator ↑ is changed and counter t is not increased.

Figure 7 shows an algorithm logical schema built with the help of adjacency matrix (Figure 6). The derived algorithm is then loaded into a software module, which is then used to administrate the business game.

$$H_1\ A_0\ P_1{\uparrow}^1\ P_2{\uparrow}^2\ P_3{\uparrow}^3\ \omega{\uparrow}^4\ {\downarrow}^2\ A_1\ P_4{\uparrow}^1\ P_5{\uparrow}^5\ P_6{\uparrow}^6\ \omega{\uparrow}^7\ {\downarrow}^3\ A_2\ P_7{\uparrow}^1\ P_8{\uparrow}^7\ P_9{\uparrow}^8\ \omega{\uparrow}^9\ {\downarrow}^4\ A_3\ P_{10}{\uparrow}^1\ P_{11}{\uparrow}^{10}\ \omega{\uparrow}^{11}\ {\downarrow}^5\ A_4\ P^{12}{\uparrow}^1\ \omega{\uparrow}^7\ {\downarrow}^6\ A_5\ \omega{\uparrow}^1{\downarrow}^8\ A_6\ P_{13}{\uparrow}^1\ \omega{\uparrow}^7\ {\downarrow}^9\ A_7\ \omega{\uparrow}^1\ {\downarrow}^{11}\ A_8\ \omega{\uparrow}^1\ {\downarrow}^{10}\ A_9\ \omega{\uparrow}^1\ {\downarrow}^7\ A_{0.0}\ P_{14}{\uparrow}^1\ P_{15}{\uparrow}^{12}\ \omega{\uparrow}^{13}\ {\downarrow}^{13}\ A_{0.1}\ \omega{\uparrow}^{14}\ {\downarrow}^{12}\ A_{0.2}\ \omega{\uparrow}^{14}\ {\downarrow}^{14}\ A_{0.k}\ P_{16}{\uparrow}^1\ P_{17}{\uparrow}^9\ \omega{\uparrow}^6\ {\downarrow}^1\ K_1$$

Figure 7. Algorithm logical scheme

Thus, the step-by-step transition from UBP to business game automate model has been considered.

Conclusion

One of the problems faced by business game developers is the problem of complicated domain formalization. It is needed to derive a range of models to get from a real business process of an organization to a set of business game resources. Business game control is performed with the help of automate module, which uses automate model of the business game in form of algorithm logical schema as input data. The automate model is acquired as a result of gradual transition from unified business process model to unified training business process, which

serves as a basis for business game scenario graph. The construction of automate model is finished by building the algorithm logical schema derived from analysis of vertex-adjacency matrix that corresponds to the scenario graph. This paper proposes construction algorithms of unified business process, scenario graph and algorithm logical scheme.

Bibliography

[Deryabin, 2013] Deryabin A.I., Shestakova L.V., Vikentyeva O.L.. The Construction of competency-based business game operational model // International Journal "Information Technologies & Knowledge". 2013. Vol. 7. No. 4. P. 303-313.

[Mazanek, 2008] Mazanek S. Visual Languages. MetaEdit+ : [Электронный документ] (http://visual-languages.blogspot.com/2007/11/metaedit.html).

[Wells, 1990] R. A. Wells. Management Games and Simulations in Management Development: An Introduction // Journal of Management Development. 1990. Vol. 9, Issue 2. P.4-6.

[Zichermann, 2011] Zichermann G. and Cunningham C. Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps [Book]. - Sebastopol, California : O'Reilly Media, 2011.

[Викентьева, 2014] Викентьева, О.Л. Формализация предметной области при проектировании деловой игры] / О.Л. Викентьева, А.И. Дерябин, Л.В. Шестакова // Информатизация и связь. – № 1, 2014. – С. 58-61.

[Викентьева, 2013] Викентьева, О.Л. Концепция студии компетентностных деловых игр [Электронный ресур] / О.Л. Викентьева, А.И. Дерябин, Л.В. Шестакова // Современные проблемы науки и образования. – 2013. – № 2; URL: http://www.science-education.ru/108-8746 (дата обращения: 03.04.2013).

[Сухов, 2009] Сухов А.О. Среда разработки визуальных предметно-ориентированных языков моделирования // Математика программных систем: Межвуз. сб. научн. тр. / Перм. ун-т. Пермь, 2008. С. 84-94.

Authors' Information

Alexander Deryabin *– National Research University Higher School of Economics, City of Perm, Perm, Russia, e-mail: paid2@yandex.ru*

Major Fields of Scientific Research: General theoretical information research, Multi-dimesional information systems

Lidia Shestakova *– National Research University Higher School of Economics, City of Perm, Perm, Russia, e-mail: L.V.Shestakova@gmail.com*

Major Fields of Scientific Research: General theoretical information research, Multi-dimensional information systems

Olga Vikentyeva *– National Research University Higher School of Economics, City of Perm, Perm, Russia, e-mail: oleovic@rambler.ru*

Major Fields of Scientific Research: General theoretical information research, Multi-dimensional information systems

THREE RD MODELS FOR TWO-HOP RELAY ROUTING WITH LIMITED PACKETS LIFETIME IN AD HOC NETWORKS

Irma Aslanishvili

Abstract*: The rules for preparing the manuscripts for the International Journals (IJ) and International Book Series (IBS) of the ITHEA International Scientific Society (ITHEA ISS) are outlined. We study mobile communication of networks, the ad hoc networks, Ad hoc networks are complex distributed systems that consist of wireless mobile or static nodes that can freely and dynamically self-organize. The parameter of the queuing models depends on the node mobility pattern.*

The main finding is that the expected relay buffer size depends on the expectation and the variance of the nodes contact time. Such analysis is done for the three dimensional random walks models over a circle; expected relay buffer size depends on the expectation and the variance of the nodes contact time.

First model-The source node transmits a packet only once (either to the relay or to the destination node). Thus, the source node does not keep a copy of the packet once it has been sent. When the source node transmits a packet to the destination node (when their locations permit such a transmission), the source node transmits packets that it has not transmitted before. The source node has always data to send to the destination node. This is a standard assumption, also made in [GMPS04, GT02, GK00], because we are interested in the maximum relay throughput of the relay node. This shows: the first relay node performs a Random walk and the source and destination are fixed, second the source, the destination, and the relay node move inside a square according to the RD model. Second model-The relay node is moving according to a symmetric random walk (RW) on a circle of 4R + 2w steps.

Third model - Three nodes: a source a destination and a relay source, nodes are moving a cording a symmetrical Random Walk over a circle.

Keywords*: Ad Hoc Networks, MANETs protocols, Routing protocols, packet, source node, Relay routing, finite memory, Relay Buffer (RB), RB occupancy, Destination.*

Introduction

We consider the Routing protocols in Ad Hoc Networks. The network consists of three types of nodes, source, destination, and relay nodes. The objective is to study the behavior of the relay buffer as a function of the nodes mobility models. We find the expected Relay Buffer size, in the heavy traffic case, embedded at certain instants of time. This expectation is called the event average. Note that the expected Relay Buffer size in the heavy traffic case serves also as an upper bound of the expected Relay Buffer size. Further, we show numerically that under the mobility models considered the event average converges toward the time average of the RB as the load of the relay buffer tends to one. This will be done for three different mobility models: Random Walk, Random Direction, and Random Way point.

Routing Protocols in Ad Hoc Networks

We have to note that in Ad hoc networks each node acts as a router for other nodes. The traditional link-state and distance-vector algorithms do not scale well in large MANETs. This is because periodic or frequent route updates in large networks may consume a sign if- cant part of the available bandwidth, increase channel contention and require each node to frequently recharge its power supply. To overcome the problems associated with the link-state and distance-vector algo-rhythm's a number of routing protocols has been proposed for MANETs. These protocols can be classified into three different groups: proactive, reactive and hybrid. In proactive routing protocols, the routes to all destinations are determined at the start up, and maintained by means of periodic route update process. In reactive protocols, routes are determined when they are required by the source using a route discovery process. Hybrid routing protocols combine the basic properties of the first two classes of protocols in to one. In proactive routing protocols, each node maintains information on routes to every other node in the network. The routing information is usually kept in different tables. These ta- bless are periodically updated if the network topology changes. The difference between these protocols lies in the way the routing information is updated and in the type of information kept at each routing table.

Reactive or on-demand routing protocols have been designed to reduce the overhead in proactive protocols. The overhead reduction is accomplished by establishing routes on- demand, and by maintaining only active routes. Route discovery usually takes place by flooding a route request packet through the network. When a node with a route to the destination is reached, a route reply packet is sent back to the source. Representative reactive routing protocols are: Dynamic Source Routing , Ad hoc On Demand Distance Vector, Temporally Ordered Routing Algorithm, Associativity Based routing, Signal Stability Routing [2].

Single source, destination, and relay nodes

The state of the relay node at time t is represented by the random variable $S_t \in \{-1, 0, 1\}$ where: S_t S_t = 1 if at time t the relay node is neighbor of the source, and if the destination is a neighbor neither of the source nor of the relay node In other words, when S_t = 1, the source node sends relay packets to the relay node at time t; S_t = −1 if at time t the relay node is a neighbor of the destination, and if the source is a neighbor neither of the destination nor of the relay node. When S_t = −1 the relay node delivers relay packets (if any) to the destination. 〖-S〗 _t = 0 otherwise. Mobiles have finite speeds. Assume that the relay node may only enter state 1 (resp. −1) from state 0: if S_t = S_t then necessarily S_t = 0, S_t = 1 or = −1.

The source node transmits a packet only once (either to the relay or to the destination node). Thus, the source node does not keep a copy of the packet once it has been sent. When the source node transmits a packet to the destination node (when their locations permit such a transmission), the source node transmits packets that it has not transmitted before.

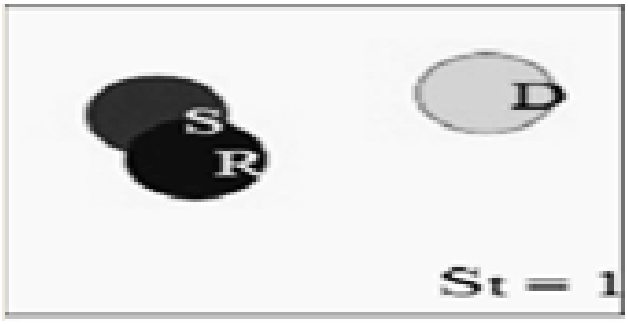


Figure 1.1.

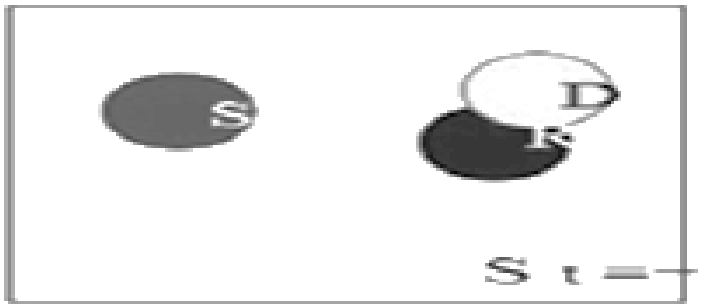


Figure 1.2.

Figure 1.1. and Figure 1.2. are the process $\{S_t\}$.

This is because when S_t= 1, the relay node receives data to be relayed from the source node at rate r_s; − it decreases at rate r_d if S_t = −1 and if the RB is non-empty. This is because if S_t= −1, and if there is any data to be relayed, then the relay node sends data to the destination node at rater_d. Let $\{Z_n\}_n$(Z_1<Z_2< • • •) (2) denote the consecutive jump times of the process { S_t= , t ≥0}. An instance of the evolution of S_t and B_t as a function of t is displayed in Figure 1.2. The evolution of the discrete indexed process $\{S_{zkn}$, k ≥ 1} consists of sequences of 1, 0 and −1. Without loss of generality assume that the process $\{S_t$, t ≥ 0} isaright-continuousprocess [6].

Packet Round Trip Time We assume that the source is ready to transmit the packet to the destination at time t = 0. The delivery time T_α, is the time after t = 0 when the destination node receives the packet.

Denote by B_t the RB occupancy at time t. The $r_v B_t$ evolves as follows: it increases at rate r_s if S_{-t} = 1. This is because when S_t = 1, the relay node receives data to be relayed from the source node at rate r_s; − it decreases at rate r_d if S_{-t}= −1 and if the RB is non-empty. This is because if S_{-t} = −1, and if there is any data to be relayed, then the relay node sends data to the destination node at rate r_d.

Assume that the relay node may only enter state 1 (resp. −1) from state 0: if S_{-t} = S_t then necessarilyS_t = 0 if S_{t-} = 1 or S_{t-} = −1. Let B_t be the RB occupancy at time t. Based on the definition of S_t, B_t increases at rate r_s if S_{t-} B_t = 1, decreases at rate if S_t = −1 and if the RB is non-empty, and B_t remains unchanged in all other cases.

Let $\{Z_n\}_n$ (Z_1<Z_2< • • •) (2) denote the consecutive jump times of the process { S_t= , t ≥0}. Define a cycle as the interval of time that starts at t, $S_t = Z_k$, for some k with S_t = 1, and (necessary) $S_{(t-)}$ = 0 and S_{zk-2} = −1, and ends at the smallest time t + τ such that $S_{(t+\tau)}$ = 1 and $S_{(s+\tau)}$ = −1 for some s < τ . Let $S_{(t+\tau)}$ denotes the duration of the n^{th} cycle. Let W_n denote time at which the n^{th} cycle begins. Let

$$\sigma_n \triangleq \int_{t=W_n}^{W_n} 1_{\{S_t\ =1\}}\, dt \tag{1}$$

be the amount of time spent by the relay node in state 1 during the n^{th} cycle. Similarly, let

$$\sigma_n \triangleq \int_{t=W_n}^{W_n} 1_{\{S_t\ =-1\}}\, dt \tag{2}$$

be the amount of time spent by the relay node in state −1 during the n^{th} cycle. Let B_n be the RB occupancy at the beginning of the n^{th} cycle, i.e. $\widetilde{B_n}$ = B_{W_n}. Clearly,

$$\widetilde{B_{n+1}} = \left[\widetilde{B_n} + r_s\sigma_n - r_d\ \sigma_n\right] \tag{3}$$

where [x]+ = max(x, 0). In other words B_{n+1} can be interpreted as the workload seen by the $(n+1)$ stcustomer, and $rs\sigma n$ is the service requirement of the nth customer [2].

Relay buffer behavior

The impact of the first mobility model on the relay buffer occupancy is studied. Assume that the mobility models under consideration have stationary node location distributions. The plan is to view this system as a GI/G/1 queue in heavy traffic and then to look at the effect of mobility patterns on the relay buffer occupancy. It is known from heavy-traffic analysis that the tail behavior (the large diva- tin exponent) of the buffer occupancy is determined by the variance of the service and inter-arrival times. Moreover, it is also to be understood that the effective arrival process to the RB in the second model, i.e.

$$\int_{u=Z_n}^{Z_{n+1}} 1_{\{S_{(u)=1}\}}\,du \tag{4}$$

$$\sigma_n = \int_{u=W_n}^{W_{n+1}} 1_{\{S_{(u)=1}\}}\,du \tag{5}$$

Clearly, α larger relay buffer occupancy would imply that the amount of time required to deliver all the packets would be composed of many contact periods between the relay node and the destination, hence there can be several inter-visits between the relay node and the destination required to deliver the packets. This implies that we cannot study the delay incurred by the nodes by considering only one inter-visit time (or the meeting time) or only one contact time. This shows that the buffer behavior (hence the delays) will depend on both contact times and the inter-visit times.

First model-The source node transmits a packet only once (either to the relay or to the destination node). Thus, the source node does not keep a copy of the packet once it has been sent. When the source node transmits a packet to the destination node (when their locations permit such a transmission), the source node transmits packets that it has not transmitted before. The source node has always data to send to the destination node. This is α standard assumption, also made in [GMPS04, GT02, GK00], because we are interested in the maximum relay throughput of the relay node. This shows: the first the relay node performs a Random walk and the source and destination are fixed, second the source, the destination, and the relay node move inside a square according to the RD model. The relay node performs a Random walk and the source and destination are fixed, second the source, the destination, and the relay node move inside a square according to the RD model.

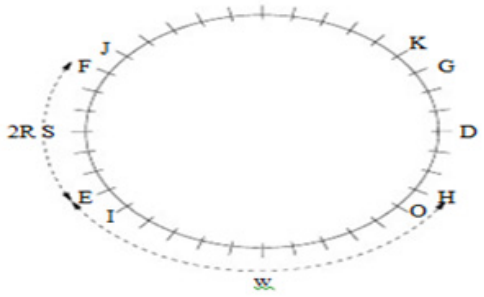


Figure 1.3.

Second model-The relay node is moving according to α symmetric random walk (RW) on a circle of 4R + 2w steps (Figure 1.3). The RW step size is fixed and is equal to μ meters. The speed of the relay node is assumed to be constant and equal to V, so the time required to jump from one step to the next one, is equal to μ/V seconds.

The source and the destination are fixed (not in movement), and they are located as shown in Figure 1.3. The quantities w and R are assumed to be integers. Also, the data transmission between source and destination only takes place through the relay node. When the relay node becomes a neighbor of the source (when passing points E or F), it starts to accumulate data at rate r_s. When the relay node enters the neighborhood of the destination, via points G or H, it delivers the data to the destination at rate r_d. Once in the interval [E, F], the relay node remains there for a random amount of time before exiting via points E or F. Symmetry implies that this time has the same distribution whether the relay node enters [E, F] through the point E or F. Similar is the case for the segment[G, H]. We call this (random) time the contact time between the relay node and the source (or the destination). Once the relay node exits [E, F], it either enters [J, K] or [I, O]. Now, the relay node stays in this region for a random amount of time (during which it neither receives nor transmits), and then either reenters [E, F] or enters [G, H] [1].

The number of times that the relay node enters [E, F] without entering [G, H] is denoted by the r_v L, and is geometrically distributed with parameter p, independent of whether the relay node exited [E, F] via E or F, that is

$$P(L = k) = (1 - p)p^{(1-k)}) \quad (6)$$

The parameter p is the probability that a symmetric random walker starting at point J hits point F before reaching G.

Let A_j, j ≥ 1, be independent and identically distributed random variables represent- ing the first time that a random walker, starting at point F, exits [E, F]. [R is the transmission range of source, destination, and relay node. Hence, the service requirement of a customer in the G/G/1 queue of Section (5) is σ, where

$$\sigma = \sum_{j=1}^{L} A_j \quad (7)$$

In the following, A denotes a generic $r_v$with the same distribution as A_j.

We find that

$$E[A] = 2R\frac{\mu}{V} \quad (8) \qquad \text{Var}[A] = (\frac{\mu}{V})^2.\frac{4R}{3}(2R + 1)(R + 1) \quad (9)$$

$$E[L]=\omega \quad (10) \qquad \text{Var}[L] = \omega(\omega - 1) \quad (11)$$

Since L is independent of A, we get

$$E[\sigma]=E[A]E[L] = 2R\omega\frac{\mu}{V} \quad (12)$$

$$\text{Var}[\sigma]=\text{Var}[A]E[L] + (E[A])^2\text{Var}[L] = 4R\omega(\frac{\mu}{V})^2.(\omega R + \frac{1}{3}(2R^2 + 1) \quad (13)$$

And now we find that when $r_s \approx r_d$ with $r_s < r_d$ and $r_s E[\sigma_n] \approx r_d E[\alpha_n]$ our conclusion is that the stationary waiting time is exponentially distributed with

$$E[\widetilde{B_n}] \approx \frac{r_d^2 \text{Var}(a_n) + r_s^2 \text{Var}(\sigma_n)}{2(r_d E[a_n] - r_s E[\sigma_n]} \quad (14)$$

Where we have used the fact that σ_n and σ_d are identically distributed. Now we find that the expected relay buffer size depends on the expectation and the variance of the nodes contact time. Such analysis is done for the one dimensional random walk over a circle. There is second model Two- hop rout between two nodes s and d [4].

Figure 1.4.

Observe that P^N is a function of $(\frac{R}{N})^2$. Note that u (i) independent of the mobility model of the nodes, and that in the case RD mobility [5].

Consider scenario of a third model. Three nodes: a source a destination and a relay source. Nodes are moving a cording α symmetrical Random Walk over a circle. It follows from p = r_s/ r_d<1. Figure 1.4. plots the evolution of relay node buffer with time for different values p. It is evident when p=1,0. The buffer occupancies process is unstable.

Figure 1.4. Time-evolution of relay node buffer for random Walk is third model over a circle for different values of ration $p = r_s/r_d$.

Conclusion

This exemplar is meant to be a model for manuscript format. Please make your manuscript look as much like this exemplar as possible. The behavior of the relay buffer of the two-hop relay routing in mobile ad hoc networks we tolled in these three RD models. The parameters of the queuing models depend on the node mobility pattern.

The main finding are in these three models the expected relay buffer size depends on the expectation and the variance of the nodes contact time. The source node transmits a packet only once (either to the relay or to the destination node). Thus, the source node does not keep a copy of the packet once it has been sent. When the source node transmits a packet to the destination node (when their locations permit such a transmission), the source node transmits packets that it has not transmitted before.

The source node has always data to send to the destination node. This is a standard assumption, also made in [GMPS04, GT02, GK00], because we are interested in the maximum relay throughput of the relay node.

When the destination node comes within the transmission range of the relay node, and if the destination and the relay node are outside transmission range of the source node, then the relay node sends the relay packets (if any packets in its RB) to the destination node at a constant rate d.

Such analysis is done for the one dimensional random walk over a circle. Relay Routing models are like models of M .L .Tsetlin who supposed that the elementary behavioral models can be singled out from the complex behavior and elementary problem can be formulated, any complex behavior based on a finite storage space..

Bibliography

[1] Oleg Namicheishvili, Hamlet Meladze, IrmaAslanishvili, Transactions. Two models for two-hop relay Routing with limited Packet Lifetime -Georgian Technical University. AUTOMATED CONTROL SYSTEMS -54-58 No 1(10), 2011

[2] R. Groenevelt, P. Nain and G. Koole, The Message Delay in Mobile Ad Hoc Networks, Proc. of Performance 2005, Juan-les-Pins, France, October 2005. Published in Performance Evaluation, Vol. 62, Issues 1-4, October 2005, pp. 210-228.

[3] M. Grossglauser and D. Tse, Mobility Increases the Capacity of Ad hoc Wireless Networks, IEEE/ACM Transactions on Networking, Vol. 10, No. 4, August, 2002, pp. 477-486.

[4] E. Zhang and G. Neglia and J. Kurose and D.Towsley. Performance Modeling of Epidemic Routing, Umass Computer Science Technical Report 2005-44.

[5] A. Al Hanbali, P. Nain and E. Altman, Performance Evaluation of Packet Relaying in Ad Hoc Networks. INRIA Research Report RR-5860, Mars2006.

[6] Irma Aslanishvili, One model for two-hop relay Routing with limited Packet Lifetime The Conference for International Synergy in Energy, Environment, Tourism and contribution of Information Technology in Science, Economy, Society and Education era-7.ISSN 1791-1133 www.erateipier.gr

Authors' Information

Aslanishvili Irma *– I.Javakhishvili Tbilisi State University - Faculty of Exact and Natural Sciences- teacher of informatics and computer sensor networks; T: +995 555340031*

E-mail :Irma.aslanishvili@tsu.ge

Scientific Research: General theoretical information research, information systems and computer sensor networks

COMPUTERIZED MODEL FOR APHASIA ASSESSMENT, BASED ON KERTEZS TEST

Iliya Pendzhurov, Polina Mihova, Atanas Dashovski

***Abstract**: Aphasia is an acquired communication disorder caused by brain damage that impairs a person's ability to understand, produce and use language. Furthermore, it is a multimodal disorder which includes additional neurological impairments such as apraxia of speech or dysarthria [Keshree et al, 2013]. Aphasia tests are used to classify and diagnose aphasia and they are needed for a standard classification of aphasia and to compare studies performed in different centers. The aim of this paper is to present an original authors solution, which performs a digital version of the Kertész test.*

***Keywords**: aphasia, WAB, software solution*

Introduction

The Western Aphasia Battery is a comprehensive test of language function for individuals with aphasia and aged 18-89 years. Test administration time is 30-60 min, depending on the severity of the patient's aphasia and coexisting deficits (e.g., apraxia, dysarthria). As stated in the test manual, the aim of the WAB is to "evaluate the main clinical aspects of language function, content, fluency, auditory comprehension, repetition, naming, reading, writing, and calculations." The WAB is designed to test all language modalities: reading, writing, listening, speaking, and gestural communication. It also tests what Shewan and Kertesz originally referred to as "higher cortical functioning", including a block design subtest as well as Raven's Colored Progressive Matrices and tests of drawing, and calculation. The purpose of this paper is to present a new kind of solution - computer assisted testing for aphasia, which is quite new and unique for Bulgarian medical practice.

Material and methods

Aphasia test batteries have been used by clinicians to assess persons with aphasia (PWA) for nearly a century. As interest in aphasia rehabilitation grew between 1960 and 1982, and objective measures were needed to measure the effects of treatment, several test batteries were developed which have a widespread use globally. Among them are the Minnesota Test for Differential Diagnosis of Aphasia, the Porch Index of Communicative Ability, the Boston Diagnostic Aphasia Examination, and the Western Aphasia Battery. The Western Aphasia Battery provides the diagnostic goals of Minnesota Test for Differential Diagnosis of Aphasia. [Keshree et al, 2013]

The test of Kertész (WAB) is a tool for diagnosing subtypes aphasia, extent of damage aphasia and evaluating the results of treatment. The test is a comprehensive survey, providing an opportunity to involve highly structured observations in the process of diagnosis.

The test consists of two parts (language and executive) containing several subtests. The language part of the test gives coefficient of Aphasia (CA), covering as necessary for the classification of types of Aphasia by several sub tests. The test itself does not classify cases cross aphasia symptoms.

The second part contains tests for reading and writing (Display language factor), Praxis, drawing, stacking cubes, calculus and colored progressive matrices Raven, by which, in combination with results of the first part is derived factor of cognitive functioning KKF (aphasiology.pitt.edu/archive/00000795/01/14-07.pdf).

In frequent cases in clinical practice are those with mixed symptomatology. KKF helps to determine whether (and to what extent) the disturbances in other higher cortical functions, which in turn strongly supports the diagnosis in the presence of cross symptoms and an individual treatment plan for the patient.

On the other hand, information technologies – part of the eHealth era, are becoming more and more widely used in health care. The digitization of “medical paper flow" reduces the needed for storage resources and gives a positive impact to the efficiency of the entire workflow. After a research, we found out that at the Bulgarian market absents specialized solutions for speech therapy with integrated form for administration of the Kertész test.

According to the National Health Development Strategy 2008-2013 the following are outlined as key priorities:

- ✓ Providing health services on-line;
- ✓ Implementation of electronic health cards.
- ✓ Implementation of personal electronic health records.
- ✓ Implementation of integrated software applications for processing and sharing information in real time, including: electronic directions, recipes, expert findings, laboratory and diagnostic data, etc.
- ✓ Development of complex integration models, working with external applications and systems.
- ✓ Creation of hospital information systems for electronic medical records.
- ✓ Construction of the infrastructure required for the normal functioning of the healthcare system - networks that connect devices and other devices.
- ✓ Construction of appropriate infrastructure for the deployment of telemedicine applications.

eHealth, by definition, is a rapidly developing field where medical informatics, public health, supply of healthcare services and information using modern information and communication technologies interact. It features technology development to improve health services at local, regional and global levels.

The standard medical practice, as such, exists from the Hippocrates times - face to face contact with a suffering patient, personal experience to suggest accurate treatment and legitimacy to the laws in Bulgaria are just some of the advantages and established standards of practice. On the other hand, the paper documentation that is still the practice in our healthcare system, the possibility of intentional or accidental error, lack of sufficient practical experience in this particular case, show these are just some of the factors that present the current status of our system as old-fashioned, unsatisfactory and risky for the health of the patient.

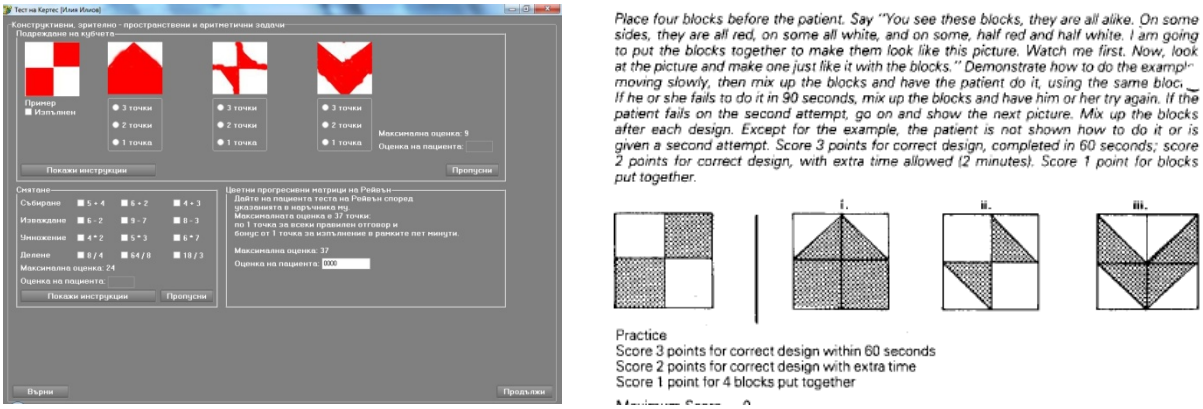
However, medical practice, based and ensured with PC and software technologies is facing many other problems that lead to fear and reluctance from consumption of new kinds of working models. It requires modern standards of employment, security for the patient by several expert opinions, prevention and minimization of the possibility of errors in the final diagnosis, the provision of care 7 days a week, 24 hours a day. Of course, the difficulties and obstacles in this scheme of work related to the willingness of medical experts, placed in a competitive regime, the lack of ethical and legal frameworks that limit abuses and skills required to work with information technology are only a few of the barriers to implement that service. At Table 1 we demonstrate the positive and negative factors for both standard medical practice and medical practice, assisted by software and hardware solutions.

Table 1 Comparative analysis between standard medical practice and assisted one by PC and software solutions

Standard medical practice		Medical practice assisted from PC and softwares	
For	**Against**	**For**	**Against**
tradition of medical practice		simultaneous operation of	costs for equipment
changing with the generations		several organizations	
tested in time – fixed and accepted as routine practice		simplified and standardized process	knowledge of electronic standards and requirements
paper does not require any technology	risk of information loss	improved health services	new public nature of the
			profession
face-to-face contact	expenditure of financial and time resources to	provision of care anytime,	problems with persuasiveness
	remote patients and	anywhere and by anyone	and reliability
	their families		
subjective – the record is done by one author	possible errors	transfer of various data formats	Perceptions of staff to
	or omission		work with the new system
Law legitimacy	opportunity for abuse and manipulation of information	various forms of diagnostic techniques	competency requirements for more than narrow specialization
written responsibility, verified with personal signature	delay in time	consultation with multiple	competitive moment
		specialists together	
results of communication	poor results of aging assets	Provision of care 24 hours at home	ongoing commitment
with patients			
		education from a distance	Pricing - Who decides?
		in real time	
		human interaction - PC	PC equipment
		objectivity of opinions	ethical issues
		reduce professional isolation	Lack of political will to implement
		increased confidence by	Institutional will
		specialist	
		providing the best experts	prudence
		in the field	
		new working standards ,	lack of legal framework
		fast and efficient , information transfer	
		Real teamwork	

The presented here electronic health record is a client -server application and it requires connection to a MySQL Server database.In practice, the essential function of any system is to provide the user a convenient way to manage records in a database table .

The screens below (Figure 1 – Figure 2) demonstrate the digital representation of Kertesz paper`s version, which is over 20 pages in non-digital format.



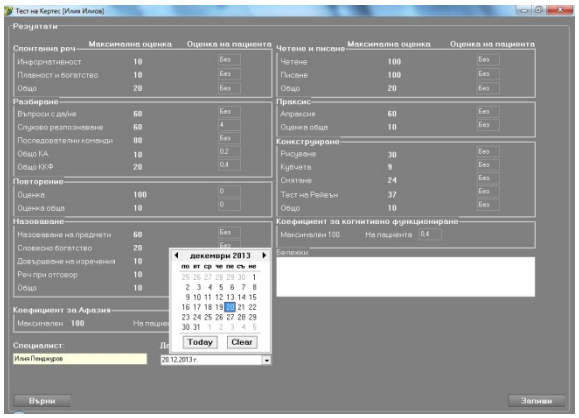
Place four blocks before the patient. Say "You see these blocks, they are all alike. On some sides, they are all red, on some all white, and on some, half red and half white. I am going to put the blocks together to make them look like this picture. Watch me first. Now, look at the picture and make one just like it with the blocks." Demonstrate how to do the exampl moving slowly, then mix up the blocks and have the patient do it, using the same bloc If he or she fails to do it in 90 seconds, mix up the blocks and have him or her try again. If the patient fails on the second attempt, go on and show the next picture. Mix up the blocks after each design. Except for the example, the patient is not shown how to do it or is given a second attempt. Score 3 points for correct design, completed in 60 seconds; score 2 points for correct design, with extra time allowed (2 minutes). Score 1 point for blocks put together.

i. ii. iii.

Practice
Score 3 points for correct design within 60 seconds
Score 2 points for correct design with extra time
Score 1 point for 4 blocks put together

Figure 1. Screen "Moving blocks" from Kertesz test computerized model vs paper version

On Figure1 is presented a comparison between digital and paper version of Nonlinguistic Skills assessment, which are Drawing, Block Designand CalculationPraxis.



	Maximum	*Patient's Subscores*	*Total For AQ*
Spontaneous Speech			
Information Content	10		
Fluency	10		
Total	20		
Comprehension			
Yes/No Questions	60		
Auditory Word Recognition	60		
Sequential Commands	80		
Total (Divide By 20 For AQ) (Divide By 10 For CQ)	10 / 20		
Repetition	100		
Total (Divide By 10)	10		
Naming			
Object Naming	60		
Word Fluency	20		
Sentence Completion	10		
Responsive Speech	10		
Total (Divide By 10)	10		
Aphasia Quotient (Add Totals And Multiply By 2 For AQ)			
Reading And Writing			
Reading	100		
Writing	100		
Total (Divide By 10)	20		
Praxis	60		
Total (Divide By 6)	10		

Figure 2. Screen Score sheet from Kertesz test computerized model vs paper model

On Figure 2 is demonstrated a comparison between digital and paper version of Spontanious speech assessment, which are the Linguistic Skills most frequently Affected by Aphasia. It consists of the following parts: Content, Fluency, Auditory, Comprehension, Repetition and Naming, Reading, Writing.

Discussion

Computer technologies offer new avenues for treatment and expression for people with aphasia. Based on web browsing and searching for similar software solutions, we have found 55 products (http://www.aphasiasoftwarefinder.org/spelling%20and%20writting?tid[]=31), containing different separate parts from the general Kertezs test, which would speech-language pathologists can use with clients with aphasia. They are compared by the following parameters: Results Recorded, Free Trial Tutorials, Personalized, Advice, Record Self, Easy To Use and Number of Exercises.

Also, we have found two other desktop solutions (PsychBook CMS and Therapy Notes), which are compared below.

The current project is designed to be easily integrated in our specialized Center for treatment of communication and emotionally-behavioral disorders in childhood, New Bulgarian University. It is first and at the moment the only electronic solution for Kertezs test for Bulgaria.

Organization and requirements of innovatory unit have to be secured and nucleus, namely "Patient Management", which provides complete patient record, ability to search, elaborate, investigate and archive records and data.

The proposed solution provides collected and processed information in a knowledge base that guarantees its customers the necessary statistical information, as well as copyright forms and methods of searching and numerous data references.

According to the task, they represent authors' model in identical to the paper copy, which operates daily in the laboratory. PsychBook CMS is an appropriate solution for a single practice with easy to use interface. Major business activities are Invoicing, Financial Statements, Planning graphic. Besides the administrative patients information in PsychBook CMS is integrated full patient's history, presented in usable form. The solution also performs the ability to prepare treatment plan.

PsychBook CMS is available in both versions for Macintosh and Windows operating systems. It uses local database, therefore it is not required internet connection.

Therapy Notes - Web based platform suitable for medium-sized venues. Support various positions with limited access makes it easy adaptable to the structure of the organization. There are well-integrated templates for guiding therapeutic documentation.

Function	*PsychBook CMS*	*Therapy Notes*	*Our project*
Multiuser	No	Yes	Yes
Export	No	Yes	Yes
Attachment of files	Yes	Yes	Yes
Schedule	Yes	Yes	No
Treatment plan	Yes	Yes	Yes
Investigation tools	No	No	No
Reports	Yes	Yes	No

Figure 3. Comparative analysis of Kertez test computerized models

Conclusion

Aphasia following an acquired neurological insult necessitates an in-depth evaluation of the primary and secondary language symptoms. Of all the tools available for aphasia diagnosis, the Western Aphasia Battery [Keshree et al, 2013] has proved to be one of the most comprehensive test batteries for describing the aphasia symptom complex. Several authors have pointed out the need for language-specific tools for the assessment of aphasia. Based on the presented analysis and the surrounding era of "e-" and mSolutions in medicine and healthcare, we can conclude that the author`s solution is necessary tool for Bulgarian medical practice.

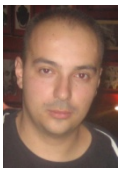
Bibliography

[Keshree et al, 2013] Keshree N. K., S. Kumar, S. Basu, M. Chakrabarty, T. Kishore, A. Y. Jung "Adaptation of the western aphasia battery in bangle", Psychology of Language and Communication, 2013, Vol. 17, No. 23.

[Leonard L. 2004] Leonard LaPointe, Aphasia and Related Neurogenic Language Disorders, ISBN 9781588902269, Thieme Medical Pub, 2004, http://www.aphasiasoftwarefinder.org/spelling%20and%20writting?tid[]=31 , last visited 20.03.2014

[Risser A.H., 2003] Assessment of aphasia Author: Otfried Spreen, Publisher: Oxford; New York: Oxford University Press, 2003, aphasiology.pitt.edu/archive/00000795/01/14-07.pdf, last visited 20.03.2014

Authors' Information



Iliya Pendzhurov *– PhD, lecturer at NBU, Montevideo str. N21, Sofia-1618, Bulgaria; e-mail: penjurov@nbu.bg*

Major Fields of Scientific Research: Data bases, Developmental of information systems in medicine and healthcare

Polina Mihova *– PhD, lecturer at NBU , Montevideo str. N21, Sofia-1618, Bulgaria; e-mail: pmihova@nbu.bg*

Major Fields of Scientific Research: Telemedicine, Integration of information systems in medicine and healthcare

Atanas Dashovski *– bachelors degree student at NBU, Montevideo str. N21, Sofia-1618, Bulgaria; e-mail: nasso_svt@abv.bg*

ORGANIZING OF KNOWLEDGE BASES FOR EMERGENCY SITUATIONS USING SOCIAL NETWORKS

Oleksandr Kuzomin, Bohdan Tkachenko, Mariia Lysenko

***Abstract**: This paper describes technology of automated survey system for emergencies eyewitness that helps to collect data about emergencies.*

***Keywords:** Social Networks, Emergencies Prevention, Automatic Survey, Language Analysis, Special Language.*

Introduction

Procuring the security of human life, which is exposed to the risk of the natural disasters or environment changes, which may cause catastrophes (such as floods, flooding coastal zone of the Atlantic ocean, landslides, avalanches, fires, etc.) requires improving quality and amount of the information provision in Situation Centers [Kuzemin, 2004a].

At the moment there are many various services, which have their own databases with the information about known emergencies. But even specialized services not always have full and almost accurate information. That's why there is need in additional sources or ways to get new of more complete information about emergencies.

These problems can be solved with the help of systematic expansion of Databases through intelligent surveys, which automatically conduct a survey in social networks. Such surveys can collect information about emergencies that have already happened, about the assumptions that have caused them or immediately before emergency, and also about success of disaster management.

In the last years the fundamentally new civil society actors in the form of local structures - communities, or Online Communities began to play active role in prevention and elimination of the emergencies. Online Communities are relatively unstable group of people, interacting via the Internet-communications, provided by Internet services and having common interests and carrying out joint activities in virtual space [Parinov, 2000]. Habbermas thinks that the main characteristics of a network of civil society are "openness" as "establishment of broad, multi-dimensional relationships" and also "spontaneity" – "free formation, flow, and constant change" structures [McConnell, 2007].

Social networks are developed on a global scale. [Rimskiy, 2009] Social resources of online communities allow them to implement a number of functions by which the actualization of online communities as subjects of action in emergencies happens. According to a sociological study of the American Red Cross, dedicated to the role of social networks in information work in emergencies, 42% of respondents which received information from Internet, 6% of them already received information about emergencies from Twitter, 14%. In March 18 1999, American programmer Brad Fitzpatrick opened service of online blogging "Live Journal". Initially, project was planned as blog platform for communicating for friends of founder, but later grew into one of the most popular sites in the world. Today Live Journal is positioned as social network. In March 11, 2010 on site it is recorded more than 25 million registered users (in 2005 it was 8.9 millions, in June 12007 – 13 million).

For empowering situational centers [Kuzemin, 2004b] in predicting and disaster management, can play role above-mentioned social networks, which besides developing volunteer tasks for providing information support can be used to expand the database and knowledge about the processes in the emergency area. Herewith, is

possible to get a versatile amount of system knowledge, which, using that knowledge precedent will reduce the time to search for alternative solutions and the adoption of effective measures for disaster management.

Proposed solutions

Greater involvement of social networks in choosing effective control decisions can be obtained by organizing a wide valid survey with the help of Social Networks. This survey should include structured sequence of questions about solving problem.

In the result of survey in Social Networks can be identified qualitative parameters, which represent a set of alternative decisions or actions. Every decision that a decision-chooser person is offered (DCP) has three main characteristics: purpose of the action; a description of the action; a means of its implementation. All these characteristics can be obtained from structured responses from Social Networks which can answer the following three questions: "Why?" or "What should be the result?" - The answer will characterize the goal (purpose of action); "What should I do?" - Answer will give description of the action; "How to do?" - The answer will be to characterize the means for performing (the ability and opportunity).

Foundation for building a "decision tree" based on system issues to identify knowledge for actions aimed for eliminating or preventing disaster has the following basic short questions: "What?", "Why?", "How?". Answers to these questions will create a knowledge base of the situational center.

Developing situation-linguistic model of the environment in the emergency area is aimed for creating production-framing hierarchy. Architecture of production-framing hierarchy is based on a frame-based knowledge, which will be based on framing inheritance hierarchy with attitude and active slots, and the output is production rules. This approach allows a natural way to combine in one model the statistical knowledge about emergency values in the form of slots, structural domain knowledge in the form of an inheritance hierarchy.

Thus, the frame-based systems can be represented as:

$$W : S \to I,$$

where I is a multiplicity of frames, $S = \{S_i\}, i = \overline{1,n}$ — finite set of slots of type $\langle v, d, \{D_j\}\rangle$, including the current value of the slot $v = \langle v_1, v_2, \dots, v_l\rangle \in T$ and default value $d = \langle d_1, d_2, \dots, d_k\rangle \in T$ of procedures demons $\{D_j\}$.

Inheritance relationship: induced by slot with reserved name parent: $F: G \Leftrightarrow \|F\ (parent)\| = G$. Typical for frame-based systems operation of specification for frame like on the model, is implemented by implicit inclusion to the model the rule $F(parent) \longleftarrow match(F, G)$. When considering multiple inheritances, slot parent is supposed as list type, and $F: G \Leftrightarrow G \in \|F\ (parent)\|$.

Conclusion about the values of the output parameter (resulting frame - I_D) is performed on condition of clear input values (request frame - I_c, consisted of subset $v \times d$) has the following form:

$$L: \begin{Bmatrix} IF\ \langle v_1^1, \dots, v_1^1;\ d_1^1, \dots, d_k^1\rangle\ THEN\ \overset{P}{1} \\ IF\ \langle v_1^2, \dots, v_1^2;\ d_1^2, \dots, d_k^2\rangle\ THEN\ \overset{P}{2} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ IF\ \langle v_1^n, \dots, v_1^n;\ d_1^n, \dots, d_k^n\rangle\ THEN\ \overset{P}{n} \end{Bmatrix} \quad (1)$$

The resulting frame I_p represents aggregate of v and d, belonged to different frames. Thus I_p is output of the production system, which is presented as follows:

$$\langle S_1^p, S_2^p, \dots, S_n^p \rangle \to I^p.$$

To obtain this goal need to solve the following problems:

- Implementation of the construction of predicate queries and their modification, which will be the formally logical unit description and study the updating and modification of data and knowledge bases;
- Determine the rules of inference based on data and knowledge bases.

Under the concept of a database will mean a set of facts that we get from social networks as a result of the survey. The main ideas of this approach are considered under specific SQL implementations or implementations of Internet technologies.

Description of the Knowledge Base

All decisions in the subject area are taken on the basis of analysis of the conclusions of experts and specialists with work experience. Information system Knowledge Base examined according to [Slipchenko, 2004] as set of information entities of atomic predicates from some information space $\Re$. All changes that occur in the knowledge base are considered as a consequence of the modification of predicate queries Q_m. The basis of these predicate queries is a set of modification predicate rules:

$$Q_m \leftrightarrow (K_B) \ll \left\| \begin{matrix} K_{B-(X)} \\ K_{B+(X)} \end{matrix} \right. ,$$

where $X \in \Re$.

$K_{B+(X)}$ means, that atomic predicate $^{\circ}$ should be included in knowledge base K_B, K_B means, that X should be excluded from the knowledge base; $(K_B) \ll$ means modification of the knowledge base at the level of logical consistency of predicate rules; $K_{B\pm(X)}$ means the ability to modify not only the knowledge base, but also protection of the user based on the descriptors; $\ll$ considered as a complex arrow, features of are examined by the theory of categories.

Knowledge Extraction

Knowledge can be represented in the form of production rules of the type [3]:

$$\text{if } X_1 \& \dots \& X_K \; then \; X_{K+1} \& \dots \& X_{K+L} ,$$

where $X_1 \dots X_K$, $X_{K+1} \dots X_{K+L}$ - some predicates.

Definition 1. Content of knowledge if $X_1 \& \dots \& X_K \; then \; X_{K+1} \& \dots \& X_{K+L}$ is called set $W = \Pi_1 \times \Pi_2 \times \dots \times \Pi_{K+L}$. Arbitrary element of this set is an element of content knowledge.

Content of condition knowledge is called a set $W_1 = \Pi_1 \times \Pi_2 \times \dots \times \Pi_K$. Arbitrary element of this set is called an element which contains conditions of knowledge.

Content of consequence knowledge is called a set $W_2 = \Pi_{K+1} \times \Pi_{K+2} \times \dots \times \Pi_L$. Arbitrary element of this set is called an element which contains consequence of knowledge.

Definition 2. Under the probability p_i of element the knowledge content $w_i \in W$ we will mean the probability of event, consisting in the fact that all predicate constants up the w_i will take a logical value "AND" by substituting the value of objects instead of arguments from areas of truth predicate variables that make up this knowledge.

Element w_j which contains knowledge is a vector, components of which are the values of predicate variables, included into the knowledge. We can associate a vector $z_j = z_j(1), \dots, z_j(K+L)$ from R^{K+L} with an element w_j which contains knowledge.

The function of the distribution of knowledge is a function from $\mathrm{K+L}$ arguments: $F(y) = F(y(1), y(2), \dots, y(K+L))$ with domain of definition R^{K+L} and taking values in the space R^1. It is defined by the formula $F(y) = \sum_{z_j \le y} p_j$, where z_j displaying the element of the content of knowledge w_j in R^{K+L}. $z_j < y$ expression is understood as an act of: $z_j(i) < y(i), i = 1, \dots, K+L$.

Definition 3. As distance between comparable knowledge 3H1 and 3H2 we will call Hellinger distance $d(G, Q)$ between two probability distributions of their content elements $G = \{p_{11}, p_{12}, \dots, p_{1r}, \dots\}$ and $Q = \{p_{21}, p_{22}, \dots, p_{2r}, \dots\}$, which is calculated by the following formula:

$$d(G, Q) = \sum_j (\sqrt{p_{1j}} - \sqrt{p_{2j}})^2 \quad (2)$$

Calculating the distance between knowledge, can be solved the problem when the input of the knowledge base is supplied some new knowledge. Required to determine which of the subsets of the knowledge should include this new knowledge. To solve this problem we should calculate the distance between the new knowledge and all available knowledge, and then take it to a subset containing such knowledge, for which the distance is minimum.

Decision making

Despite the fact that decision-making is carried out in a selected subset of knowledge for complex systems and processes adequate mathematical description of the decision is absent or are rather cumbersome mathematical constructions which optimization and practical use in real-time is impossible. This problem can be solved, using algorithms, based on models, simulating the decision-making process by experienced expert [Kuzemin, 2005]. For a large number of models decision-making as the mathematical apparatus can be used the theory of fuzzy sets. When choosing solutions in situational centers the aim is to choose the design stages of design options or parameter value from a fairly small predetermined set determined as mentioned earlier, using the formula (2). For modeling decision-making process proposed to use decision-making models, based on fuzzy rule modus ponens, inductive output fuzzy and fuzzy expert information of the second kind. This will be used Inductive Output type [4]:

$$L': \begin{Bmatrix} IF\ B'_1\ THEN\ A'_1 \\ IF\ B'_2\ THEN\ A'_2 \\ \dots\dots\dots\dots \\ IF\ B'_m\ THEN\ A'_m \end{Bmatrix} \quad (3)$$

Clear-cut statements A' and B' have the form:

$$A': \langle \beta_w\ is\ w' \rangle;\ B': \langle \beta_v\ is\ v' \rangle;$$

$$w' = (x, y, z, \dots) \in X \times Y \times Z \times \dots, v' \in V$$

In this scheme of output statements about the values of the input parameters are sending to schema itself (statement A') and a consequence inside the system L' of the statement (statement A_j) and statements about the values of output parameters are the consequence of output circuits (3) (statement B'), but sending circuit inside L' (statement B'_j). Therefore, to select the output v parameter values, based on the rules modus ponens it should convert the output circuit (3) in the form:

$$\left.\begin{matrix} L' \\ A' - true \end{matrix}\right\} \to B' \; is \; true$$

For this purpose is proposed to convert system of the statements of the second type in the equivalent system of the first type, using the contra positive rule, according to which for arbitrary expressions A and B saying "$IF\ A\ THEN\ B$" and "$IF\ \bar{A}\ THEN\ \bar{B}$" are equivalent.

Applying the rule to the expressions $L'_j, j = 1, \dots, m$ contrapositive of the second type, we obtain

$$\langle IF\ A'_j\ THEN\ B'_j \rangle = \langle IF\ \overline{A'_j}\ THEN\ \overline{B'_j} \rangle$$

where statements $\overline{A'_j}$ and $\overline{B'_j}$ can be considered as statements of $\langle \beta_w\ is\ \alpha_k \rangle$ and $\langle \beta_v\ is\ \alpha_m \rangle$ in which the values $\alpha_{w_j^*}$ and $\alpha_{v_j^*}$ are determined by membership functions $\mu_{w_j^*}$ and $\mu_{v_j^*}$ which are in addition to $\mu_{w_j}$and μ_{v_j}:

$$\mu_{w_j^*}(w) = 1 - \mu_{w_j}(w), \forall w \in W = X \times Y \times Z \times \dots;\ \mu_{v_j^*}(v), \forall v \in V$$

Knowledge base construction

Emergency description $Sit' = \{Sit_i\}\ (i = \overline{1,n})$ consists from the set of micro situations − s_i, which are formed by concepts – environmental elements that are being presented by survey participants:

$$Sit' = \langle e_i, K_e, X, \Im, \Sigma \rangle$$

where Σ − required resources (volumes and methods for the prevention and elimination of emergency situations); $\Im$− a plurality of control actions; part of the situation that is defined by pair $\langle e_i, K_e \rangle$ is called linguistic (qualitative, semantic unit) micro situation of central concept – e, which is based around micro situations.

$X = \{x_j, (j = \overline{1,m})\}$ – quantitative indicators about natural phenomena (e.g., atmospheric pressure, temperature, etc.), constant parameters of the environment (for example, the slope of the mountain, covering the slope, etc.) in a controlled area and resources that respondent was able to identify (e.g., the number of serviceable mobile means for evacuation, mobile facilities for aerial reconnaissance, etc.). Set $K_e = \{Ot_i, Pvt_k\}$ is geo-information context of the micro situation for linguistic, central concept e. Set $K_c = \{k_{e_i}\}, i = \overline{1,m}$ expresses relationship Ot_i of the central concept e with other minor concepts Pvt_k that are being participating in current micro situation. Relationship Ot_i is some dependency of the central concept from minor concepts. For current micro situation as the minor concept can act concept that is a central concept to another micro situations.

For the text, obtained in the survey through social networks - must be extracted from the set of candidates to the central concepts $cPvt$ that will be associated with a variety of micro situations Sit' precedents that represent knowledge about the previously observed emergency or disaster that was observed by survey participant.

Stages of building standard or observed micro situations:

1. Description of the current situation of the domain in the form of narrative text;
2. Identification of concepts obtained from the descriptions from the categories;
3. Search relations between these concepts;
4. Building the description in a language of representation of micro situations.

Concept Pvt for considered category $eKat$ receive a result of executing the function to identify the concepts ERp

$$ERp(TextExp) \to Pvt_i$$

where Pvt_i – identified concepts, $i = \overline{1,n}$.

And we have

$$Sit' \Longrightarrow \{cuP, Ot\}$$

where Sit' micro situation, for which are corresponded: cuP – a set of concepts, which are entities and Ot – a set of concepts that express the relationship between the other concepts.

Step 1. Choose from a variety of candidates cuP set of central concepts or precedents - $cPvt$ (nouns that are the subject of a sentence $TextExp$).

$$ERp1(cuP) \rightarrow cPvt_i$$

where $ERp1$ is a function of identifying candidates into central concepts - $cPvt$

Step 2. Choose context from relations.

$$OtA = {}^{cuP}/_{\tilde{n}Pvt}$$

for candidates received in the central concepts $cPvt$. Task is to extract a subset of relations (associations) $Ot \subset OtA$.

Elements of the set of relations Ot are the main, active and additional. Each of the set of relations $OtP_1 \in Ot$ will have relation to a central concept $cPvt_i \in cPvt$.

Step 3. Forming set of micro situations $Sit_i' = \langle cPvt_i, Ot_{cPvt_i} \rangle$ - precedents, in which elements of the sets Ot_{cPvt_i} are elements of the set Ot. At this stage received micro situations still are not complete as to the elements of Ot not yet mapped to the secondary concepts.

Step 4. Search secondary concepts. As secondary concepts are any elements of the set of candidates cuP regardless of whether they were in the set of central concepts $cPvt$, set of relations Ot or any one of them. In most cases, secondary concepts derived from additions to verbal forms, previously selected. In general, secondary concepts - are the concepts that are referenced by the association.

The decision rule is represented as follows:

$$pravResh = \hat{i} Prizn_i$$

where $Prizn_i$ – separate characteristic.

When calculating the value of a decision rule $pravResh$ each of signs $Prizn_i$ is assigned a value of true if this feature is present in the concepts and false otherwise. Denote a text description of the situation Sit obtained from an expert as $TextExp$. The notion of a primary domain text representation $etExp$ a concept expressed in terms of existing categories. The process of identifying concepts Pvt from the $TextExp$ is a function that displays text on a set of concepts:

$$ERp(TextExp) \rightarrow Pvt_i$$

where ERp– function to identify the concept; Pvt_i – identified concepts, $i = \overline{1,n}$.

For micro situation Sit' we have two sets of candidates: cuP – a set of concepts, which are entities and Ot – a set of concepts that express the relationship between the other concepts.

Conclusion

The proposed approach is the basis for establishing an expanded knowledge base situation center for forecasting and disaster management. Using the model representation and technology knowledge base allow you to create precedent knowledge base for decision-making for considerably less time using extensive interviews eyewitnesses disaster through social networks.

Acknowledgement

The paper is published with the financial support from the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA (www.ithea.org) and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine (www.aduis.com.ua).

Bibliography

[Kuzemin, 2004a] Kuzemin A. Situation centres in modern sate // International Journal on Information Theories&Applicatios.– Bulgaria, 2004. – Vol. 11, №1.– P. 79 - 82

[Kuzemin, 2004b] Kuzemin A., I. Yanchevskiy, Sorochan M., Torojev A. The use of situation representation when searching for solution in computer aided design systems // International Journal on Information Theories&Applicatios.– Bulgaria, 2004. – Vol. 11, №1.– P. 82 - 88.

[Kuzemin, 2005] Kuzemin A., Sorochan M., Yanchevskiy I., Torojev A. The use of situation representation when searching for solutions in computer aided design systems International Journal // Information Theories & Applications. – Bulgaria. – 2005. – Vol.11. – P. 101 – 107.

[McConnell, 2007] McConnell B., Huba J. Citizen Marketers: When People Are the Message

[Parinov, 2000] Онлайновые сообщества: методы исследования и практическое конструирование. http://rvles.ieie.nsc.ru/~parinov/autoref.htm (accessed: 31.03.2014)

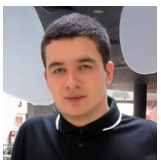
[Rimskiy, 2009] Римский В. Воздействие сети Интернет на социальную активность, формирование и развитие идентичностей // Вестник общественного мнения. 2009. № 1.

[Slipchenko, 2004] Slipchenko M., Kuzemin A., Sorochan M., Torojev A Situation centers in modern state // Proc. of the Second International Conference i.TECH. – Sofia. – Bulgaria: ITHEA. – 2004. – P. 96-98.

Authors' Information

Oleksandr Kuzomin *– Prof. Dr., Department of Informatics, Office of Innovations & Marketing, Kharkiv National University of Radio Electronics; Kharkiv, Ukraine; tel.: +38(057)7021515; e-mail: kuzy@daad-alumni.de*

Major Fields of Scientific Research: General theoretical information research, Decision Making, Emergency Prevention, Data Mining, Business Informatics.



Bohdan Tkachenko *– Master of Kharkiv National University of Radio Electronics; Kharkiv; Ukraine; tel.: +380 97 238 61 31; e-mail: bogus.weber@gmail.com*

Major Fields of Scientific Research: General theoretical information research, Knowledge Discovery and Engineering, Business Informatics.

Mariia Lysenko *– Master of Kharkiv National University of Radio Electronics; Kharkiv; Ukraine; tel.: +380 68 349 29 09; e-mail: trendskaut@gmail.com*

Major Fields of Scientific Research: General theoretical information research, Knowledge Discovery and Engineering, Business Informatics.

TABLE OF CONTENTS