ITHEA

International Journal INFORMATION THEORIES & APPLICATIONS

2010 Volume 17 Number 3

International Journal
INFORMATION THEORIES & APPLICATIONS
Volume 17 / 2010, Number 3

International Journal “INFORMATION THEORIES & APPLICATIONS” (IJ ITA)
is official publisher of the scientific papers of the members of
the ITHEA International Scientific Society

IJ ITA welcomes scientific papers connected with any information theory or its application.
IJ ITA rules for preparing the manuscripts are compulsory.
The **rules for the papers** for IJ ITA as well as the **subscription fees** are given on *www.foibg.com/ijita*.

Responsibility for papers published in IJ ITA belongs to authors.

General Sponsor of IJ ITA is the **Consortium FOI Bulgaria** (www.foibg.com).

International Journal “INFORMATION THEORIES & APPLICATIONS” Vol. 17, Number 3, 2010

Printed in Bulgaria

Edited by the Institute of Information Theories and Applications FOI ITHEA, Bulgaria, in collaboration with:
V.M.Glushkov Institute of Cybernetics of NAS, Ukraine,
Institute of Mathematics and Informatics, BAS, Bulgaria,
Universidad Politécnica de Madrid, Spain.

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

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

CONSTRAINT CONVEXITY TOMOGRAPHY AND LAGRANGIAN APPROXIMATIONS

Levon Aslanyan, Artyom Hovsepyan, Hasmik Sahakyan

Abstract: This paper considers one particular problem of general type of discrete tomography problems and introduces an approximate algorithm for its solution based on Lagrangian relaxation. A software implementation is given as well.

Keywords: discrete tomography, lagrangian relaxation.

ACM Classification Keywords: F.2.2 Nonnumerical Algorithms and Problems: Computations on discrete structures.

Introduction

Discrete tomography is a field which deals with problems of reconstructing objects from its projections. Usually in discrete tomography object T, represents a set of points in multidimensional lattice. Some measurements are performed on T, each of which contains projection, which calculates number of points of T along parallel directions. Given finite number of such measurements it is required to reconstruct object T, or if it is not possible to find unique reconstruction, construct an object which satisfies given projections. The object existence problem even by given 3 non-parallel projections is NP-complete [1].

In recent years discrete tomography draws huge attention because of the variety of mathematical formulations and applications. Theory of discrete tomography is widely used particularly in the field of medical image processing, which is based on so called computerized tomography.

Lets consider 2-dimensional lattice and horizontal and vertical projections only. Object T can be represented as a $m \times n$ $(0,1)$ matrix, where 1s corresponds to points in T. Vector of row sums corresponds to horizontal projection and vector of column sums to vertical projection. So the problem of reconstructing the object by given horizontal and vertical projections is equivalent to the $(0,1)$-matrix existence problem with given R and S row and column sums. The latter problem was solved independently by Gale and Ryser in 1957. They gave sufficient and necessary condition for such a matrix existence and also proposed an algorithm for the matrix construction. Same problem with condition of rows inequality was investigated in [6].

In many cases orthogonal projections does not contain enough information for the objects unique reconstruction.

That's why often we consider different classes of such problems, where we impose additional constraints, for instance of geometrical nature. Such constraints narrow the solutions set but at the same time could make the problem hard to solve. Typical examples of such constraints are convexity and connectivity.

We say that matrix has row (or horizontal) convexity feature if all ones in the row forms a continuous interval. Same way we define column (or vertical) convexity. Connectivity is the feature of moving between 1s in neighboring cells. In our case we consider only vertical and horizontal connectivity (not diagonal).

Existence problem for connected matrices is NP-complete [2]. Existence problems for horizontally or vertically convex, and for both horizontally and vertically convex matrices are also NP-complete [3].

Different authors proved that horizontally and vertically convex and connected matrices reconstruction problem can be solved in polynomial time. Given description shows how sensitive are this kind of problems to input conditions. We see that existence problem's complexity changes along with adding new constraints. At the same time there are a lot of other notations of the problem for those the complexity is not even known. Particularly that means that they also lacks easy solution algorithms.

So we consider several problems in the field of discrete tomography, propose ways for constructing such matrices that satisfy constraints (convex or nearly convex, satisfying given parameters or having values near to given parameters). Further we will formulate the problems as optimization problems and give ways for their approximation, based on the integer programming relaxation. The question is that integer programming model is known for being used to reformulate known NP complex optimization problems. This model's (precise or approximate algorithms construction) investigation is very important and often this model is used to approximate optimizations problems [4, 6]. Implemented algorithms and software package based on that algorithms give an ability to make calculations either for tomography problem or for similar problems, such that those calculations might guide us or give approximate or precise solutions.

In this paper we will consider one problem from the field of discrete tomography, horizontally convex matrix existence problem.

Horizontally convex matrix existence problem

Since 1's in the horizontally convex matrix are in neighboring position then if we count the number of 1's in the matrices rows, that number for convex matrices will be maximum for the ones with same parameters. That's why problems that are often considered are related to number of neighboring 1's, their constraints and optimization.

$R=(r_1,\cdots,r_m)$, $S=(s_1,\cdots,s_n)$, $R'=(r_1',\cdots,r_m')$ vectors are given. Is there a $m\times n$ $X=\{x_{i,j}\}$ matrix such that R is row sum vector for that matrix and S is column sums vector, and number of neighboring 1's in row i is equal to r_i'.

$$\begin{cases}\sum_{i=1}^{m} x_{i,j}=s_j,\ j=1,\cdots,n\\ \sum_{j=1}^{n} x_{i,j}=r_i,\ i=1,\cdots,m\\ \sum_{j=1}^{n-1}\min(x_{i,j},x_{i,j+1})=r_i',\ i=1,\cdots,m\\ x_{i,j}\in\{0,1\}\end{cases}$$

In other words the problem is following, find the matrix with horizontal convexity in the class of $(0,1)$ matrices with given row and column sums. This problem is NP-complete, since for the case when $r_i'=r_i-1,\ i=1,\cdots,m$ it's equivalent to the horizontally convex matrix existence problem. Given particular case just require the matrix to be horizontally convex by neighboring 1s in the rows.

As we already mentioned lot of combinatorial problems are suitable to represent as integer linear optimization problems. Lets reformulate our problem as integer programming problem.

Lets define $y_{i,j} \in \{0,1\}$ variables the way that it provides neighboring 1's in row i.

$$(y_{i,j} = 1) \Leftrightarrow (x_{i,j} = 1) \,\&\, (x_{i,j+1} = 1),\ i = 1,\cdots,m;\ j = 1,\cdots,n-1$$

This can be done by satisfying conditions

$$\begin{cases} y_{i,j} \le x_{i,j} \\ y_{i,j} \le x_{i,j+1} \\ y_{i,j} \ge x_{i,j} + x_{i,j+1} - 1 \end{cases}$$

So we reformulate the problem in the following way.

$R = (r_1,\cdots,r_m)$, $S = (s_1,\cdots,s_n)$, $R^{'} = (r_1^{'},\cdots,r_m^{'})$ vectors are given: Is there a $m \times n$ $X = \{x_{i,j}\}$ matrix such that

$$\begin{cases} (1)\sum_{i=1}^{m} x_{i,j} = s_j,\ j = 1,\cdots,n \\ (2)\sum_{j=1}^{n} x_{i,j} = r_i, i = 1,\cdots,m \\ (3)\begin{cases} y_{i,j} \le x_{i,j} \\ y_{i,j} \le x_{i,j+1} \\ y_{i,j} \ge x_{i,j} + x_{i,j+1} - 1 \end{cases} \quad i = 1,\cdots,m,\ j = 1,\cdots,n-1 \\ (4)\sum_{j=1}^{n-1} y_{i,j} = r_i^{'},\ i = 1,\cdots,m \\ (5) x_{i,j} \in \{0,1\},\ y_{i,j} \in \{0,1\} \end{cases}$$

Lagrangean relaxation and variable splitting

So we have horizontal row convex matrix existence problem, which is reformulated as linear integer programming problem I. We also know that problem I is NP-complete. To solve this problem we will use a method based on Lagrangian relaxation.

Obviously if we drop some of the constraints we will get problems relaxation. Assume that we can call one or several constraints hard in the since that by dropping those constraints we can solve resulted integer programming problem more easily. Constraints dropping could be embedded in more common method which is called Lagrangian relaxation. We can apply Lagrangian relaxation to given method in various ways. One of the ways, which we will use here is following, if the problem can be splitted to subproblems, which have common variables, first split those variables and then relax their equality constraint.

So, we take two set of variables $x^h_{i,j}$ and $x^v_{i,j}$ by duplicating $x_{i,j}$ variables, and reformulate our problem as

$$\begin{cases} (1)\sum_{i=1}^{m} x^v_{i,j} = s_j, j = 1,\cdots,n \\ (2)\sum_{j=1}^{n} x^h_{i,j} = r_i, i = 1,\cdots,m \\ (3)\begin{cases} y_{i,j} \le x^h_{i,j} \\ y_{i,j} \le x^h_{i,j+1} \\ y_{i,j} \ge x^h_{i,j} + x^h_{i,j+1} - 1 \end{cases} \quad i = 1,\cdots,m,\ j = 1,\cdots,n-1 \\ (4)\sum_{j=1}^{n-1} y_{i,j} = r_i^{'}, i = 1,\cdots,m \\ (5) x^h_{i,j}, x^v_{i,j} \in \{0,1\}, y_{i,j} \in \{0,1\} \\ (6) x^h_{i,j} = x^v_{i,j} \end{cases}$$

We split our original problem using variable splitting to two problems, each of which has its own variable set and which would be independent without constraint (6). From this point of view constraint (6) is the hardest one. We will relax constraint (6) using Lagrangian relaxation with coefficients $\lambda_{i,j}$.

We get following problem $VSI(\lambda)$, and its optimal value is $v^{VSI}(\lambda)$.

$$\begin{cases} \max(\sum_{i,j} \lambda_{i,j}(x^h_{i,j} - x^v_{i,j})) \\ (1)\sum_{i=1}^{m} x^v_{i,j} = s_j, j = 1,\cdots,n \\ (2)\sum_{j=1}^{n} x^h_{i,j} = r_i, i = 1,\cdots,m \\ (3)\begin{cases} y_{i,j} \le x^h_{i,j} \\ y_{i,j} \le x^h_{i,j+1} \\ y_{i,j} \ge x^h_{i,j} + x^h_{i,j+1} - 1 \end{cases} \quad i = 1,\cdots,m,\ j = 1,\cdots,n-1 \\ (4)\sum_{j=1}^{n-1} y_{i,j} = r_i^{'}, i = 1,\cdots,m \\ (5) x^h_{i,j}, x^v_{i,j} \in \{0,1\}, y_{i,j} \in \{0,1\} \end{cases}$$

Then using same method we can further split the problems into subproblems for rows and columns, which itself is reducing to the finding of simple path, with given number of edges and biggest weight on directed graph.

We can approach the problem in other way, by relaxing constraint (3) we would split the problem into two subproblems with $x_{i,j}$ and $y_{i,j}$ variables. But this paper is limited with first approach.

Obviously problem $VSI(\lambda)$ is relaxation of problem I, hence $v^{VSI}(\lambda)$ is upper limit for value of I. Find best upper limit means to solve Lagrangian dual problem which is

$$v^{VSD} = \min_{\lambda} v^{VSI}(\lambda)$$

This is convex non-differential optimization problem: There are different methods for solving this problem. One of them is subgradient optimization method. Subgradient optimization on each step calculates the value of $v^{VSI}(\lambda)$ for given $\lambda_{i,j}$, in this case that equals to solving following m independent problems

$$\begin{cases} \max(\sum_{j=1}^{n} c_j x_j) \\ \sum_{j=1}^{n} x_j = r \\ \begin{cases} y_j \le x_j \\ y_j \le x_{j+1} \qquad j = 1, \cdots, n-1 \\ y_j \ge x_j + x_{j+1} - 1 \end{cases} \\ \sum_{j=1}^{n-1} y_j = r' \\ x_j \in \{0,1\}, y_j \in \{0,1\} \end{cases} \qquad (*)$$

We will try to solve these problems using algorithm for finding simple path on acyclic directed graph with biggest cost and given number of edges.

Decomposed problem on graph and the solution

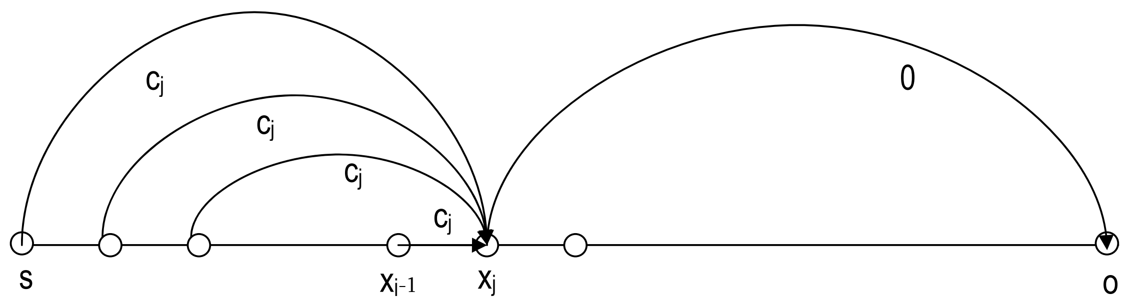
We consider directed graph $G = (V, E)$ which vertex set consists of vertexes for each x_j variable plus s source and o destination. We define edges in following way

(s, x_j) with weight c_j

$(x_i, x_j), i \le j-1$ with weight c_j

(x_j, o) with weight 0

Consider the paths form s to o. Only r variables corresponding to x_j vertexes, are 1's according to (*) and among them r' is neighboring 1's. Hence we are interested only in those paths from s to o that have only r vertexes and there are only r' with neighboring 1's. We need to find among those paths, the one that has maximum weight. Now by assigning 1's to variables corresponding to vertexes we will get solution to the problem (*).



Now lets give algorithmic description.

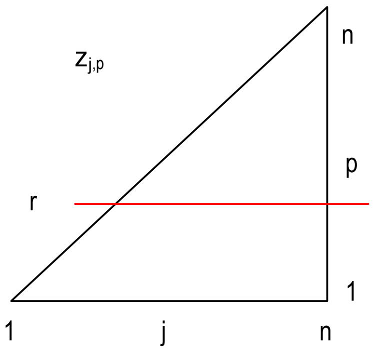
Let $z(j,p)$ is weight of the longest path from s to x_j vertex with p vertexes on it. Lets $w(j,p,q)$ is weight of the longest path from s to x_j which has p vertexes on it and there are q neighboring vertexes with corresponding variables equal to 1. In this case $z(j,p)$ and $w(j,p,q)$ can be calculated the following way. First of all consider $z(j,p)$

$$z(j,1) = c_j$$
$$z(j,p) = \max_{u<x_j}(z(u,p-1)+c_j)$$

And the optimal value we're looking for is $z(o,r) = \max_j z(x_j, r)$.

This part of the problem is solving in the following way.

For given n and r $Z_{j,p} = [z_{j,p}]$ array is constructed, where $j = 1,\dots,n$ and for j $p = 0,\dots,j-1$. In reality for fixed r its enough to consider $p = 1,\dots,r$ layers, but $p = 1,\dots,n$ will satisfy calculations needed for all r.



First of all $z_{1,1}$ value is calculated. That's equal to c_1. All values of row $p=1$ are calculated in the same way $z_{j,1}=c_j$. To calculate $z_{j,p}$ by our formula we need to know values for $p-1$ and for all $1,\dots,j-1$ indexes. But in row $p-1$ first non-zero value is in $j=p-1$ position, which is on diagonal. So calculations can be done sequentially on $p=1,\dots,r,\dots$ rows and in rows in order $j=p,\dots,n$. This constructs are needed for software implementation and these give ability to measure number of operations in calculation. It doesn't exceed n^3, which means polynomial complexity.

Maximal weight paths can be stored in a separate array. They can be stored as 0,1 vectors or as indexes of non zero elements which however won't significantly decrease number of computations.

Now lets calculate values of $w(j,p,q)$. First of all lets consider edge values. From $w(j,p,q)$ we have maximal weight path from s to x_j which has p vertexes and there are q pairs with neighboring 1's. $q \le p-1$ and lets p's are decreased up to $q+1$. $w(j,q+1,q)$'s can be non-zero starting from $j \ge q+1$. For bigger q's and smaller j's $w(j,p,q)$'s are equal to 0.

Interestingly q can't be very small. If $p > \left]\frac{j+1}{2}\right[$ then q can't be 0 (at least 2 vertexes must have neighboring indexes).

Let $\tau = \left[\frac{j+1}{3}\right]$. In that case τ vertex pairs still might not be neighbors, which gives 2τ vertexes. After that any new vertex addition would add 2 new pairs.

Now lets consider common case. For calculating $w(j,p,q)$ lets consider class where for j $p \le j$ and for j,p pairs $q \le p-1$. This class is larger than needed but in reality it doesn't differ much from the minimal class which is necessary for calculations. For slight transition of edge values class is zeroed before performing calculations. Lets investigate value of $w(j,p,q)$. We do chain calculations and on each step consider 2 cases $x_{j-1}=1$ and $x_{j-1}=0$. So we get following values

$w(j-1,p-1,q-1)+c_j$ and $\max\limits_{u<x_j-1}(w(u,p-1,q)+c_j)$

We are interested in maximum of these values.

In $w(j-1, p-1, q-1)+c_j$ all indexes are less than preceding and we assume that this value is already calculated in previous steps. For calculating $\max\limits_{u<x_j-1}(w(u, p-1, q)+c_j)$ we do next step in chain calculation.

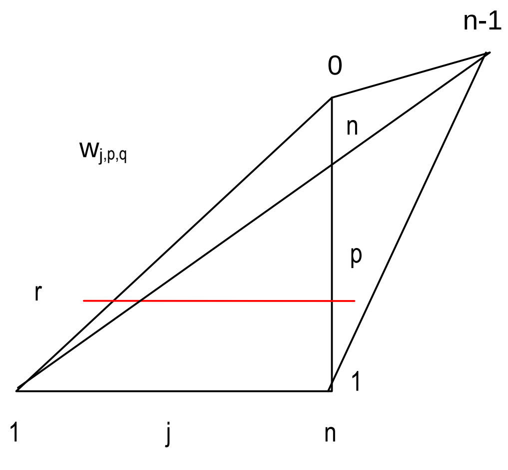
$$w(u-1, p-2, q-1)+c_j$$

$$\max_{v<u-1}(w(v, p-2, q)+c_j)$$

And the needed optimal value is $w(o,r,r')=\max\limits_j w(x_j,r,r')$. This problem practically can be solved in following way.

For given n, r, r' we construct the class given above, array $W_{j,p,q}=[w_{j,p,q}]$, where for $j=1,...,n$ and j $p=1,...,j$ and for pair j,p $q=0,...,p-1$.

In reality for fixed r it's enough to consider $p=1,...,r$ layers and for q all values where $q \le r-1$. But calculations must be done in such sequence to be executable.



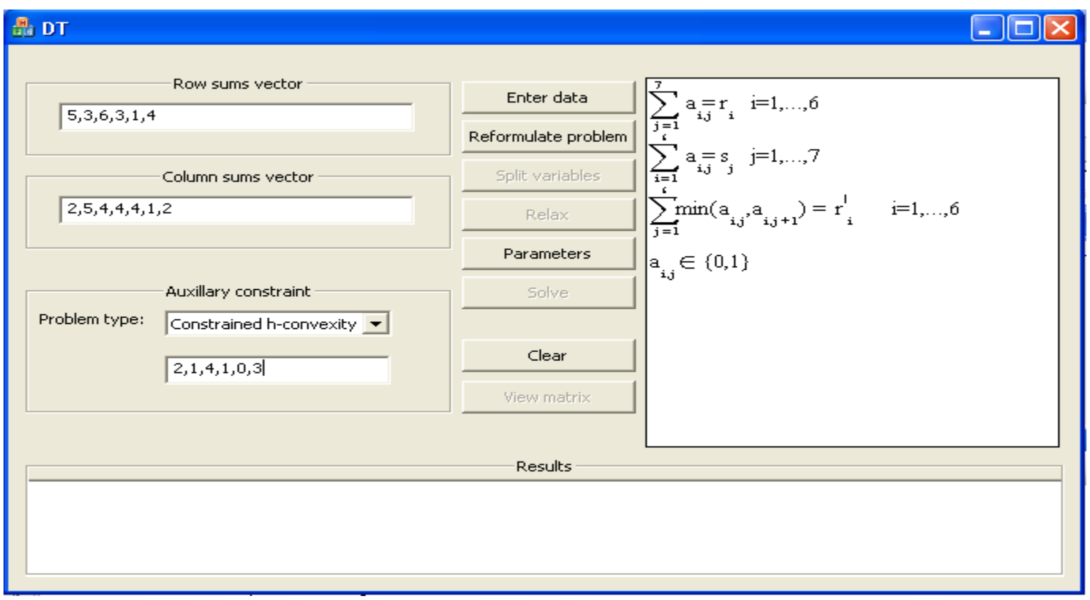
First $w_{1,1,0}$ values are calculated, $w_{1,1,0}=c_1$, all values in row $p=1$ are calculated in the same way $w_{j,1,0}=c_j$. More, $q=0$ values were already considered. To calculate $w_{j,p,q}$ based on our formula we need to know values for $p-1$ and all $1,..., j-1$. But in layer $p-1$ with current q value is either 0 or already calculated. Then calculations can be done in layers $p = 1,...,r,...$ sequentially and in layers in order of $j = p,...,n$. Given constructions are needed for software implementation and give ability to measure number of calculations. Those are not more than n^4 which means polynomial complexity.

Maximal weight paths that we're looking for could be stored in separate array as 0,1 vectors or as array of indexes with non-zero values, which however won't significantly lower number of calculations.

Software implementation

Based on given methods a software system with an UI was implemented, which can be used to solve some problems from the field of discrete tomography based on Lagrangian relaxation.

There are several fields which are used for data input. Since we are solving problems in the field of discrete tomography so input data are projections, in our case row sums and column sums. Also we are giving specific problem description by additional constraints. So we have special fields for that purpose. Then there is special control which can be used to reformulate given problem as mathematical programming problem. Then we can choose one or several constraints which we want to relax. Also we can do variable splitting etc. And there is an output window which is used for displaying results. For example value of Lagrangian Dual or variables difference as a result of splitting.



In given example as a problem is considered horizontal row convexity existence problem.

Now lets describe one of the main classes in the implementation, ProblemBase abstract class. This class is base for all problem. Class encapsulates problem data. It also has several virtual functions which are used for problem solution. For example function which reformulates the problem as mathematical programming problem, chooses constraints for relaxation. Important function in ProblemBase is Solve method, which invokes the method for specific problem. Since most of the problems are reducing to relatively easy problems on graphs and methods for those solutions can be used for different problems we put those methods in a separate library.

Bibliography

1. Gardner R.J., Gritzmann P., Prangenberg D., On the computational complexity of reconstructing lattice sets from their X-rays. Technical Report (970-05012), Techn. Univ. Munchen, fak. f. math, 1997.

2. G.J. Woeginger. The reconstruction of polyominoes from their orthogonal projections. Inform. Process. Lett., 77:225-229, 2001.

3. E. Barcucci, A. Del Lungo, M. Nivat, and R. Pinzani. Reconstructing convex polyominoes from horizontal and vertical projections. Theoret. Comput. Sci., 155:321-347, 1996.

4. G. Dahl and T. Flatberg. Lagrangian decomposition for reconstructing hv-convex (0, 1) matrices, Report 303, University of Oslo, p. 1-13, 2002.

5. M. Guignard and S. Kim. Lagrangian decomposition: a model yielding stronger lagrangian bounds. Math. Prog., 39:215-228, 1987.

6. H.J. Ryser. Combinatorial properties of matrices of zeros and ones. Canad. J. Math., 9:371-377, 1957.

7. D. Gale. A theorem on flows in networks. Paci_c J. Math., 7:1073-1082, 1957.

Authors' Information

Levon Aslanyan – Head of Department, Institute for Informatics and Automation Problems, P.Sevak St. 1, Yerevan 14, Armenia, e-mail: lasl@.sci.am

Artyom Hovsepyan – Researcher, Institute for Informatics and Automation Problems, P.Sevak St. 1, Yerevan 14, Armenia, e-mail: artyom.hovsepyan@gmail.com

Hasmik Sahakyan – Leading Researcher, Institute for Informatics and Automation Problems, NAS RA, P.Sevak St. 1, Yerevan 14, Armenia, e-mail: hasmik@ipia.sci.am

A NEW ALGORITHM FOR THE LONGEST COMMON SUBSEQUENCE PROBLEM

Vahagn Minasyan

Abstract: *This paper discusses the problem of determining the longest common subsequence (LCS) of two sequences. Here we view this problem in the background of the well known algorithm for the longest increasing subsequence (LIS). This new approach leads us to a new online algorithm which runs in $O(d \log n)$ time and in $O(n)$ space where n is the length of the input sequences and d is the number of minimal matches between them. Using an advanced technique of van Emde Boas trees the time complexity bound can be reduced to $O(d \log \log n)$ preserving the space bound of $O(n)$.*

Keywords: *longest common subsequence, longest increasing subsequence, online algorithm.*

ACM Classification Keywords:*G.2.1 Discrete mathematics: Combinatorics*

Introduction

Let $A = a_1 \cdots a_i \cdots a_m$and $B = b_1 \cdots b_j \cdots b_n$, $1 \leq m \leq n$, be two sequences over some alphabet Σ of size s, $s \geq 1$. A sequence $C = c_1 \cdots c_k \cdots c_l$, $1 \leq l$, over Σ is called a *subsequence* of A, if C can be obtained from A by deleting some of its elements, that is if exists a set of indices $\{i_1, \ldots, i_k, \ldots, i_l\}$ such that $1 \leq i_1 < \cdots < i_k < \cdots < i_l \leq m$ and $c_k = a_{i_k}$ for $1 \leq k \leq l$. C is said to be a *common subsequence* of A and B, if it is a subsequence of both sequences A and B;C is said to be a *longest common subsequence (LCS)* of A and B, if it has the maximum length among all common subsequences of A and B; that length is called the *LCS length* of A and B. In general the longest common subsequence is not unique.

The *Longest Common Subsequence Problem (LCS Problem)* is to determine a LCS of $A$and B. Oftenthe problem of determining theLCS length is also referred to as LCS Problem. This is due to the fact that most of algorithms intended to find the LCS length can easily be modified to determine a LCS [Bergroth, 2000]. In this paper we will concentrate on determining the LCS length rather thandetermining an actual LCS.The first known solution of the LCS Problem is based on dynamic programming [Cormen, 2009].For $1 \leq i \leq m$ and $1 \leq j \leq n$ denote by $l_{i,j}$ the LCS length of sequences $a_1 \cdots a_i$and $b_1 \cdots b_j$; thus $l_{m,n}$ is the LCS length of A and B. Note that thefollowing recursion holds for $l_{i,j}$:

$$l_{i,j} = \begin{cases} 0 \text{ if } i = 0 \text{ or } j = 0 \\ l_{i-1,j-1} + 1 \text{ if } a_i = b_j \\ \max\{l_{i-1,j}, l_{i,j-1}\} \text{ if } a_i \neq b_j \end{cases} \quad (1)$$

Based on this relation it is easy to construct an algorithm which fills an array of size $m \times n$, where (i,j)-th cell contains the value of $l_{i,j}$. As it follows form (1) such algorithm has to fill the rest of array before obtaining the value of (m,n)-th cell, so it will determine the LCS length of sequences A and Bin $\Theta(mn)$ time and $\Theta(mn)$ space ($\Theta(1)$ time for filling each cell and $\Theta(1)$ space for holding each cell).A simple trick can be used to make this algorithm require only $\Theta(m+n)$ space to obtain the value of the (m,n)-th cell [Cormen, 2009]. Here we give some definitions which will be used later in the paper. For $1 \leq i \leq m$ and $1 \leq j \leq n$ the pair (i,j) is called *matching* between sequences A and B if $a_i = b_j$; it is called *minimal (or dominant) matching* if for every other matching (i',j') such that $l_{i,j} = l_{i',j'}$ it holds $i' > i$ and $j' \leq j$ or $i' \leq i$ and $j' > j$.Note that if m' and n' are two integers such that $m \leq m'$and $n \leq n'$, then theLCS Problem for two sequences of size m and n is asymptotically not harder than the LCS Problem for two sequences of size m' and n'. Indeed, given two sequences of size m and n and an algorithm which solves the LCS Problem for two sequences of size m' and n', we can lengthen the given sequences (by appending to themsymbols which don’t occur in the initial sequences) up to size m' and n' respectively and pass the resulting two sequences to the given algorithm. It is easy to see that such algorithm will solve the LCS Problem for two sequences of size m and n in asymptotically the same time and space bounds as the given algorithm solves the LCS Problem for two sequences of size m' and n'. This means that each lower bound for the LCS Problem for two sequences of size m andeachupper bound for the LCS Problem for two sequences of size n are respectively lower and upper bounds for the LCS Problem for two sequences of size m and n (recall that $m \leq n$). At [Aho, 1976] the LCS Problem is examined using the decision tree model of computation where the decision tree vertices represent “equal-unequal” comparisons. There it is shown that each algorithm solving the LCS Problem and fitting this model has time complexity lower bound of $\Omega(ms)$, where s is the number of distinct symbols occurring in the sequences (i.e. the alphabet size). This means that the LCS Problem with unrestricted size of the alphabet has time complexity lower bound of $\Omega(mn)$, as such LCS Problem can be viewed as an LCS Problem with restricted alphabet of size $(m+n)$.In practice the underlying encoding scheme for the symbols of the alphabet implies a topological order between them. Algorithms which take into account this fact don’t fit the decision tree model with “equal-unequal” comparisons examined at [Aho, 1976]. At [Masek, 1980] it is presented an algorithm which applies the “Four Russians” trick to the dynamic programming approach, thusit doesn’t fit the model examined at [Aho, 1976] and has time complexitybound of$O(mn/\log m)$.This bound isasymptotically the best known for general case LCS Problem[Cormen, 2009].

Previous Results

Lot of algorithms have been developed for the LCS Problem that, although not improving the time complexity bound $O(mn)$, exhibit much better performance for some classes of sequences A and B [Bergroth, 2000]. Consider the special case when the alphabet Σ consists of first n integers, i.e. $\Sigma = \{1, ..., j, ..., n\}$, and the sequences A and B are two permutations of Σ. It is easy to check that this case can be reduced to the case where B is the identical permutation (by replacing b_j by j for $1 \leq j \leq n$ in both sequences A and B we will get two sequences which are equivalent to the initial ones with respect to the LCS Problem). In this case each LCS of A and B is an increasing sequence of some of first n integers and each such sequence is a LCS of $A$and B. Thusin the case when A and $B$are permutations the LCS Problem is reduced to the problem of determining a longest increasing subsequence of permutation B. The *Longest Increasing Subsequence (LIS) Problem* is to determine a non decreasing subsequence of maximum length in the given sequence of integers. The LIS Problemcan be solved in $O(n \log n)$time [Fredman, 1975], and using advanced data structures like van Emde Boas trees [Cormen, 2009]thistime bound can be reduced to $O(n \log\log n)$. Thus these bounds apply to the LCS Problem in the case of permutations.Also there are many algorithms for the general case LCS Problem which except$m$ and $n$are also sensitive for other parameters like the LCS length, the alphabet size, the number of matches and the number of minimal matches. A survey on such algorithms is given at [Bergroth, 2000].The table below gives a brief remark of some of known algorithms for the LCS Problem. There l denotes the LCS length, s denotes the alphabet size, r denotes the number of all matches and d denotes the number of minimal matches. It is known [Baeza-Yates, 1999] that for two random sequences of length n the expected LCS length is $O(n)$ and the expected number of minimal matches is $O(n^2)$ [Tronicek, 2002]. This means that (except the 5th) none of the algorithms mentioned in the table hastime complexity upper bound less than$O(mn)$not only in the worst case but also in the average case.

No.	Year	Authors	Time Complexity	Ref.
1	1974	Wagner, Fischer	$O(mn)$	[Cormen, 2009]
2	1977	Hunt, Szymansky	$O(m + r \log l)$	[Hunt, 1977]
3	1977	Hirschberg	$O(ln)$	[Hirschberg, 1977]
4	1977	Hirschberg	$O(l(m - l) \log n)$	[Hirschberg, 1977]
5	1980	Masek, Paterson	$O(mn/\log m)$	[Masek, 1980]
6	1982	Nakatsu et al.	$O(n(m - l))$	[Nakatsu, 1982]
7	1984	Hsu, Du	$O(lm \log(n/l))$	[Hsu, 1984]

8	1986	Myers	$O(n(n-l))$	[Myers, 1980]
9	1987	Apostolico, Guerra	$O(m\log n + d\log(2mn/d))$	[Apostolico, 1987]
10	1987	Apostolico, Guerra	$O(lm\log(2n/m))$	[Apostolico, 1987]
11	1990	Chin, Poon	$O(ns+\min\{lm, ds\})$	[Chin, 1990]
12	1990	Wu, Manber, Myers	$O(n(m-l))$	[Wu, 1990]
13	1992	Apostolico et al.	$O(n(m-l))$	[Apostolico, 1992]
14	1994	Rick	$O(ns+\min\{lm, l(n-l)\})$	[Rick, 1994]
15	1994	Rick	$O(ns+\min\{lm, ds\})$	[Rick, 1994]
16	2002	Goeman, Clausen	$O(ns+\min\{lm, l(n-l)\})$	[Goeman, 2002]

All these algorithms are developed in the background of building the $m \times n$ array mentioned in the dynamic programming approach, and they purport to perform fewer operations in order to obtain the (m,n)-th cell of that array. In this paper we view the LCS Problem inanother background, namely the background of the classical algorithm for the LIS Problem described at [Fredman, 1975]. For sure each term we deal with in this background has its direct analogue in the background of the $m \times n$ array; however our approach can be justified by the fact that it leadsus to simpler constructions and an $O(d\log n)$ algorithm for the LCS Problem which can be reduced to $O(d\log\log n)$ if using van Emde Boas trees (details are in the next section). Initially algorithms from 10th to 16th require $O(ns)$ space, but at [Apostolico, 1987] a trick is introduced which can be used to reduce the space complexity to $O(n)$, however in this case the time complexity bounds increase by a multiplicative factor of $\Theta(\log s)$. The 9th algorithm requires $O(ns)$ space but that trick cannot be used to reduce this space complexity bound[Apostolico, 1987]. Recall that d is the number of minimal matches. It can be checked that $d \leq l(m-l)$ [Rick, 1994] and it is known that in average it holds $d = \Theta(mn)$ [Tronicek, 2002]. This means the 9th,10th and 14th algorithms mentioned in the table above have better time complexity bounds than the others mentioned there. The algorithm we present here has better time complexity bound than 10th and 14th in case when $s = \omega(\log n)$ (or $s = \omega(\log\log n)$ if the van Emde Boas trees are used), and it has better space complexity bound than 9th in cases when $s = \omega(1)$ (see [Cormen, 2009] for the ω-notation). Roughly speaking the algorithm we present here has better time and space complexity bounds than the ones mention in the table above when the alphabet size if relevantly larger. We present the algorithm in the next section.

The New Algorithm

First we will discuss the algorithm for the LIS Problem presented at [Fredman, 1975]. That algorithm is an *online algorithm* meaning that it sequentially handles the elements of the input sequence and determines the LIS length of the sequence handled so far.Online algorithms have advantage that they can run on dynamically changing input data. For instanceunlike the Selection Sort, the Insertion Sort algorithm can maintain the sorted list upon the appendingof the next element to the input list [Cormen, 2009]. Thus such algorithms are defined as update procedures which are to be performed upon the appending of the next element. Now back to the LIS Problem.Let $A = a_1 \cdots a_i \cdots a_m$ be a sequence of integers and let x be an integer which is being appended to A. We will describe an online algorithm which determines the LIS length of A' which is A appended by x. Denote by l the LIS length of A and by l' the LIS lengthof A'. Note that $l' = l$or $l' = l + 1$. For $1 \le k \le l$ there are increasing subsequences of length $k$in A. Let $x_k$be the minimumof their last elements. It is easy to check that

$$x_1 \le \cdots \le x_k \le \cdots \le x_l \qquad (2)$$

We denote by x'_k the analogue of $x_k$inA': for $1 \le k \le l'$ let x'_k denote the minimum of the last elements of increasing subsequences of length k of A'. In order to obtain an online algorithm for the LIS Problem we will describe how to determine values $(x'_k)_{k=1}^{l'}$ based on values $(x_k)_{k=1}^{l}$. Firstly note that $l' = l + 1$ if and only if $x_l \le x$, and if so then $x'_{l+1} = x$. It is easy to check that this claim can be generalized for any $1 \le k \le l$: let r denote $l + 1$ if $x_l \le x$ and otherwise let r be the least index such that $x < x_r$. It is easy to check that for $k - r$ it holds $x'_k = x$and otherwise $x'_k = x_k$. Thus we have described a way how to obtain values $(x'_k)_{k=1}^{l'}$ based on values $(x_k)_{k=1}^{l}$. Next the online algorithm for the LIS Problem is described. The algorithm maintains the values $(x_k)_{k=1}^{l}$ in an array **endpoints**. Upon the appending of the next element x to sequence A the algorithm just searches for the index r mentioned above and updates the value at that index.

LIS-update

Input: the next element **x** of sequence A

Output:the LIS length of the sequences$A$ handled so far

Method:

1. **if** (**x** *has no upper bound in* **endpoints**) **then do**
2. *append* (**endpoints**, **x**)
3. **done else do**
4. **endpoints**[*upper_bound* (**endpoints**, **x**)] = **x**
5. **done**
6. **output** *size* (**endpoints**)

Note that each call of this procedure requires $\Theta(\log l)$ time where l is the LIS length of the sequence handled so far. Thus we have described an online algorithm for the LIS Problem which runs in $\Theta(m \log l)$ time and in $\Theta(l)$ space where m is the length of the sequence handled so far and l is the LIS length of that sequence. Next we will present an online algorithm for the LCS Problem which determines the LCS length of two sequences of length $m$and n, $m \le n$, in $O(d \log n)$ time where d is the number of minimal matches between the input sequences. As for the LCS Problem there are two input sequences some clarification is needed regarding the notion of online algorithms.By an online algorithm for the LCS Problem we mean an algorithm which can accept the next element of either of the two input sequences and provide the LCS length of the two sequences handled so far. Let $A = a_1 \cdots a_i \cdots a_m$ and $B = b_1 \cdots b_j \cdots b_n$ be two sequences over some alphabet Σ of size s and let $y \in \Sigma$ be a symbol being appendedto B. We will describe an online algorithm which determines the LCS length of $A$and B', where B' is Bappended by y. Denote by l the LCS length of A and B and by l' the LCS length of $A$and B'. Note that $l' = l$or $l' = l + 1$. For $1 \le k \le l$ there are subsequences of length $k$common to $A$and B. Let i_k be the minimum index such that there is a subsequence of length $k$common to$A$ and B ending at $i_k$in A.It is easy to check that

$$i_1 < \cdots < i_k < \cdots < i_l \quad (4)$$

Similarly for $1 \le k \le l$ we define j_k as the minimum index such that there is a subsequence of length k common to A and B ending at $j_k$in B, and we get

$$j_1 < \cdots < j_k < \cdots < j_l \quad (5)$$

We will call the indices at (4) *thresh indices* or *thresh values* of sequence$A$with respect to B and the indices at (5) thresh indices or thresh values of sequence B with respect to A. Let for $1 \le k \le l'$ i'_k be the thresh values of sequence$A$with respect to B' and j'_k be the thresh values of B' with respect to A. In order to obtain an online algorithm for the LCS Problem we will describe how to determine indices $(i'_k)_{k=1}^{l'}$ and $(j'_k)_{k=1}^{l'}$ based on indices $(i_k)_{k=1}^{l}$and $(j_k)_{k=1}^{l}$. Firstly note that $l' = l + 1$ if and only if there is some index r, $l < r \le m$, such that $x_r = y$, and if so then i'_{l+1} is the minimum of such r-s. It is easy to check that this claim can be generalized for any $1 \le k \le l$: if r is the first occurrence of y in A after i_{k-1} and $r < i_k$ then $i'_k = r$ and otherwise $i'_k = i_k$ (see Figure 1).

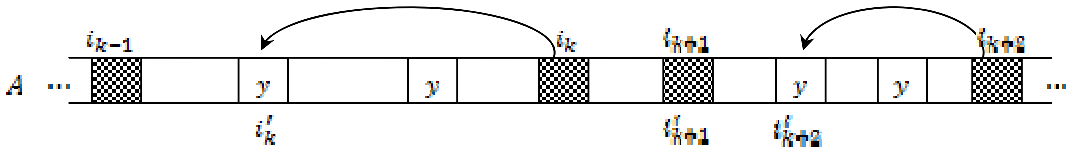


Figure 1

Thus we have described a way how to obtain sequences A and B' and their thresh indices based on sequences A and B and their thresh indices.So during this some thresh values are updated and the others are not.A trivial approach would be to handle all thresh values and update them if they has to be updated, however better would be to handle only those thresh values which has to be updated.Let r be the first occurrence of y in A aftersome i_k. Note that the least thresh value exceeding i_k which has to be updated is the first occurrence of thresh value after r. This means that while searching for the first occurrence of y (after some thresh value) the thresh values can be ignored.Also note that the thresh values of B' with respect to A, i.e. the $(j'_k)_{k=1}^{l'}$, can be obtained easily:there is a new thresh value there if and only if $l' = l + 1$ and if so then $j'_{l+1} = n + 1$. It can be checked that each update of a thresh value corresponds to a minimal match.Next the algorithm is presented. It consists of two update procedures: one for calling upon the appending the next element to sequence A and another upon the appending the next element tosequence B. We will restrict only on the second one as the first one can be obtained just by swapping symbols “A” and “B” in the text of the procedure. The algorithm maintains the sequences A and B in arrays $\mathbf{sequenceA}$ and $\mathbf{sequenceB}$ respectively and for each symbol z of alphabet Σ it maintains the set of occurrences of z in A and B in binary search trees $\mathbf{layersA}[z]$ and $\mathbf{layersB}[z]$ respectively.The algorithm also maintains the thresh indices $(i_k)_{k=1}^{l}$ and $(j_k)_{k=1}^{l}$ in binary search trees $\mathbf{threshA}$ and $\mathbf{threshB}$ respectively. Following is the update procedure which is to be called upon the appending the next element to sequence B.The procedure uses two temporary variables $\mathbf{p}$ and $\mathbf{q}$ which correspondto the next and previous values of updating thresh indices.

LCS-updateB

Input: the next element $\mathbf{y}$ of sequence B

Output:the LCS length of the sequences A and B handled so far

Method:

1. $\mathbf{p} = 0$
2. $\mathbf{q} = 0$
3. **while** (**true**) **do**
4. **if** (**q** *has no upper bound in* **layersA**[**y**]) **then break**
5. **p** = *upper_bound* (**layersA**[*y*], **q**)
6. *erase* (**layersA**[**sequenceA**[**p**]], **p**)
7. **if** (**p** *has no upper bound in* **threshA**) **then break**
8. **q** = *upper_bound* (**threshA**, **p**)

9. *insert* (**layersA[sequenceA[q]], q**)
10. *erase* (**threshA, p**)
11. *insert* (**threshA, q**)
12. **done**
13. **if (p** *has no upper bound in* **threshA) then do**
14. *insert* (**threshA, p**)
15. *insert* (**threshB,** *size* (**sequenceB**))
16. **done else do**
17. *insert* (**layersB[y],** *size* (**sequenceB**))
18. **done**
19. *append* (**sequenceB, y**)
20. **output** *size* (**threshA**)

Note that each iteration of the while loop at lines 3-12 updates a thresh value ($\mathbf{p} \leq \mathbf{q}$ at the end of each iteration) and the operations carried out during each iteration require $\Theta(\log n)$ time as they are performed on binary search trees. Recall that each update of the thresh value corresponds to a minimal match, so we have described an online algorithm for the LCS Problem which runs in $\Theta(d \log n)$ time and in $\Theta(m + n)$ space where m and n are the lengths of the sequences handled so far and d is the number of minimal matches between that sequences. These bounds can be improved if using van Emde Boas trees [Cormen, 2009] instead of binary search trees.van Emde Boas tree is a data structure that for some a priory fixed integer w can store some of first 2^w integers,itsupports operations of insertion deletion and search for the upper bound with worst case time complexity bound of $\Theta(\log \log n)$and it requires $\Theta(2^w)$ space regardless the number of integers stored in it. At [Cormen, 2009] it is shown how this data structure can be modified to require only $\Theta(n)$ space where n is the number of stored elements (there the modified data structure is called y-fast trie). In this case the operations of insertion and deletion do not have worst case time complexity bound of $\Theta(\log \log n)$ but this bound holds for the amortized time complexity. This fits with our needs as we perform $\Theta(d)$ insertions and deletions, thus we conclude that if using these modified van Emde Boas trees then the algorithm presented in this paper will run in $\Theta(d \log \log n)$ time and in $\Theta(n)$ space.

Bibliography

[Aho, 1976]A. Aho, C. Hirschberg, J. Ullman. Bounds on the Complexity of the Longest Common Subsequence Problem. Journal of the Association for Computing Machinery, Vol. 23, No. 1, 1976, pp. 1-12.

[Apostolico, 1987]A. Apostolico, G. Guerra. The longest common subsequence problem revisited. Algorithmica 2, 1987, pp. 315-336.

[Apostolico, 1992]. Apostolico, S. Browne, C. Guerra. Fast Linear-Space Computations of Longest Common Subsequences. Theoretical Computer Science, Vol. 92, 1992, pp. 3-17.

[Baeza-Yates, 1999] R. Baeza-Yates, R. Gavalada, G. Navarro, R. Scheihing.Bounding the Expected Length of Longest Common Subsequences and Forestes.Theory of Computing Systems, Vol. 32, 1999, pp. 435-452.

[Bergroth, 2000]L. Bergroth, H. Hakson, T. Ratia.A Survey of Longest Common Subsequence Algorithms.Proceedings of the Seventh International Symposium on String Processing Information Retrieval, 2000, pp. 39-48.

[Chin, 1990]. F. Chin, C. Poon. A Fast Algorithm for Computing Longest Common Subsequences of Small Alphabet Size.Journal of Information Processing, Vol.13, No. 4,1990, pp. 463-469.

[Cormen, 2009] T. Cormen, C. Leiserson, R. Rivest, C. Stein. Introduction to algorithms (Third Edition), pp. 43-65, pp. 390-397, pp.531-561, 2009.

[Fredman, 1975] M. Fredman. On Computing the of Longest Increasing Subsequences. Discrete Mathematics, Vol. 11, No. 1, 1975, pp. 29-35.

[Goeman, 2002]H. Goeman, M. Clausen. A new practical linear space algorithm for the longest common subsequence problem.Kybernetika, Vol. 38, No. 1, 2002, pp. 45-66.

[Hirschberg, 1977]D. Hirschberg.Algorithms for the Longest Common Subsequence Problem.Journal of Association for Computing Machinery, Vol. 24, No. 4, 1977, pp. 664-675.

[Hsu, 1984]W. Hsu, M. Du. New algorithms for the LCS problem. Journal of Computer and Syste Sciences, Vol. 29, 1984, pp. 133-152.

[Hunt, 1977]J. Hunt, T. Szymanski. A fast algorithm for computing longest common subsequences.Communications of Association for Computing Machinery, Vol. 20, No. 5, 1977, pp. 350-353.

[Masek, 1980] W. Masek, M. Paterson. A Faster Algorithm Computing String Edit Distances. Journal of Computer and System Sciences, Vol. 20, No. 1, 1980, pp. 18-31.

[Myers, 1986]E.Myers. An $O(ND)$ Difference Algorithm and its Variations. Algorithmica 1, 1986, pp. 251-266.

[Nakatsu, 1982]N. Nakatsu, Y. Kambayashi, S. Yajima. A Longest Common Subsequence Algorithm Suitable for Similar Text Strings. Acta Informatica Vol. 18, 1982, pp. 171-179.

[Rick, 1994] C. Rick. New Algorithms for the Longest Common Subsequence Problem.Research Report No. 85123-CS, Department of Computer Science, University of Bonn, 1994.

[Tronicek, 2002]Z. Tronicek.Common Subsequence Automatation.Research Report DC-2002-02, Department of Computer Science and Engineering, Czech Technical University, 2002.

[Wu, 1990] S. Wu, U. Manber, G. Myers, W. Miller.An $O(NP)$ Sequence Comparison Algorithm.Information Processing Letters, Vol. 35, 1990, pp. 317-323.

Authors' information

VahagnMinasyan – Postgraduate student, Yerevan State University, faculty of Informatics and Applied Mathematics, department of Discrete Mathematics and Theoretical Informatics, Armenia, 0025, Yerevan, 1st Alex Manoogian; e-mail: vahagn.minasyan@gmail.com

INTERFERENCE MINIMIZATION IN PHYSICAL MODEL OF WIRELESS NETWORKS

Hakob Aslanyan

Abstract: Interference minimization problem in wireless sensor and ad-hoc networks is considered. That is to assign a transmission power to each node of a network such that the network is connected and at the same time the maximum of accumulated signal straight on network nodes is minimum. Previous works on interference minimization in wireless networks mainly consider the disk graph model of network. For disk graph model two approximation algorithms with $O(\sqrt{n})$ and $O((opt \ln n)^2)$ upper bounds of maximum interference are known, where n is the number of nodes and opt is the minimal interference of a given network. In current work we consider more general interference model, the physical interference model, where sender nodes' signal straight on a given node is a function of a sender/receiver node pair and sender nodes' transmission power. For this model we give a polynomial time approximation algorithm which finds a connected network with at most $O((opt \ln n)^2 / \beta)$ interference, where $\beta \geq 1$ is the minimum signal straight necessary on receiver node for successfully receiving a message.

Keywords: interference, wireless networks, graph connectivity, set cover, randomized rounding.

ACM Classification Keywords: C.2.1 Network Architecture and Design - Network topology, G.2.2 Graph Theory - Network problems.

Introduction

We consider interference minimization problem in energy limited wireless networks (wireless sensor and ad-hoc networks) where recharging or changing the energy sources of nodes is not feasible and sometimes due to environmental conditions not possible. In such networks it is important to consider the minimization of energy consumption of algorithms running on network nodes. By decreasing energy consumption we increase nodes operability time and as a result networks' lifetime. In different wireless sensor network (WSN) applications definition of networks' lifetime may be different (till all the nodes are alive, network is connected, given area is monitored by alive nodes, etc). In current work we tend to decrease energy consumption of nodes by decreasing the maximum interference of network algorithmically. Wireless communication of two nodes which is experiencing the third one is called interference. High interference on a receiver node (high value of accumulated signal straights on a node) makes difficulty to determine and accept the signals dedicated to it, this makes necessity for sender node to retransmit the signal until it is successfully accepted by receiver node, which is extra energy consumption and should be avoided.

Interference Minimization in Disk Graph Model of Wireless Networks

Consider a set of spatially distributed nodes, where each node equipped with radio transmitter/receiver and the power of nodes' transmitter is adjustable between zero and nodes' maximum transmission level. In disk graph model of network assumed that by fixing a transmission power for a node we define a transmission radius/disk of a node, i.e. the transmitted signal is reachable and uniform in any point of transmission disk of node and is zero outside of it. In this model two nodes considered connected if they are covered by each others transmission disks and interference on a given node defined as the number of transmission disks including that node. The overall interference of network is the maximum interference among all the nodes forming the network. The main weakness of disk graph model is the assumption that the radio coverage area is a perfect circle.

Assigning a transmission powers to a given set of spatially distributed nodes such that nodes form a connected network with assigned transmission powers while the interference of network is minimal called interference minimization problem in wireless networks.

One particular case of interference minimization problem described above is studied in [Rickenbach, 2005]. Authors considered the problem in one dimensional network, where all the nodes are distributed along the straight line, and named it a highway-model. For this model they showed that intuitive algorithm, which connects each node with its closest left and right nodes, can give a bad performance. An example of network where intuitive algorithm has worst performance is the exponential node chain, where distance between two consecutive nodes grows exponentially ($2^0, 2^1, \ldots, 2^{n-1}$). They also gave two algorithms for one dimensional case of interference minimization problem. The first algorithm, for a given set of distributed nodes, finds a connected network with at most $O(\sqrt{\Delta})$ interference where Δ is interference of uniform radius network under consideration and is $O(n)$ in some network instances. The second one is an approximation algorithm with $O(\sqrt[4]{\Delta})$ approximation ratio. By applying computational geometry and ε-net theory to ideas given in [Rickenbach, 2005], [Halldorsson, 2006] proposes a algorithm which gives $O(\sqrt{\Delta})$ interference bound for maximum interference in two and $O(\sqrt{\Delta log \Delta})$ for any constant dimensional network. Authors of [Aslanyan, 2010] give iterative algorithm based on linear program relaxation techniques which guaranties $O((opt \ln n)^2)$ interference bound for networks of n nodes, opt here is the optimal interference value for given instance of network. Logarithmic lower bound for interference minimization problem in disk graph model of networks under the general distance function is proven in [Bilo, 2006] by reducing minimum set cover to minimum interference problem.

Interference Minimization in Physical Model of Wireless Networks

Again, consider a set of spatially distributed wireless nodes, where each node has a radio transmitter/receiver with adjustable power level. In physical model of wireless networks we refuse the assumption that the signal coverage of a node is a perfect circle and assume that the signal straight on any given point (node) of network is a function of sender node, the node in question and the level of transmitted signal. In this model we are also given a constant β which is a signal acceptance threshold, i.e. it assumed that receiver node accepts the signal if it's straight is at least β. By this mean two nodes considered connected if their signals' straights are at least β on each other. Interference on a given node defined as a sum of signal straights on that node and interference of networks is the maximum interference among all the nodes forming the network.

The disk graph model can be deduced from physical model if we consider a signal straight function which for every node and its transmission level draws a disk and outputs a positive constant for every node within that disk and zero for the rest. Another example of signal straight function is $f(u,v,\xi) = \xi / d(u,v)^{\alpha}$ where u and v are sender and receiver nodes respectively, ξ is the transmission power of u, $\alpha \in [2,6]$ is the path lost exponent and $d(u,v)$ is the distance between nodes u and v [Pahlavan, 1995].

Interference minimization problem defined in a same way as for disk graph model.

Assign a transmission powers to a given set of spatially distributed wireless nodes such that nodes form a connected network with assigned transmission powers and the interference of network is minimal.

Our result is a deterministic polynomial time algorithm for interference minimization problem in wireless networks under the physical model of wireless networks in consideration, which for given network of n wireless nodes finds a connected network with at most $O((opt \ln n)^2/\beta)$ interference.

Formal Definitions

Consider a set V of n wireless nodes spatially distributed over a given area where nodes have adjustable transmission power and it can be fixed between zero and nodes' maximum transmission power. For any node $u \in V$ denote the range of feasible transmission powers by $R_u = [0, \xi_u^{max}]$, where ξ_u^{max} is the maximum transmission power for node u, and define a signal straight function $\phi_u : V \times R_u \to R^+$ where $\phi_u(v,\xi)$ is the signal straight of node u on node v when u uses the transmission power ξ. We assume that the signal straight function satisfies to following conditions

1. for any $\xi_1, \xi_2 \in R_u$, from $\xi_1 \geq \xi_2$ it follows that $\phi_u(v,\xi_1) \geq \phi_u(v,\xi_2)$
2. for given $\eta \in R^+$ it is easy to find a $\xi \in R_u$ (if exists) such that $\phi_u(v,\xi) = \eta$

Suppose that for any node u the suitable transmission power ξ_u is fixed, then any two nodes u and v considered connected if $\phi_u(v,\xi_u) \geq \beta$ and $\phi_v(u,\xi_v) \geq \beta$ where $\beta \geq 1$ is the signal acceptance threshold of network. Interference on a given node u is the accumulated signal straight of all the nodes forming the network $I(u) = \sum_{v \in V \setminus \{u\}} \phi_v(u,\xi_v)$ and $I(V) = \max_{v \in V} I(V)$ is the overall network interference. At this point interference minimization problem can be formulated as follows:

Given a spatially distributed set of wireless nodes, assign a suitable transmission power to each node such that the network is connected and the interference of network is minimal.

This is the formulation of interference minimization problem by transmission power assignment.

Consider a network graph $G = (V,E)$ where $E = \{(u,v) \mid u,v \in V, \phi_u(v,\xi_u^{max}) \geq \beta, \phi_v(u,\xi_v^{max}) \geq \beta\}$ i.e. in graph G two vertexes/nodes are incident if their maximum transmission powers are enough for communicating with each other. By this mean interference minimization problem is formulated as follows.

For a given network graph $G = (V,E)$ find a connected spanning subgraph $H = (V,E')$ such that the interference of network computed by the selected set of edges is minimal.

Formally, having the subgraph $H = (V,E')$ it is correct to further extract transmission power for any node u as a minimum power such that u can communicate with all of its neighbors in H, $\xi_u = min_{\xi}\{\xi \mid \phi_u(v,\xi) \geq \beta \text{ for all } v \text{ that } (u,v) \in E'\}$, which avoids unnecessary interference.

Set Covering and Interference Minimization

In the classical set cover problem a set S and a collection C of subsets of S are given, it is required to find a minimum size sub collection C' of C such that the union of sets of C' is S. In a decision version of set cover problem a positive integer k is given and the question is if it is possible to choose at most k subsets from collection C such that the union of chosen sets is S. It is well known that decision version of set cover problem is NP-complete and in polynomial time the optimal solution can not be approximated closer than with a logarithmic factor [Johnson, 1974]. Several variants of set cover problem have been studied [Kuhn, 2005; Garg, 2006; Demaine, 2006; Guo, 2006; Mecke, 2004; Ruf, 2004; Aslanyan, 2003].

Being motivated by interference minimization problem in cellular networks the minimum membership set cover (MMSC) problem has been investigated in [Kuhn, 2005]. In MMSC a set S and a collection C of subsets of S are given, it is required to find a subset C' of C such that the union of sets in C' is S and the maximum covered element of S is covered by as few as possible subsets from C'. In a decision version of MMSC problem a positive integer k is given and the question is if it is possible to choose a sub collection of C such that the union of chosen sets is S and each element of S is covered by at most k different subsets. [Kuhn, 2005] Contains the proofs of NP-completeness of decision version of MMSC problem and non-approximability of MMSC optimization problem by factor closer than $O(\ln n)$ unless $NP \subset TIME(n^{O(\log\log n)})$. Also, by using the linear program relaxation and randomized rounding techniques, [Kuhn, 2005] gives a polynomial time algorithm, which approximates the optimal solution of MMSC with logarithmic factor $O(\ln n)$.

Minimum partial membership partial set cover (MPMPSC) problem has been proposed in [Aslanyan, 2010] and used for developing interference minimization algorithm for wireless networks (disk graph model under consideration). In MPMPSC a set $S = S_1 \cup S_2$, consisting of two disjoint sets S_1 and S_2, along with collection C of subsets of S are given, it is required to find a sub collection C' of C such that the union of sets in C' contains all the elements of S_1 and the maximum covered element of S_2 is covered by as few as possible subsets from C'. In a decision version of MPMPSC problem a positive integer k is given and the question is if it is possible to choose a sub collection of C such that the union of chosen sets contains all the elements of S_1 and each element of S_2 is covered by at most k different subsets. It is known that the decision version of MPMPSC problem is NP-Complete and that the deterministic polynomial time algorithm exists which approximates the optimal solution of optimization version of MPMPSC by logarithmic factor $O(log(max\{|S_1|,|S_2|\}))$ which asymptotically matches the lower bound [Aslanyan, 2010]. The approximation algorithm for MPMPSC is achieved by applying the same techniques which has been applied in [Kuhn, 2005] for solving the MMSC.

Being motivated by interference minimization problem in physical model of wireless networks we consider a weighted minimum partial membership partial set cover (WMPMPSC) problem which is a generalization of MPMPSC. In WMPMPSC a set $S = S_1 \cup S_2$, consisting of two disjoint sets S_1 and S_2, along with collection C of subsets of S are given. In each subset from C the elements of S_2 have weights in $[0,1]$. The same element of S_2 may have a different weights in different sets of C. It is required to find a sub collection C' of C such that the union of sets in C' contains all the elements of S_1 and the accumulated, among the subsets of C', weight of a node which has the maximum accumulated weight, is as small as possible. In a decision version of WMPMPSC problem a positive number k is given and the question is if it is possible to choose a sub collection of C such that the union of chosen sets contains all the elements of S_1 and the accumulated, among the chosen sets, weight of each node is at most k. It is easy to see that in WMPMPSC we get a instance of MPMPSC when each node has a weight 1 in all the sets of C. This last statement proves the NP-Completeness of the decision version of WMPMPSC and the logarithmic lower bound for optimization version of the problem.

LP Formulations

Let C' denote a subset of the collection C. To each subset $C_j \in C$ we assign a variable $x_j \in \{0,1\}$ such that $x_j = 1 \Leftrightarrow C_j \in C'$. For C' to be a set cover for S, it is required that for each element $u \in S$ at least one set C_j with $u \in C_j$ is in C'. Therefore, C' is a set cover for S if and only if for all $u \in S$ it holds that $\sum_{C_j \ni u} x_j \geq 1$. Let z is the maximum membership over all the elements caused by the sets in C'. Then for all $u \in S$ it follows that $\sum_{C_j \ni u} x_j \leq z$. Then the integer linear program IP_{MMSC} of MMSC problem can be formulated as:

$$\begin{aligned}
&\text{minimize} && z && \\
&\text{subject to} && \sum_{C_j \ni u} x_j \geq 1, && u \in S && (1)\\
& && \sum_{C_j \ni u} x_j \leq z, && u \in S && (2)\\
& && x_j \in \{0,1\}, && C_j \in C && (3)
\end{aligned}$$

Integer linear program IP_{MPMPSC} of MPMPSC would be:

$$\begin{aligned}
&\text{minimize} && z && \\
&\text{subject to} && \sum_{C_j \ni u} x_j \geq 1, && u \in S_1 && (4)\\
& && \sum_{C_j \ni u} x_j \leq z, && u \in S_2 && (5)\\
& && x_j \in \{0,1\}, && C_j \in C && (6)
\end{aligned}$$

After introducing the weight function $w: C \times S_2 \to [0,1]$, where $w(C_j, u)$ is the weight of u in subset C_j, the integer linear program $IP_{WMPMPSC}$ of WMPMPSC can be formulated as:

$$\begin{aligned}
&\text{minimize} && z && \\
&\text{subject to} && \sum_{C_j \ni u} x_j \geq 1, && u \in S_1 && (7)\\
& && \sum_{C_j \ni u} x_j w(C_j, u) \leq z, && u \in S_2 && (8)\\
& && x_j \in \{0,1\}, && C_j \in C && (9)
\end{aligned}$$

By applying randomized rounding technique to IP_{MMSC} with relaxation of constraints (3), [Kuhn, 2005] gives a deterministic polynomial time approximation algorithm with $(1+O(1/\sqrt{z'}))(\ln(n)+1)$ approximation ratio for MMSC problem, where z' is the optimal solution for IP_{MMSC} relaxation. Later on [Aslanyan, 2010] states that by applying the same randomized rounding technique to IP_{MPMPSC} with relaxation of constraints (6) gives a deterministic polynomial time approximation algorithm with $(1+O(1/\sqrt{z'}))(\ln(\max\{|S_1|,|S_2|\})+1)$ approximation ratio for MPMPSC problem, where z' is the optimal solution for IP_{MPMPSC} relaxation. In current work we state that the same randomized rounding technique can be applied to $IP_{WMPMPSC}$ with relaxation of constraints (9) to achieve a deterministic polynomial time approximation algorithm with $(1+O(1/\sqrt{z'}))(\ln(\max\{|S_1|,|S_2|\})+1)$ approximation ratio for WMPMPSC problem, where z' is the optimal solution for $IP_{WMPMPSC}$ relaxation. The proof of the last statement is presented in the Appendix of this work. To sum up, we have the following theorem.

Theorem 1. *For WMPMPSC problem, there exists a deterministic polynomial-time approximation algorithm with an approximation ratio of* $O(log(max\{|S_1|,|S_2|\}))$ [1]

Approximation Algorithm for Interference Minimization in Physical Model of Wireless Networks

Algorithm takes a network graph $G=(V,E)$ with n vertices as an input and after logarithmic number of $k \in O(\log n)$ iterations returns connected subgraph $G_k \subseteq G$ where interference of network corresponding to the graph G_k is bounded by $O((opt \cdot \ln n)^2/\beta)$, where $n=|V|$ is the number of network nodes and opt is the interference of minimum interference connected network.

Algorithm starts the work with the graph $G_0=(V,E_0)$ where $E_0=\varnothing$. On the l^{th} iteration, $l \geq 1$, algorithm chooses a subset $F_l \subseteq E \setminus E_{l-1}$ of new edges and adds them to already chosen edge set $E_{l-1}=\cup_{i=1}^{l-1}F_i$. As a consequence of such enlargement of edge set, interference on graph vertices may increase in some value depending on F_l. Algorithm finishes the work if the graph $G_l=(V,E_l)$ is connected otherwise goes for the next iteration. Below we present how algorithm chooses the set of edges $F_l \subseteq E \setminus E_{l-1}$ on the l^{th} iteration. Algorithms' quality, i.e the final maximal interference on nodes (its upper estimate) is bounded by the accumulated through the iterations interferences which we try to keep minimal. Let $G_{l-1}=(V,E_{l-1})$ is the graph obtained after the $(l-1)^{th}$ iteration, and has the set of connected components $C(G_{l-1})=\{C_{l-1}^1,\ldots,C_{l-1}^{k_{l-1}}\}$. Denote by $H_{l-1} \subseteq E \setminus E_{l-1}$ the set of all edges which have their endpoints in different connected components of G_{l-1}. On the l^{th} stage of algorithm a subset of H_{l-1} is selected to further reduce the number of connected components which finally brings us to a connected subgraph. In this way we build the collection $T(C(G_{l-1}),H_{l-1})$ of special sets as follows. Starting with H_{l-1} we add to the set $T(C(G_{l-1}),H_{l-1})$ of l^{th} stage specific weighted subsets

[1]See the Appendix A for the proof.

$T^l(u,v) = \{C_{l-1}^u, C_{l-1}^v\} \cup V$ defined by all $(u,v) \in H_{l-1}$, where u belongs to connected component C_{l-1}^u and v belongs to C_{l-1}^v. By selection of u and v we have that C_{l-1}^u and C_{l-1}^v are different. By definition of connectivity nodes u and v can communicate with each other if their signal transmission powers ξ_{uv} and ξ_{vu} satisfy to $\phi_u(v,\xi_{uv}) \geq \beta$ and $\phi_v(u,\xi_{vu}) \geq \beta$, where β is the signal acceptance threshold. To avoid unnecessary energy consumption and to reduce interference it would be right to adjust transmission powers ξ_{uv} and ξ_{vu} such that $\phi_u(v,\xi_{uv}) = \beta$ and $\phi_v(u,\xi_{vu}) = \beta$, this is possible to do because of the second property of the signal straight function ϕ. Then the noise of the link (u,v) on any node t can be calculated as $w((u,v),t) = \phi_u(t,\xi_{uv}) + \phi_v(t,\xi_{vu})$ which would be the weight $w(T^l(u,v),t)$ of node t in the subset $T^l(u,v)$. And so $T^l(u,v)$ is a composite set which includes two labels for components C_{l-1}^u and C_{l-1}^v and all the vertices in V along with the weights, which are the interference increase on nodes if the edge (u,v) is selected as a communication link. In terms of WMPMPSC the labels of connected components will compose the set S_1 and weighted V will be the set S_2.

Figure 1 demonstrates connected components that are input to the stage l, and the set H_{l-1} of all cross component edges.

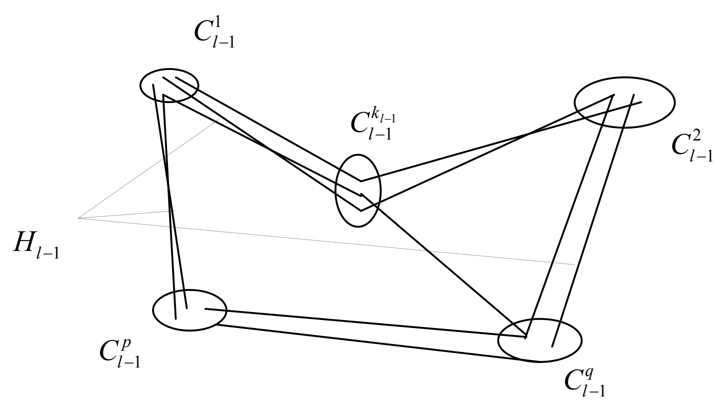


Figure 1: Connected components that are input to the l-th stage of the algorithm

After constructing $T(C(G_{l-1}), H_{l-1})$ we normalize the weights of elements by dividing all the weights by the maximum weight $w_{max} = \max_{t,(u,v)\in H_{l-1}} w((u,v),t)$ and solve the WMPMPSC on the set $C(G_{l-1}) \cup V$ and collection of subsets $T(C(G_{l-1}), H_{l-1})$, where condition for elements of $C(G_{l-1})$ is *to be covered* and for elements of V is *to have minimum accumulated weight*. Finally, based on the solution $W(C(G_{l-1}), H_{l-1}) \subseteq T(C(G_{l-1}), H_{l-1})$ of WMPMPSC we build the set F_l of network graph edges, selected at the l^{th} iteration of algorithm by adding to F_l all the edges $(u,v) \in H_{l-1}$ such that $T^l(u,v) \in W(C(G_{l-1}), H_{l-1})$ and multiply all the weights by w_{max} to receive the real interference increase.

Algorithm performance

Theorem 2. *On each iteration of algorithm the number of connected components is being reduced at least by factor of two, which bounds the total number of iterations by* $O(\log n)$.

Proof. For each connected component $C_{l-1}^{u} \in C(G_{l-1})$ of graph G_{l-1} the solution $W(C(G_{l-1}), H_{l-1})$ of WMPMPSC solved at l^{th} iteration contains at least one set $T^{l}(u,v) \in W(C(G_{l-1}), H_{l-1})$ such that $C_{l-1}^{u} \in T^{l}(u,v)$ (as $W(C(G_{l-1}), H_{l-1})$ is a cover for the set $C(G_{l-1})$). And as each set $T^{l}(u,v) \in W(C(G_{l-1}), H_{l-1})$ contains exactly two connected components, then by adding the edge (u,v) to our solution, we merge those two connected components into one (connecting by the edge (u,v)). So every connected component merges with at least one other component, which reduces the number of connected components at least by factor of 2.

Lemma 1. *Network corresponding to the graph* $G^{l} = (V, F_{l})$, *where* F_{l} *is the edge set obtained on the* l^{th} *iteration of algorithm, has interference in* $O((opt^{2} \cdot \ln n)/\beta)$.

Proof. Consider the set of connected components $C(G_{l-1}) = \{C_{l-1}^{1}, \ldots, C_{l-1}^{k_{l-1}}\}$ of l^{th} iterative step of algorithm. Let E_{opt} is the set of the edges of some interference optimal connected network for our problem (edges of connected network with optimal interference opt). Then there is a subset $E_{opt}^{l} \subseteq E_{opt}$ which spans connected components $C(G_{l-1})$ and the network of the graph $G_{opt}^{l} = (V, E_{opt}^{l})$ has interference not exceeding the opt.

Fact 1. *The maximal vertex interference due to the spanner* E_{opt}^{l} *of* $C(G_{l-1})$ *is at most* opt.

Now let us build the set collection $T_{opt}(C(G_{l-1}), E_{opt}^{l}) = \{T^{l}(u,v)/(u,v) \in E_{opt}^{l}\}$.

Fact 2. $T_{opt}(C(G_{l-1}), E_{opt}^{l})$ *is a sub collection of* $T(C(G_{l-1}), H_{l-1})$ *built on the* l^{th} *iteration of algorithm and is a cover for* $C(G_{l-1})$, *i.e.* $T_{opt}(C(G_{l-1}), E_{opt}^{l})$ *is a solution for the WMPMPSC problem, with some value* z^{*}, *solved on the* l^{th} *iteration of algorithm, not necessary optimal.* Now consider the matrix P_{opt}^{w} representing the transmission signals on some node w caused by communication links of E_{opt}^{l}.

$$P_{opt}^{w} = \begin{pmatrix} P_{u_1u_1}^{w} & P_{u_1u_2}^{w} & \cdots & P_{u_1u_j}^{w} & \cdots & P_{u_1u_n}^{w} \\ P_{u_2u_1}^{w} & P_{u_2u_2}^{w} & \cdots & P_{u_2u_j}^{w} & \cdots & P_{u_2u_n}^{w} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \\ P_{u_iu_1}^{w} & P_{u_iu_2}^{w} & \cdots & P_{u_iu_j}^{w} & \cdots & P_{u_iu_n}^{w} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \\ P_{u_nu_1}^{w} & P_{u_nu_2}^{w} & \cdots & P_{u_nu_j}^{w} & \cdots & P_{u_nu_n}^{w} \end{pmatrix}$$

where

$$P^{w}_{u_i u_j} = \begin{cases} 0, & \text{if } i = j \text{ or } (u_i, u_j) \notin E^{l}_{opt} \\ \phi_{u_i}(w, \xi_{u_i u_j}), & \text{otherwise} \end{cases}$$

is the signal straight of node u_i on node w when u_i uses the transmission power $\xi_{u_i u_j}$ (communicates with node u_j).

Fact 3. *and the sum of the matrix elements will give the interference increase we count (the real interference increase is the sum of the maximal elements from each row) on node w by edge set E^{l}_{opt}. Due to the Fact 1 and signal acceptance threshold β for any vertex u_i the number of sets $T^{l}(u_i, v) \in T_{opt}(C(G_{l-1}), E^{l}_{opt}, w)$ will not exceed the $\lfloor opt/\beta \rfloor$, in other words the number of non zero elements on each row of matrix P^{w}_{opt} is bounded by $\lfloor opt/\beta \rfloor$.*

Fact 4. *The interference increase on node w by the edge set E^{l}_{opt} can be calculated as $\sum_{i=1}^{n} \max_j P^{w}_{u_i u_j}$ and due to the Fact 1 it doesn't exceed the opt.*

From facts 3 and 4 it follows that the sum of the matrix elements is bounded by opt^2/β, which means that the optimal value of WMPMPSC problem solved on the l^{th} iteration of algorithm is bounded by opt^2/β and therefor by Theorem 1 the interference increase by the edge set F_l is bounded by $O(opt^2 \cdot \ln n/\beta)$.

Theorem 3. *The network built by WMPMPSC relaxation algorithm has at most $O((opt^2 \cdot \ln^2 n)/\beta)$ interference.*

Proof. The proof is in combination of Theorem 2 and Lemma 1.

Conclusion and Future Work

In current work we considered the interference minimization problem in physical model of wireless networks and proposed a polynomial time approximation algorithm which for a given set of wireless nodes creates a connected network with at most $O((opt \cdot \ln n)^2/\beta)$ interference. In some WSN applications network considered as functional while it is connected, therefore in future works on interference minimization the k-connectivity of network should be considered. Also considering the problem in Euclidean spaces, which is a realistic case for WSNs, may give a better approximation ratio.

References

[Shannon, 1949] C.E.Shannon. The Mathematical theory of communication. In: The Mathematical Theory of Communication. Ed. C.E.Shannon and W.Weaver. University of Illinois Press, Urbana, 1949.

[Aslanyan, 2003]H. Aslanyan. Greedy Set Cover Estimations. Computer Science and Information Technologies (CSIT), :143--144, 2003.

[Aslanyan, 2010] H. Aslanyan and J. Rolim. Interference Minimization in Wireless Networks. IEEE/IFIP International Conference on Embedded and Ubiquitous Computing (EUC 2010), :444--449, 2010.

[Bilo, 2006] D. Bilo and G. Proietti. On the complexity of minimizing interference in ad-hoc and sensor networks. 2nd International Workshop on Algorithmic Aspects of Wireless Sensor Networks, ALGOSENSORS, LNCS, 4240:13--24, 2006.

[Demaine, 2006] E.D. Demaine and U. Feige and M.T. Hajiaghayi and M. R. Salavatipour. Combination can be hard: Approximability of the unique coverage problem. Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), :162--171, 2006.

[Garg, 1997] N. Garg and V.V. Vazirani and M. Yannakakis. Primal-dual approximation algorithms for integral flow and multicut in trees. Algorithmica 18:3-20, 1997.

[Guo, 2006] J. Guo and R. Niedermeier. Exact algorithms and applications for Tree-like Weighted Set Cover. Discrete Algorithms, 4(4):608--622, 2006.

[Halldorsson, 2006] M. M. Halldorsson and T. Tokuyama. Minimizing Interference of a Wireless Ad-Hoc Network in a Plane. Algorithmic Aspects of Wireless Sensor Networks, ALGOSENSORS, 4240/2006:71--82, 2006.

[Johnson, 1974] D. Johnson. Approximation algorithms for combinatorial problems. Journal of Computer and System Sciences, 9:256--278, 1974.

[Kuhn, 2005] F. Kuhn and P. von Rickenbach and R. Wattenhofer and E. Welzl and A. Zollinger. Interference in cellular networks: the minimum membership set cover proble. 11th International Computing and Combinatorics Conference, COCOON, LNCS, 3595:188--198, 2005.

[Mecke, 2004] S. Mecke and D. Wagner. Solving Geometric Covering Problems by Data Reduction. 12th ESA, LNCS, 3221:760--771, 2004.

[Pahlavan, 1995] K. Pahlavan and A. H. Levesque. Wireless information networks. Wiley-Interscience, 1995.

[Rickenbach, 2005] P. von Rickenbach and S. Schmid and R. Wattenhofer and A. Zollinger. A robust interference model for wireless ad-hoc networks. 5th International Workshop on Algorithms for Wireless, Mobile, Ad Hoc and Sensor Networks (WMAN), 2005.

[Ruf, 2004] N. Ruf and A. Schobel. Set covering with almost consecutive ones property. Discrete Optimization, 1(2):215--228, 2004.

Appendix A

Here we show how randomized rounding technique used in [Kuhn, 2005] for solving the IP_{MMSC} can be used for solving $IP_{WMPMPSC}$. This section mostly presents the work of [Kuhn, 2005].

Consider a instance $(S = S_1 \cup S_2, C, w)$ of $IP_{WMPMPSC}$ and the solution vector $\underline{x}'$ and $\underline{z}'$ of $LP_{WMPMPSC}$ relaxation of $IP_{WMPMPSC}$. Consider the following randomized rounding scheme, where an integer solution $\bar{x} \in 0,1^m$ is computed by setting

$$x_i = \begin{cases} 1, & \text{with probability } p_i := \min\{1, \alpha x_i'\} \\ 0, & \text{otherwise} \end{cases}$$

independently for each $i \in \{1,...,n\}$. Let A_i be the "bad" event that the i^{th} element is not covered.

Lemma A1. *The probability that the i^{th} element remains uncovered is*

$$P(A_i) = \prod_{C_j \ni u_i} (1 - p_j) < e^{-\alpha}$$

Proof. The proof is in Lemma 1 of [Kuhn, 2005].

Let B_i be the “bad” event that the weight of the i^{th} element is more than $\alpha\beta z'$ for some $\beta \geq 1$.

Lemma A2. *The probability that the weight of the i^{th} element is more than $\alpha\beta z'$ is*

$$P(B_i) < \frac{1}{\beta^{\alpha\beta z'}} \cdot \prod_{C_j \ni u_i} (1 + (\beta^{w(j,i)} - 1) p_j) \leq \left(\frac{e^{\beta-1}}{\beta^{\beta}} \right)^{\alpha z'}$$

Proof. We use a Chernoff-type argument. For $t = \ln \beta > 0$, we have

$$P(B_i) = P\left(\sum_{C_j \ni u_i} x_j w(j,i) > \alpha\beta z' \right) = P\left(e^{t \cdot \sum_{C_j \ni u_i} x_j w(j,i)} > e^{t \cdot \alpha\beta z'} \right)$$

$$< \frac{E\left[e^{t \cdot \sum_{C_j \ni u_i} x_j w(j,i)} \right]}{e^{t \cdot \alpha\beta z'}} = \frac{1}{e^{t \cdot \alpha\beta z'}} \cdot \prod_{C_j \ni u_i} \left[p_j e^{t \cdot w(j,i)} + 1 - p_j \right]$$

$$= \frac{1}{\beta^{\alpha\beta z'}} \cdot \prod_{C_j \ni s_i} \left[1 + (\beta^{w(j,i)} - 1) p_j \right] \leq \frac{1}{\beta^{\alpha\beta z'}} \cdot \prod_{C_j \ni s_i} e^{p_j (\beta^{w(j,i)} - 1)}$$

$$\leq \frac{1}{\beta^{\alpha\beta z'}} \cdot \prod_{C_j \ni s_i} e^{p_j (\beta - 1) w(j,i)} = \frac{1}{\beta^{\alpha\beta z'}} \cdot e^{(\beta-1) \cdot \sum_{C_j \ni u_i} p_j w(j,i)} \leq \left(\frac{e^{(\beta-1)}}{\beta^{\beta}} \right)^{\alpha z'}$$

The inequality and equality in the second line results by application of the Markov inequality and because of the independence of the x_j. The equality and inequality in the third line hold because $t = \ln \beta$ and $1 + x \leq e^x$. For the inequalities in the last line we apply $\beta^x - 1 \leq (\beta - 1)x$ for $\beta \geq 1$, $x \in [0,1]$ and $\sum_{C_j \ni u_i} p_j w(j,i) \leq \alpha z'$.

Denote the probability upper bounds given by Lemmas $A1$ and $A2$ by $\overline{A}_i$ and $\overline{B}_i$:

$$\overline{A}_i := \prod_{C_j \ni s_i} (1 - p_j) \quad \text{and} \quad \overline{B}_i := \frac{1}{\beta^{\alpha\beta z'}} \cdot \prod_{C_j \ni u_i} (1 + (\beta^{w(j,i)} - 1) p_j).$$

In order to bound the probability for any “bad” event to occur, we define a function P as follows

$$P(p_1, \ldots, p_m) := 2 - \prod_{i=1}^{n} (1 - \overline{A}_i) - \prod_{i=1}^{n} (1 - \overline{B}_i).$$

Lemma A3. *The probability that any element is not covered or has a weight more than $\alpha\beta z'$ is upper-bounded by $P(p_1, \ldots, p_m)$:*

$$P\left(\bigcup_{i=1}^{n} A_i \cup \bigcup_{i=1}^{n} B_i \right) < P(p_1, \ldots, p_m).$$

Proof. The proof is in Lemma 3 of [Kuhn, 2005].

The following shows that if α and β are chosen appropriately, $P(p_1,\ldots,p_m)$ is always less than 1.

Lemma A4. *When setting* $\alpha = \ln(\max\{|S_1|,|S_2|\})+1$, *then for* $\beta = 1+\max\{\sqrt{3/z'}, 3/z'\}$, *we have* $P(p_1,\ldots,p_m) < 4/5$.

Proof. The proof is in Lemma 4 of [Kuhn, 2005].

Lemmas A1 – A4 lead to the following randomized algorithm for the WMPMPSC problem. As a first step, the linear program $LP_{WMPMPSC}$ has to be solved. Then, all x_i' are rounded to integer values $x_i \in \{0,1\}$ using the described randomized rounding scheme with $\alpha = \ln(\max\{|S_1|,|S_2|\})+1$. The rounding is repeated until the solution is feasible (all elements are covered) and the weight of the integer solution deviates from the fractional weight z' by at most a factor $\alpha\beta$ for $\beta = 1+\max\{\sqrt{3/z'}, 3/z'\}$. Each time, the probability to be successful is at least $1/5$ and therefore, the probability of not being successful decreases exponentially in the number of trials.

We will now show that $P(p_1,\ldots,p_m)$ is a pessimistic estimator and that therefore, the algorithm described above can be derandomized. That is, P is an upper bound on the probability of obtaining a "bad" solution, $P<1$ (P is a probabilistic proof that a "good" solution exists), and the p_i can be set to 0 or 1 without increasing P . The first two properties follow by Lemmas A3 and A4, the third property is shown by the following lemma.

Lemma A5.

For all i , either setting p_i to 0 or setting p_i to 1 does not increase P :

$$P(p_1,\ldots,p_m) \ge \min\{P(\ldots,p_{i-1},0,p_{i+1},\ldots), P(\ldots,p_{i-1},1,p_{i+1},\ldots)\}$$

Proof. The proof is in Lemma 5 of [Kuhn, 2005].

Lemmas A3, A4 and A5 lead to an efficient deterministic approximation algorithm for the WMPMPSC problem. First, the linear program $LP_{WMPMPSC}$ has to be solved. The probabilities p_i are determined as described above. For α and β as in Lemma A4, $P(p_1,\ldots,p_m) < 4/5$. The probabilities p_i are now set to 0 or 1 such that $P(p_1,\ldots,p_m)$ remains smaller than $4/5$. This is possible by Lemma A5. When all $p_i \in \{0,1\}$, we have an integer solution for $IP_{WMPMPSC}$. The probability that not all elements are covered or that the weight is larger than $\alpha\beta z'$ is smaller than $P < 4/5$. Because all p_i are 0 or 1 , this probability must be 0 . Hence, the computed $IP_{WMPMPSC}$ -solution is an $\alpha\beta$ -approximation for WMPMPSC.

Authors' Information

Hakob Aslanyan – PhD student and research assistant; Computer Science Department, University of Geneva, Battelle Batiment A, route de Drize 7, 1227 Geneva, Switzerland; e-mail: hakob.aslanyan@unige.ch

Major Fields of Scientific Research: combinatorial optimization, graph theory, approximation and exact algorithms, hardness of approximation, network design and connectivity

ON MEASURABLE MODELS OF PROMOTION OF NEGENTROPIC STRATEGIES BY COGNITION

Pogossian Edward

Abstract: Could models of mind be independent from living realities but be classified as mind if the mind uses the same criteria to form the class mind? In the paper a constructive view on the models of mind, cognizers, is presented and the measurable criteria and schemes of experiments on mentality of cognizers are discussed.

Keywords: modeling, cognition, measures, negentropic, strategies.

ACM Classification Keywords: A.0 General Literature.

1. Introduction

Due mind forms models of any realities including itself raises the question whether models of mind can be mental not being living realties (LR), assembled from LR or developed from the springs of LR?

In other words, whether are models of mind which do not depend from LR but are classified as mind possible if mind uses the same criteria when forms the class mind?

To answer the question constructive models of mind and criteria of measuring their mentality as well as the exhaustive experiments on revealing the truth are needed.

In what follows a measurable approach to the models of mind, cognizers, is presented and the criteria and experiments of testing of mentality of cognizers are questioned.

This approach to refining of cogs continues the approach started in [Pogossian,1983] and continued in [Pogossian,2005,2007] on interpretation of the recognized views on mind [Flavell,1962,Neuman,1966, Botvinnik,1984, Atkinson1993, Pylyshin,2004, Roy,2005, Winograd, 986,Mendler,2004,] by models having unanimous communalized meanings followed by experiments on validity of those models.

The paper describes the author’s view on mental behavior and traditionally we should address to the readers by using words "our view" ,"we think», etc.

On the other hand, mental behavior, we assume, is identified with ourselves and we plan to discuss personalized and communalized constituents in communications.

That explains why we find possible in the paper to use the pronoun "I" for the mind along with "we" and "our" when they seem to be appropriate

2. A View on Mind

2.1. I am a *mind* and I am able to interpret, or *model* the *realities* I *perceive,* including myself, evaluate the quality, or *validity* of models and use those models to promote my *utilities*.

The models are composed from cause-effect relationships between realities, particularly between realities and utilities, and any composition of those relationships comprise the *meanings* of the realities.

The basic, or *nucleus* utilities and meanings are inborn while mind incrementally enriches them by assimilating and accommodating by Piaget [Flavel,1962, Mandler, 2004] cause-effect relationships between realities and already known utilities and meanings *solving* corresponding *tasks* and *problems* .

By Piaget "Mind neither starts with cognition of itself nor with cognition of the meanings of realities but cognizes their interactions and expanding to those two poles of interactions mind organizes itself organizing the world" [Flavell,1962].

As much coincide ontology, or *communalized* (vs. *personalized*) meanings of realities with meanings of their models and as much those meanings are *operational*, i.e. allow to reproduce realities having equal with the models meanings, so better is the validity of the models.

In what follows a personalized model of mind, a view W, and a communalized version of W , *cognizers,* are presented with discussion of the validity of cognizers and schemas to meet the requirements.

2.2.1. Minds are algorithms for promoting by certain *effectors* the utilities of *living realities* (LR) in their games against or with other *players* of those games.

The players can be LR, assembles of LR like communities of humans or populations of animals as well as can be some realities that become players because not voluntarily but they affect LR inducing games with environments or the units like programs or devices that have to be tested and response to the actions of engineers . To compare and discuss some hypothetic mental realities like Cosmic Mind by Buddhists and Solaris by Stanislaw Lem are considered as players as well. Note, that descriptions of religious spiritual creatures resemble algorithm ones.

2.2.2. A variety of economic, military, etc. games can be processed by players. But all LR in different ways play the main negentropic games against overall increase of the entropy in the universe [Shrodinger,1956].

In those negentropic games with the environments LR and their populations realize some versified *reproduction* and on-the-job selection *strategy elaboration algorithms* (r SEA).

The parent rSEA periodically generates springs of LR where each child of the springs realizes some particular strategy of survival of those children in on going environments. LR with successful survival strategies get the chance to give a new spring and continue the survival games realizing some versions of strategies of their parents while unsuccessful LR die.

2.2.3. The utilities of LR and their assembles initially are determined by their nucleus, basic interests in the games but can be expanded by new mental constructions promoting already known utilities. For example, the nucleus utilities of LR, in general, include the codes (genetic) of rSEA and algorithms for reconstructing rSEA using their genetic codes.

2.2.4. The periods of reproduction, the power of the springs and other characteristics of rSEA are kinds of means to enhance survival abilities of LR and vary for different LR depending, particularly, from the resources of energy available to LR and the velocity of changes of the environments of LR.

2.2.5. Minds can be interpreted as one of means to enhance the survival of LR. In fact, minds realize SEA but in contrast to on-the-job performance rSEA the strategies elaborated by minds are auxiliary relatively to rSEA and are selected by a priory modeling.

Correspondingly, the nucleus of mental LR in addition to rSEA codes include codes of mind developing algorithms like the adaptation algorithms by Piaget [Flavel,1962, Mandler, 2004].

2.3. Thus, *modeling* SEA, or mSEA, do, particularly, the following:

- form the models of games and their constituents
- classify models to form classes and other mental constructions
- use mental constructions for a priori selection the most prospective strategies for the players
- elaborate instructions for the effectors of players using the prospective strategies.

The effectors transform the instructions into external and internal *actions* and apply to the environments of mSEA and mSEA themselves, correspondingly, for developing the environments and mSEA and enhancing the success of the players.

2.4. Whether are the models of mind which are not dependent from LR but are classified as mind possible if mind uses the same criteria when forms the class *mind*?

To answer to the question constructive models of mind and criteria of measuring their mentality as well as the exhaustive experiments on revealing the truth are needed.

2.5. Let's name *cognizers* the models of mind not depending from LR while the models of mental constructions name *mentals*.

Apparently, this ongoing view W on mind is a kind of cognizers, say, for certainty, *1-cognizers, 1cogs* or *cogs* in this paper.

In what follows a constructive approach to cogs, the criteria and experiments of testing of mentality of cogs are presented.

3. Basic Approaches and Assumptions

3.1. Further refining of cogs extends the approach described above on interpretation of the recognized views on mind by models having unanimous communalized meanings followed by experiments on validity of those models to mind.

3.2.1. Later on it is assumed that cogs are object-oriented programs, say in Java.

All programs in Java are either classes or sets of classes.

Therefore, it is worth to accept that cogs and their constituents, mentals, are either Java classes or their compositions as well.

3.3. Accepting the above stated assumption the experiments on quality of cogs were run for SSRGT games.

Particularly, because chess represents the class and by variety of reasons is recognized as a regular environment to estimate models of mind [Botvinnik,1984, Pogossian,1983,2007, Atkinson,1993, Furnkranz, 2001] in what follows the constructions of mentals and experiments on mentality of cogs are accompanied, as a rule, by interpretations in chess.

3.4. Following to the view W cogs elaborate instructions for the effectors of players to promote their utilities. The effectors in turn transform instructions into *actions* applied to the players and their *environments.* They can be parts of the players or be constructed by cogs in their work.

It is assumed that certain *nucleus* mentals of cogs as well as the players and their effectors are predetermined and process in discrete time intervals while mentals of cogs can evolve in time.

The fundamental question on the origin of nucleus mentals and other structures needs further profound examination.

4. Refining Constituents of Cognizers

4.1.1. In general, *percepts* are the inputs of cogs and have the structure of bundles of instances of the classes of cogs composed in discrete time intervals.

The *realities* of cogs are refined as the causes of their *percepts*.

The *environments* and the *universe* of cogs are the sets and the totality of all *realities of cogs*, correspondingly.

More in details, the bundles of instances of attributes of a class X of cogs at time t are named *X percepts at t* and the causes of X/t percepts are named *X/t realities.*

It is worth to consider *t percepts* and *percepts* as the elements of the unions of X/t percepts and t percepts, correspondingly, and assume that there may be multiple causes for the same percept.

Analogically, *t realities* and *X/t realities* are defined.

In case percepts are bundles of instances of attributes of certain classes of cogs the realities causing them are the classes represented by those attributes.

Otherwise, cogs learn about the realities by means of the percepts corresponded to realities and by means of the responses of those percepts when cogs arrange actions by effectors.

Due cogs are continuously developed they start with percepts formed by nucleus classes followed by percepts formed by the union of new constructed and nucleus classes.

4.1.2. Cogs promote utilities by using links between utilities and percepts. They continuously memorize percepts, by certain criteria unite them in classes as *concepts* and distinguish realities to operate with them using *matching* methods associated with the concepts.

In addition some concepts are nominated by *communicators* to communicate about the realities of the domains of the concepts with other cogs or minds and enhance the effectiveness of operations of cogs in the environments.

4.2.1. The base criteria to unite percepts in concepts are *cause-effect relationships* (cers) between percepts, particularly, between percepts and utilities.

For revealing cers cogs form and solve *tasks* and *problems.*

Tasks are requirements to link given percepts (or realities) by certain cers and represent those cers in frame of certain classes.

4.2.2. The basic tasks are the *utility* tasks requiring for given percepts to find utilities that by some cers can be achieved from the percepts. In chess utility tasks require to search strategies for enhancing the chances to win from given positions.

The *generalization*, or *classification* tasks unite percepts (as well as some classes) with similar values into more advanced by some criteria classes and associate corresponding matching procedures with those classes to distinguish the percepts of the classes and causing them realities.

The *acquisition* tasks create new classes of cogs by transferring ready to use classes from other cogs or minds while the *inference* tasks infer by some general rules new classes as consequences of already known to cogs classes.

The *question* tasks can be considered as a kind of formation tasks inference tasks which induce new tasks applying syntax rules of question tags to the solutions of already solved tasks.

The *modeling* tasks require revealing or constructing realities having certain similarities in *meanings* with the given ones.

Before refining meanings of realities let’s note that to help to solve the original tasks some approximating them model tasks can be corresponded.

4.2.3. *Problems* are compositions of homogeneous tasks and *solutions of problems* are procedures composing the solutions of constituent tasks.

The problems can be with *given spaces* of possible *solutions* (GSS) or without GSS, or the *discovery* ones.

Tasks formation and *tasks solving procedures* form and solve tasks types.

4.3.1. To refine the meanings of realities and mentals it is convenient to interpret the percepts, uniting them concepts, nucleus classes and the constituents of those mentals as the nodes of the *graph of mentals* (GM) while the edges of GM are determined by utility, cers, attributive, part of and other relationships between those nodes.

Then the *meaning of a percept C* can be defined as the union of the totality of realities causing C and the connectivity sub graph of GM with root in C.

The *meaning of a concept X* is defined as the union of the meanings of the nodes of the connectivity sub graph of GM with the root in X.

The *meaning of realities R c*ausing the percept C is the union of the meanings of the nodes of the connectivity sub graph of GM with the root in the percept C.

4.3.2*.* Later on it is assumed that the *knowledge* of cogs unites, particularly, the cogs, GM and their constituents.

4.4.1. Processing of percepts and concepts is going either *consciously* or *unconsciously.* While unconsciousness, usually, addresses to the *intuition* and needs the long way of research efforts for its explanation, the consciousness is associated with the named concepts and percepts in languages and their usage for communications. Particularly, the vocabularies of languages provide names of variety of concepts and realities causing those percepts.

Mind operates with percepts, concepts and other mentals while names realities causing those mentals when it should communicate.

Particularly, this ongoing description of cogs follows to the rules for named realities while internally refers to corresponding mentals.

4.4.2. When mind operates *internally* with the representations of realities it is always able to address to their meanings or *to ground* those representations [8].

For *external* communications mind uses representations of realities, *communicators,* which can be separated from the original carriers of the meanings of those realities, i.e. from the percepts of those realities, and become *ungrounded*.

The role of communicators is to trigger [12] the desired meanings in the partners of communications. Therefore, if partners are deprived of appropriate grounding of the communicators special arrangements are needed like the ones provided by ontologies. If the communicators are not sufficiently grounded well known difficulties like the ones in human-computer communications can rise.

Note, that if the model R` is a grounded reality the meaning of R' can induce new unknown aspects of the meaning of the original ones.

4.5. Realities R` *represent* realities R, or R` is a *model* of R, if meanings of R` and R intersect.

Model R` is *equal* to R if R' and R have the same meanings. The more is the intersection of the meanings of R and R` relative to the meaning of R the greater is the *validity* of R`. For measuring the validity of models a variety of aspects of the meanings of original realities can be emphasized. Particularly, descriptive or behavioral aspects of the meanings can be considered, or be questioned whether the meanings are views only of the common use or they are specifications.

5. Questioning Validity of Mind

5.1. Modeling problems require constructing realities having certain similarities in meanings with the original ones.

When those realities are problems as well cogs correspond model problems to the original ones, run them to find model solutions and interpret them back to solve the original ones.

Apparently, solutions of problems are the most valid models of those problems but, unfortunately, not always can be found in frame of available search resources.

Valid models trade off between the approximations of the meanings of solutions of problems and between available resources to choose the best available approximations.

Due of that inevitable trade off the models are forced to focus on only the particular aspects of those solutions.

If communication aspects are emphasized the *descriptive* models and criteria of validity can be in use require the realities-models be equal only by communicative means of the communities.

On-the-job or *behavioral* criteria evaluate validity of models by comparing the performances of corresponding procedures.

The records of computer programs provide examples of descriptive models while when processed programs become the subject of behavioral validity. Sorts of behavioral validity provide functional testing and question-answer ones like Turing test.

Productive behavioral validity criteria compare the results of affection of the outputs of realities and their models on the environment. Fun Newman requirement on self-reproducibility of automata [Neuman, 1966] provides an example of productive validity. In its interpretation as *reflexive reproducibility* (RR) validity that criterion requires to construct 1-models of realities able to produce 2-models equal to the 1-models and able to chain the process.

5.2. To formulate criteria of validity of cogs it is worth to summarize the refined to this end views on mind as the following:

mind is an algorithm to solve problems on promotion of utilities of LR in their negentropic games

mind is composed from certain constituent algorithms for forming and solving tasks of certain classes including the utility, classification, modeling, questioning classes

mind uses solutions of problems to elaborate instructions for certain effectors to make the strategies of LR more effective and the environments of LR more favorable to enhance the success of LR in negentropic games.

5.3. Criteria of validity of cogs to mind have to answer whether cogs have meanings that minds have about themselves.

On the long way in approaching to valid cogs a chain of inductive inferences is expected aimed to converge eventually to target validity.

Inductive inferences unite science with arts and, unfortunately, the term of their stabilization can not be determined algorithmically. Nevertheless, what can be done is to arrange those inferences with the trend to converge to the target stabilization in limit [19].

To approach to valid cogs it is worth to order the requirements to the validity of cogs and try to achieve them incrementally, step by step.

The requirements v1- v4 to validity of cogs condition them to meet the following:

v1. be well positioned relatively to known psychological models of mind

v2. be able to form and solve the utility, classification, modeling and question tasks with acceptable quality of the solutions

v.3. be able to use the solutions of tasks and enhance the success of the players

v4. be able to form acceptable models of themselves, or be able to *self modeling*

The requirements v2 - v4 follow the basic views on mind while v1 requires positioning cogs relatively, at least, to the recognized psychological models of mind to compare and discuss their strengths and weaknesses.

Note, that parent minds of LR reproduce themselves in the children minds in indirect ways using certain forms of cloning, heritage and learning procedures.

Some constituents of reproduction of LR can already be processed artificially, i.e. by regular for the human community procedures.

The requirement v4 is questioning, in fact, whether completely artificial minds, cogs, can reproduce new cogs equal themselves and to the biological ones.

5.4. What are the validity criteria to make cogs equal by meaning to mind and whether cogs valid by those criteria can be constructed?

It is a long way journey to answer to these questions and elaborate some approaches to implement.

6. Conclusion

Valid cogs, if constructed, confirm the assertion that mind is a modeling based problem formation and solving procedure able to use knowledge gained from the solutions to promote the utilities of LR in their negentropic games.

Synchronously, mental cogs provide a constructive model of mind as the ultimate instrument for cognition. Knowledge on the nature of instruments for revealing new knowledge gives a new look on the knowledge already gained or expected and raise new consequent questions.

Therefore, revealing by cogs the new knowledge on the instruments of cognition it is worth to question the new aspects of relationships between mind and the overall knowledge mind creates and uses.

Ongoing experiments on study of cogs are based on the technique of evaluating adaptive programs and their parts by local tournaments and use the game solving package with its kernel Personalized Planning and Integrated Testing (PPIT) and Strategy Evaluation units [Pogossian,1983,2005,2007].

Bibliography

[Atkinson,1993] G. Atkinson Chess and Machine Intuition. Ablex Publishing Corporation, Norwood, New Jersey, 1993.

[Botvinnik,1984] M.Botvinnik Computers in Chess: Solving Inexact Search Problems. Springer Series in Symbolic Computation, with Appendixes, Springer-Verlag: NY, 1984.

[Flavell,1962] J. Flavell The Develometal Psycology of Jean Piaget, D.VanNostrand Company Inc., Princeton, New Jersey, 1962.

[Furnkranz, 2001] J.Furnkranz Machine Learning in Games: A Survey in “Machines that Learn to Play Games”, Nova Scientific, 2001.

[Mandler,2004] Mandler J. The Foundations of Mind: Origins of Conceptual Thought. Oxford Univ. Press, 2004.

[Neuman, 1966] John von Neuman.Theory of Self-reproducing Automata. University of Illinois Press, 1966.

[Pogossian,2007] E.Pogossian. On Measures of Performance of Functions of Human Mind. 6th International Conference in Computer Science and Information Technologies, CSIT2007, Yerevan, 2007, 149-154

[Pogossian ,2006] E.Pogossian. Specifying Personalized Expertise. International Association for Development of the Information Society (IADIS): International Conference Cognition and Exploratory Learning in Digital Age (CELDA 2006), 8-10 Dec., Barcelona, Spain (2006) 151-159

[Pogossian,2005] E. Pogossian. Combinatorial Game Models For Security Systems. NATO ARW on "Security and Embedded Systems", Porto Rio, Patras, Greece, Aug. (2005) 8-18

[Pogossian,2007] E.Pogossian, V. Vahradyan, A. Grigoryan. On Competing Agents Consistent with Expert Knowledge", Lecture Notes in Computer Science, AIS-ADM-07: The Intern. Workshop on Autonomous Intelligent Systems - Agents and Data Mining, June 5 -7, 2007, St. Petersburg, 11pp.

[Pogossian,1983] E.Pogossian. Adaptation of Combinatorial Algorithms (a monograph in Russian), 293 pp. 1983. Yerevan.,

[Pylyshyn,2004] Z. Pylyshyn Seeing and Visualizing: It's Not What You Think, An Essay On Vision And Visual Imagination, http://ruccs.rutgers.edu/faculty/pylyshyn.htm,l2004.

[Roy,2005] D.Roy Grounding Language in the World: Signs, Schemas, and Meaning Cognitive Machines Group ,The Media Laboratory, MIT (http://www.media.mit.edu/cogmac/projects.html) 2005.

[Searle,1990] Searle J. Is the brain's mind a computer program? Scientific American 262, pp26-31, 1990.

[Shannon, 1949] C.E.Shannon. The Mathematical theory of communication. In: The Mathematical Theory of Communication. Ed. C.E.Shannon and W.Weaver. University of Illinois Press, Urbana, 1949.

[Shrodinger,1956] E.Shrodinger . Mind and Matter. Cambridge, 1956.

[Winograd,1986] T.Winograd, F.Flores. Understanding Computers and Cognition (A new foundation for design). Publishers, Chapter 2, pp. 11–59, Huntington, NY, 1986.

Authors' Information

Head of the Cognitive Algorithms and Models Laboratory at the Academy of Science of Armenia (IPIA) , professor at the State Engineering University of Armenia , Marshall Bagramyan Av.24, Academy of Sciences, Yerevan, 0019, epogossi@aua.am,

Major Fields of Scientific Research: models of cognition and knowledge processing, strategy elaboration, algorithms.

FAST RAY TRACING ALGORITHM FOR CLOUD SLICE VISUALIZATION

Ostroushko Andrii, Bilous Nataliya, Shergin Vadim

ABSTRACT. The three-dimensional computer graphic is widely used today in visualization systems of scientific researches, a computer tomography, producing of computer games, TV, animation, advertise manufacture etc. Especially high requirements are shown to realness of formed images. One of the most perspective methods is ray tracing. Fast ray tracing algorithm is proposed. This algorithm allows forming images of cloud slice according to the model described in [Ostroushko, 2010]. Using projecting ray's classification allows excluding rays which has not crossings with cloud. Due to low memory requirements for storage of the cloud model and intermediate data fast ray tracing algorithm can be implemented on the GPU-based systems.

Keywords: cloud, ray tracing, β-slice

ACM Classification Keywords: I.3.5 Computational Geometry and Object Modeling

Conference topic: Information Retrieval

Introduction

High flexibility of the cloud model and high performance of visualization algorithm makes it to be widely used in real time visualization systems. One of the actual requirements to modern three-dimensional computer graphic systems is high realness of formed images. High realness is provided by image resolution increasing, static and dynamic shades forming, environment transparency considering, various special effects generating. Method of synthesis has a key role in forming realistic images. Most perspective method is ray tracing, which provides extremely highly realistic images. High computational complexity is the main lack of this method. That is why developing of an effective models of elements for three-dimensional scenes and visualization algorithms for ray tracing method is an actual problem.

Modifying algorithm from [Gusyatin, 2000a] for projective rays and cloud spheres intersection parameters searching is proposed.

Using projecting ray's classification allows excluding rays which has not crossings with cloud. This approach allows considerably reduce computational job and correspondingly processing time for image synthesis. High level of concurrentability of computations allows proposed algorithm to be widely used in multiprocessors, distributed, GRID systems and clusters. Moreover, due to low memory requirements for storage of the cloud model and intermediate data fast ray tracing algorithm can be implemented on the GPU-based systems.

Problem

Solution of the equations of the projection ray, given in parametric form, and the surface equation in an implicit form is the main approach in the implementation of the ray tracing method for today. Traditional algorithm for searching the intersection point position of projective rays with cloud spheres allows substantially reduces the computation job [Hill, 2001].

However, as can be seen in [Ostroushko, 2010], the time spent on each frame synthesis is still large, so traditional algorithm can’t be employed in real-time systems. Therefore fast algorithm designing of the intersection point position of projective rays with cloud spheres searching for ray tracing method is actually needed.

Solution of the task

Traditional algorithm [Foley, 1996] allows processing each projective ray independently, that make it possible to provide high level concurrentability of the computation. It is rays classification that provides increasing effectiveness and reducing cloud synthesis time [Gusyatin, 2000a].

Modifying algorithm from [Gusyatin, 2000a] for projective rays and cloud spheres intersection parameters searching is proposed.

Step 1. Transition from cloud’s coordinate system (c.c.s.) to viewer’s one (v.c.s.) is carrying out [Hill, 2001]. For implementation this step it’s enough to re-count centers of spheres coordinates:

$$p_i = Mp_i^0,$$

where p_i^0 – center of the *i*-th sphere in c.c.s.;

p_i – center of the *i*-th sphere in v.c.s.;

M – c.c.s to v.c.s. transition matrix.

Step 2. Determination the intersection parameters of β-slice with base cloud sphere (zero-level sphere). The angle between nearby β-slices $\Delta\beta_x$ and α-slices $\Delta\alpha_y$ should not exceed the maximal angular error of the system [Gusyatin , 2000b]. Thus formed image resolution *X* x Y is computed. Implementation of the first step results that β-slice is always a plane, passing through Y-axis in v.c.s. and x-th pixels column of the formed image (Fig.1). Thus β-slice equation defined as follows:

$$x\cos(\beta_x) - z\sin(\beta_x) = 0 \tag{1}$$

where $x \in \{1,\ldots,X\}$ – column number in the formed image;

β_x – angle between the plane of β-slice *x* and Z-axis.

In order to plane and sphere has been crossed the distance d_i from β-slice’s plane to the center of the i-th sphere should not exceed the radius r_i of this sphere:

$$d_i \le r_i \tag{2}$$

The distance from point $p(x_0, y_0, z_0)$ to the plane $Ax + By + Cz + D = 0$ defined as

$$d = \frac{\left| Ax_0 + By_0 + Cz_0 + D \right|}{\sqrt{A^2 + B^2 + C^2}} \quad (3)$$

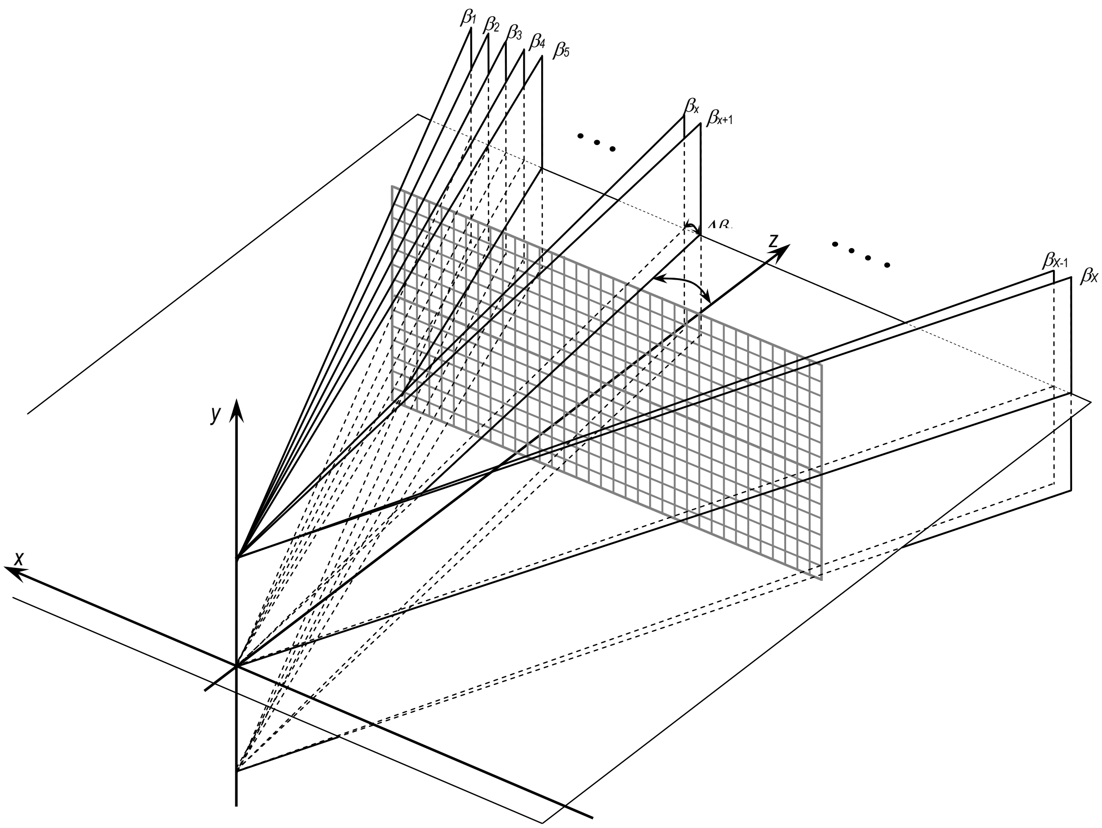


Figure 1 – Forming β-slices for ray classification

It's clear from (1) that denominator in (3) is always equal to one for any β-slice. Hence, it is sufficient to substitute the sphere's center coordinates obtained at the first step of the algorithm into (1) for verifying the inequality (2). This process can be performed for each β-slice independently. If sphere and some β-slice has not intersections the algorithm for that β-slice ends. Otherwise two lists for spheres numbers storage creates and zero as number of the base sphere puts into the first list.

Step 3. Iteration process for each β-slice begins. The list of areas crossing this β-slice is formed at each iteration. Number of sphere n^k current β-slice has intersection with extracts from the first list and numbers of spheres at the next iteration n_j^{k+1} calculates:

$$n_j^{k+1} = m\, n^k + j; \quad j \in \{1,\ldots,m\},$$

where $m = 5$ – number of spheres the previous iteration sphere is splitted into,

k – iteration number.

Until emptying the first list, condition (2) is checking for each n_j^{k+1} sphere and if true sphere's number puts into the second list. Then lists are swapped and iteration number is incremented. Accessing the required level of detalization after $k \le K$ iterations leads to passing to the next step of the algorithm.

Step 4. For each β-slice having non-empty list of spheres the range of α-slices is calculated (Fig.2). To do this, intersection parameters of plane with sphere (circle's center coordinates and radius) are determined. Center of the circle is determined by turning the center of the sphere at an angle β_x around the Y-axis:

$$x_i^{\beta_x} = z_i \cos(-\beta_x) - x_i \sin(-\beta_x) = z_i \cos(\beta_x) + x_i \sin(\beta_x);$$
$$y_i^{\beta_x} = y_i,$$

here x_i, y_i, z_i are the i-th sphere coordinates at v.c.s.,

$x_i^{\beta_x}, y_i^{\beta_x}$ – the coordinates of the center of the circle on a β-slice plane.

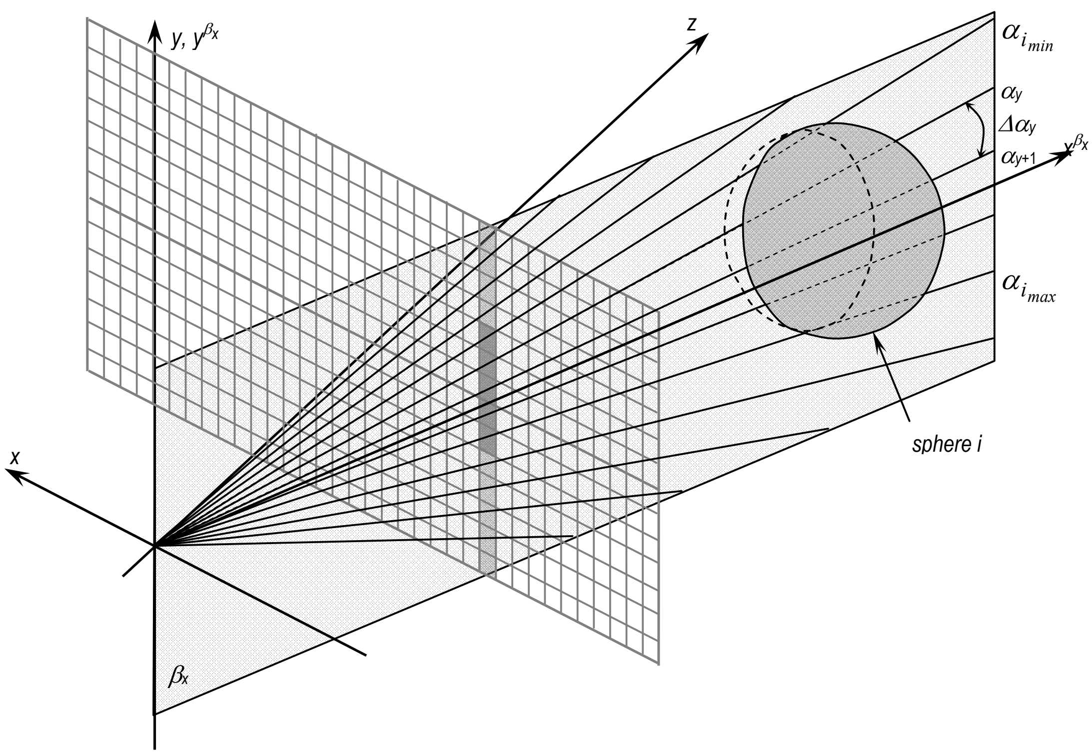


Figure 2 – Classification by α-slices

Parameters obtained while checking condition (2) are used for determining circle's radius on a β-slice plane:

$$r_i^{\beta_x} = \sqrt{r_i^2 - d_i^2}\,.$$

If the condition $x_i^{\beta_x} - r_i^{\beta_x} < 0$ is true, then the sphere is behind the viewer and should be excluded from further processing. Otherwise, the range of α-slices that intersects the circle is determined:

$$\alpha_{i_{\min}} = \frac{x_i^{\beta_x} y_i^{\beta_x} + r_i^{\beta_x}\sqrt{A}}{\left(x_i^{\beta_x}\right)^2 - \left(r_i^{\beta_x}\right)^2};$$

$$\alpha_{i_{\max}} = \frac{x_i^{\beta_x} y_i^{\beta_x} - r_i^{\beta_x}\sqrt{A}}{\left(x_i^{\beta_x}\right)^2 - \left(r_i^{\beta_x}\right)^2},$$

$$A = \left(x_i^{\beta_x}\right)^2 + \left(y_i^{\beta_x}\right)^2 - \left(r_i^{\beta_x}\right)^2.$$

If $A < 0$ the viewer is inside the sphere. Thus, the range of α-slices is bounded by α_1 to α_Y and one should process all the rays for the current β-slice.

The working range of α-slices is defined as an intersection of two ranges $\left\{\alpha_{i_{\min}}, \alpha_{i_{\max}}\right\}$ and $\left\{\alpha_1, \alpha_Y\right\}$:

$$\left\{\alpha_{i_{\min}}', \alpha_{i_{\max}}'\right\} = \left\{\alpha_{i_{\min}}, \alpha_{i_{\max}}\right\} \cap \left\{\alpha_1, \alpha_Y\right\} \quad (4)$$

Step 5. The list for storage an intersection parameters of the ray with cloud's spheres is created. These parameters are distances from viewer to intersection points l_1 and l_2, transparency and normal direction. Coordinates *x*, *y, z* could be recovered uniquely by the parameters of the projection ray and the distance *l*. Transparency and distances are necessary for the relative position of the cloud's spheres and other objects in the scene are to be taken into consideration when calculating pixel's color.

Step 6. For each projection ray that belongs into the operating range (4), points of intersection with the circle are calculated:

$$x_i^{1,2} = \frac{x_i^{\beta_x} + By_i^{\beta_x} \mp \sqrt{(x_i^{\beta_x} + By_i^{\beta_x})^2 - A(1+B^2)}}{1+B^2} \quad (5)$$

$$B = \tan\left(\alpha_y\right)$$

Coordinates of the intersection point calculated by (5) puts into output list which forms for each projection ray separately.

It should be noted that the values of *B* and $1+B^2$ can be pre-computed and tabulated, because they depend only on the angular resolution of the graphics system.

Step 7. All of the lists obtained in the previous step are sorted by the first point.

Step 8. Processing of the output list for determining the transparency of each sphere is beginning with the sphere closest to the viewer.

Transparency of the medium inside the cloud is determined by the meteorological optical range (MOR) S_M, which is defined as a visibility distance of a black body. MOR is a unique characteristic of the rejection of the radiation by the medium and is also an input parameter of the model.

For practical calculations of MOR the medium transmittance (T) is used. The specific transparency (t) is defined as transmittance of an optically homogeneous layer of medium with thickness equal to length unit. Knowing the specific transparency (t) and thickness of the layer of the medium (l) the transmittance could be found [McCartney, 1976]:

$$T = t^l \tag{6}$$

Specific transparency and MOR are related as $S_ì = \frac{\ln(0.02)}{\ln(t)}$. Thus, knowing the MOR and calculating path length of the projection ray inside the sphere one can easily obtain the transmittance.

Distances l_1 и l_2 are calculated on the basis of intersection point coordinates (5) as follows:

$$l_i^1 = x_i^1\sqrt{1+B^2}; \quad l_i^2 = x_i^2\sqrt{1+B^2}.$$

Length of area on which a projection ray passes inside a sphere is determined as $l = l_i^2 - l_i^1$. The base transmittance T_i is estimated according to (6). The value of function-modifier of cloud’s transparency is then given by [Ostroushko, 2010] :

$$f(d_i, r_i) = 1 - \frac{d_i}{r_i}.$$

The corrected transmittance T_i' of the *i*-th sphere having been calculated on the basis of the function-modifier is determined as

$$T_i' = T_i f\left(d_i, r_i\right) + 1 - f\left(d_i, r_i\right).$$

Step 9. The normal direction to sphere surface for determining color of the cloud sphere is needed. Therefore, recovering the coordinates x, y, z of the first intersection point is required:

$$x_i^1 = l_i^1\cos\left(\alpha_y\right)\sin\left(\beta_x\right); \quad y_i^1 = l_i^1\sin\left(\alpha_y\right); \quad z_i^1 = l_i^1\cos\left(\alpha_y\right)\cos\left(\beta_x\right) \tag{7}$$

On the basis of (7) the surface normal at the first intersection point is calculated: $n_{x_i} = x_i^1 - x_i$; $n_{y_i} = y_i^1 - y_i$; $n_{z_i} = z_i^1 - z_i$. Then the normal direction is modified according to following relation:

$$\vec{n}' = f(u,v)\vec{n} \tag{8}$$

where $f(u,v)$ – parametric function of the normal variation [Ostroushko, 2010].

Step 10. The transparency of the spheres is corrected. That correction is carried out in accordance with the relative position of the spheres. It is necessary for calculating the pixel color corresponding to the current projection ray. The variants of mutual location of spheres on a projection ray are shown at Fig. 3.

In the first case the correction is not executed and $T_i'' = T_i'$; $T_{i+1}'' = T_{i+1}'$.

In the second and third cases the transparency of the sections of the first sphere t_{i1} и t_{i2} should be defined, thus $T_i' = t_{i1}t_{i2}$. The new values of transparency determines as

$$T_i'' = t_{i1}; T_{i+1}'' = t_{i2}T_{i+1}' \tag{9}$$

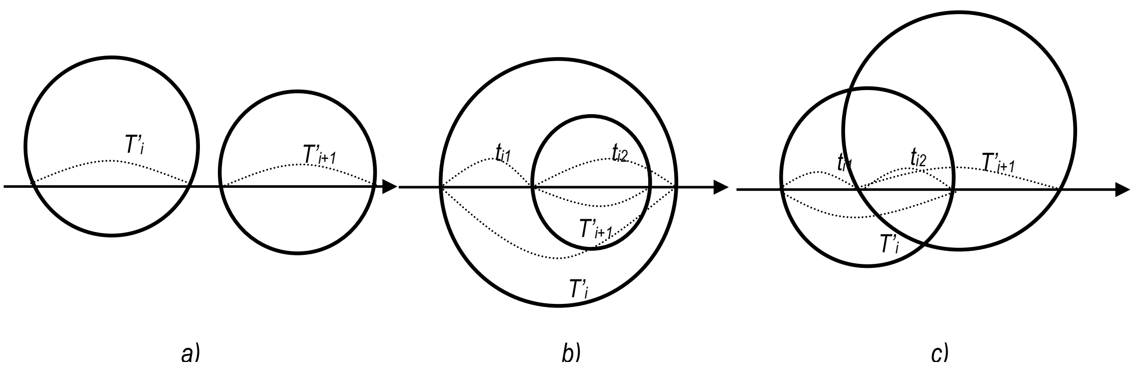


Figure 3 – Variants of mutual location of spheres:

a) spheres do not intersect; *b)* one sphere is entirely into another; *c)* spheres intersects

If there is more than two spheres that intersects simultaneously, step 10 of the algorithm repeats till the ray has any intersection with spheres. Further actions on the computation of pixel color can be performed using standard algorithms [Hill, 2001, Foley, 1996] taking into consideration the spatial position of clouds and other objects in the scene with respect to each other and light sources.

Step 11. The resulting color of the light regarding for (9) is given by

$$C = C_1\left(1 - T_1''\right) + \sum_{l=2}^{n}\left(C_l\left(1 - T_l''\right)\prod_{k=1}^{l} T_k''\right) \tag{10}$$

where C – color of the sphere with the regard for the modification of the normal at step 9,

n – number of spheres on the projection ray .

It should be noted that the amount of computational job needed at the last step of the algorithm could be reduced by checking the resulting transparency after each multiplication in (10). If $T \le 0.2$ computation can be stopped and the points remaining in the list can be excluded from further processing because, in accordance with the Weber’s law [Sedunov, 1991, Drofa, 1981], human eye stops to differ further alterations in the transparency of the clouds.

Results

Using the proposed algorithm allowed to accelerate the computational process approximately four times compared with the classical algorithm that uses a parametric description of the projection rays. Table 1 shows computational time needed for visualization of cloud image with resolution 800x600 pixels using one processor core. Personal computer with Intel CoreDuo E6600 processor and 4 GB of RAM was used.

Table 1 – Comparison of algorithms by performance

The number of spheres on the last level	Computation time, ms	
	Traditional algorithm	Algorithm with ray classification
15625	4750	1171
390625	9203	2422
9765625	22116	8844

Conclusion

The proposed algorithm uses a space partition by α-slices and β-slices, allowing to classify the rays by the presence of intersections with the spheres. Projection rays having not any intersection with the cloud are excluded from the analysis. It should be noted that using such technologies as AMD Stream or NVIDIA CUDA will allow to improve the performance essentially due to possibility of using concurrent computation either by β-slices or by α-slices while processing parameters of the intersection of projecting ray with cloud spheres, thus will enable to perform processing in real time mode.

References

[Drofa, 1981] Drofa A.S., Katsev I.L. 11. Some aspects of vision through the clouds and mists // Meteorology and hydrology. Sci. Mag. 1981. N 1. P. 101-109 (in Russian).

[Foley, 1996] Foley J.D., van Dam A., Feiner S.K., Hughes J.F. Computer Graphics (principles and practice) by Addison-Wesley Publishing Company, Inc. 1996,– 1175 p.

[Gusyatin, 2000a] Gusyatin V.M. Scene scanning algorithms in visualization systems // Radioelectronics and Computer Science. – 2000. – №1. – P. 46-49 (in Russian).

[Gusyatin, 2000b] Gusyatin V.M., Bugriy A.N. Calculation of angle error criterion of the specially designed processor in a vusualization system //Management Information System and Devices. All-Ukr. Sci. Interdep. Mag. 2000. N 112. P. 4-10 (in Russian).

[Hill, 2001] Fransis S. Hill. Computer Graphics Using OpenGL (c) Prentice Hall 2001

[McCartney, 1976] E. J. McCartney, Optics of the Atmosphere Scattering by Molecules and Particles. Wiley, New York, 1976, 176–261.

[Nikulin, 2003] E. A. Nikulin, Computer Geometry and Computer Graphics. BKhV-Peterburg, St. Petersburg, 2003. 560p. (in Russian)

[Ostroushko, 2010] A. Ostroushko, N. Bilous, A. Bugriy, Y. Chagovets Mathematical Model of the Cloud for RAY Tracing. International Journal "Information Theories and Application" Vol.17 (2010) pp.18-26.

[Sedunov, 1991] Atmosphere: A Handbook, Ed. by Yu. S. Sedunov. (Gidrometeoizdat, Leningrad, 1991) [in Russian].

Authors' Information

Andrii Ostroushko – Ph.D., Associate Professor, Kharkiv National University of Radio Electronics (KhNURE), Lenin Prospect 14, Kharkiv 61166, Ukraine; e-mail:

osa@kture.kharkov.ua

Major Fields of Scientific Research: computer graphics, ray tracing, digital visualization systems.

Nataliya Bilous – Ph.D., head of ITLCVS laboratory, professor of Kharkov National University of Radio Electronics, Lenin Prospect 14, Kharkiv 61166, Ukraine; e-mail:

belous@kture.kharkov.ua

Major Fields of Scientific Research: image processing, segmentation and recognition, building of computer vision system, computer testing and learning systems.

Vadim Shergin – Ph.D., Associate Professor, Kharkiv National University of Radio Electronics (KhNURE), Lenin Prospect 14, Kharkiv 61166, Ukraine; e-mail:

shergin@kture.kharkov.ua

Major Fields of Scientific Research: mathematical statistics, data mining, fractal analysis, image processing

WAVELET TRANSFORM IN INVESTIGATIONS OF STUDENTS EDUCABILITY DEPENDENTLY ON DEGREE OF GRAFICAL SKILLS AUTOMATION OF WRITING PROCESS

Olga Kozina, Nataliya Bilous, Mykola Zapolovskij, Vladimir Panchenko

Abstract: Application of 2-level 2-D Haar Discrete Wavelet Transform of images with special words repeated by hand is offered. Metric of similarity for images with pattern words and images with students' written words like analog of proposed patterns is considered. A new approach to objective assessment of educability like dependence on level of graphic skills of process of writing is offered. Experiment is revealed the presence of statistic significant correlation between average score in special academic subjects group or in other words between special educability and level of graphic writing skills automation for high school students.

Keywords: educability, process of writing, 2-D discrete wavelet transform, image similarity

ACM Classification Keywords: I.5 PATTERN RECOGNITION and K.3.2 COMPUTER AND INFORMATION SCIENCE EDUCATION

1 The degree of automation of movements in writing process like a feature of the higher nervous activity

Process of writing is a complex activity and all divisions of the cerebral cortex participate in its formation. Psychophysical basis of process of writing is the interaction of different analyzers – speech-motor, auditory, visual, hand-motor. The interaction of such mental activities as thinking, memory, attention, imagination, external and internal speech occurs during writing. Process of writing is a complex motor skill, which includes technical, graphical and spelling skills. Graphic skills reflect the ability to write symbols by hand on paper quickly and clearly and spelling skills reflect knowledge how to apply hand-written characters and the abilities to use these rules during process of writing for the reflecting of its content. Technical skill is manifested in the ability to hold the right techniques and methods of writing, such as proper position of the body during process of writing or the location of paper, etc. The technical and graphic skills define a character and features of the handwriting of a person.

The process of writing according to its psychological content from the outset is a conscious act to be formed in the special education arbitrarily and requires a number of special operations for its implementation in contrary to spoken language which assimilated across the imitating of another person's speech.

The type of coordination of movements during process of writing that acquired in learning and the degree of automation of these movements greatly influence on the number and nature of deviations appeared during writing, from the typical patterns of writing of characters and symbols. Insufficient degree of automation of movements during process of writing will be manifested in significant changes in the implementation of characters and their compounds than it is happening in the cases with a sufficient degree of automation of these movements [Kornev 1999, Bernstein 1990].

Processes of excitations and inhibitions are main nervous processes taking place in development of writing. Ratio of the basic properties of these processes (power, balance, mobility) affects the formation of writing. Type of higher nervous activity, depending on the degree of dominance first or second signaling system in it and the degree of precision transfer of temporal links the first signaling system in the second one, which may be different for individuals, may affect the ability of that person more or less successfully imitate of handwriting features of another person. I. Pavlov identifies three main types of people depending on the predominance of the first or second signal system in the brain [Bernstein 1990]:

a) the artistic style, characterized by a predominance of the first signaling system;

b) the thinking style, characterized by a predominance of the second signaling system;

c) the average type, characterized by a relative balance of both signaling systems.

Other things being equal, person with a predominance of the first signaling system in the state more accurately imitate the handwriting of another person and demonstrate in his process of writing to the significant set of handwriting features of another person than those with a predominance of the second signal system. It is known that persons with a more precision transfer of temporal links the first signaling system in the second signaling system able to fully and comprehensively recognize the signs of his and others' handwriting and successfully imitate the handwriting of another person [Kornev 1999].

Thus, individual differences in process of writing to different persons are a reflection not only of external learning environments, but also characteristics of their higher nervous activity.

2 organization of the experiment to identify the statistical relationship between educability and the level of graphical skills automation of writing process

Pavlov's research showed that education is the formation of temporary connections in the cerebral cortex of the brain, and skill – the system of these bonds [Podlasiy 2003].

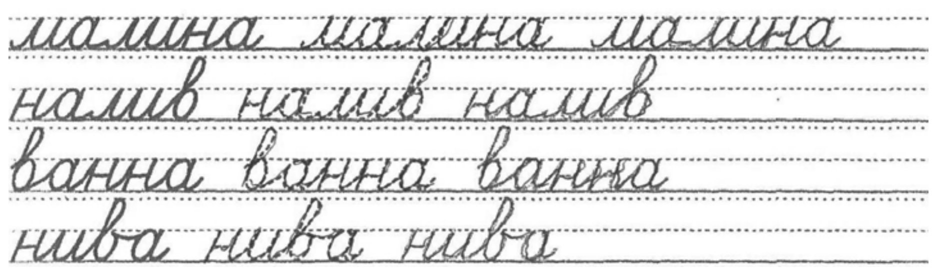
On the one hand, the motor component of process of writing requires considerable conscious effort and attention when writing, distracting him from more important tasks associated with given semantic task. On the other hand, it's known that productivity of education is directly proportional to level of educability and stability of attention of students, and it also depends on the level of their memory. Addition, the student achievements are directly proportional to educability depending on the strength of attention. Therefore, if we assume that student estimations reflect student's achievements and student with higher scores has greater educability, then the level of automation of the motor components of process of writing is a reflection of the level of attention, therefore, the level of educability that is the main hypothesis will be tested experimental.

Distinguish between general educability – the ability of mastering any material and special educability – learning ability of certain types of material (of various sciences, academic subjects, practices). The first type of educability is an indicator of the total talents of the individual while the second one is an indicator of only special.

To identify the statistical relationship between each type of educability and the level of graphical skills of writing process works from 18 third-year students of Computer and Information Technologies Faculty in the National technical university «Kharkiv Politechnical Institute» (Ukraine) were analyzed. Forms with patterns of words with different lengths in native language were given for students in several weeks. One part of such sheets is shown in Figure1a.

Spelling and fonts of handwriting patterns are conform to the norms of the Ministry of Education and Science of Ukraine and were taken from handwriting patterns for primary classes. Students were not aware of the purpose of the experiment, in other words they were not aware that the quality of this works will be compared with their educability. For students the problem had formulated such: as precisely as possible to repeat the handwriting patterns of words and to continue write to string the same words in the shortest time. Then an average score was calculated from previous 4 sessions for each student. Moreover, all subjects were divided into two groups – the general and special subjects. Each group of academic subjects is corresponds to general or special educability.

a



b

Fig.1. Examples of experimental materials: a – part of a sheet-pattern, b – part of a sheet-result

Table 1. Groups of academic subjects

General subjects	Special subjects
High Mathematics	Programming
Ukraine History	Basics of Computer Technology
Physics	Computer-aided design
Engineering Graphics	Discrete Mathematics
Theory of electro-magnetic circuit	System Programming
Philosophy. Logic	Applied theory of digital automata
Foreign Language	Computer electronics and circuitry
History of Science and Technology	Computing Algorithms and Data Structures
Cultural Studies, Ethics , Religious studies	Databases
Sport	Computers Architecture

Thus, the level of automation of graphical and motor skills of process of writing reinforced by the ability to properly carry out the assigned trivial tasks are compared with two values of average scores on general and special academic subjects.

3 Processing of experimental data

20 repeated and 20 own written three-, four- and five-symbols words for each student were selected from all obtained materials. To assess the level of graphical skills automation of writing process measure of similarity between the proposed handwriting patterns of word and both repeated word and own written same word are analyzed. Digital variants of experiment materials was received by scanning of the completed student sheets with experimental results and clean patterns sheets with same resolution. Algorithm for primary processing of digital images has been developed and implemented in Matlab. The sense of the algorithm is as follows:

- alignment of contours on patterns sheets and results sheets;

- subdividing of whole sheets on disjoint sub-images with patterns of every words;

- select a base point on the processed results sheet which will be left down vertex of the allocated area of interest with equal size dimensions to pattern sub-image of corresponding word (Figure 2);

- convert the result sub-image with analyzed word into grayscale.

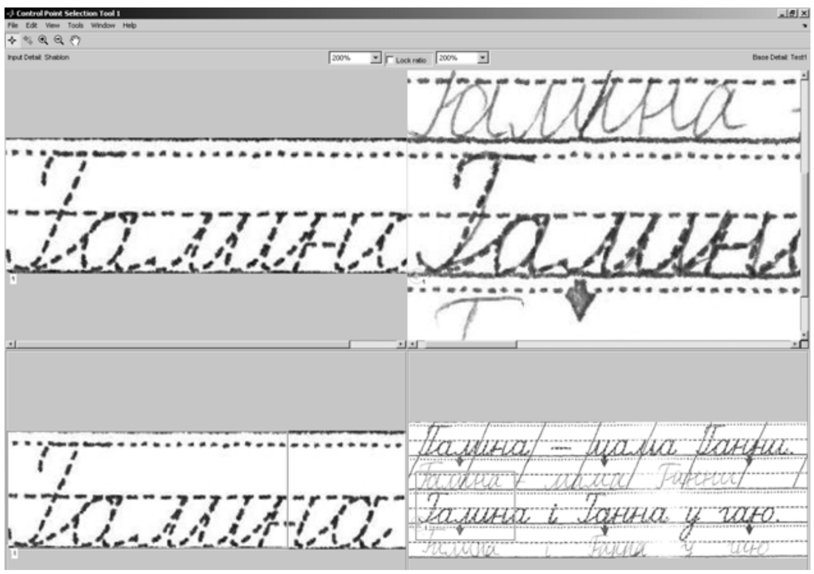


Fig. 2. Selection of interest area into results sheet

After primary processing the task of experiment results processing are considered as a problem of quantitative comparison of pattern and result sub-images with corresponding word through a specific metric called similarity measurement. A high similarity value between two images means a high level of graphical skills automation of writing process.

To images recognizing and to their similarity estimation Fourier Transform and discrete wavelet transform are successfully used [Bebis 2006, Pretto 2010, Rakhmankulov 2008].

As presented in [Burrus 1998] the Fourier Transform gives the spectral content of the whole signal, but it gives no explicit information regarding where in space those spectral components appear. A better tool for non-stationary signal analysis (whose frequency response varies in time, like in the images) is the Wavelet Transform [Iyengar 1997, Chui 1992, Mallat 1998]: it gives information about which frequency components exist and where these components appear. Properties of Discrete Wavelet Transform (DWT) coefficients were exploited in order to calculate the images similarity for our case. Every image can be represented as a linear combination of the images in the wavelet basis:

$$P(x) = \sum_k c_k \varphi_k(x) + \sum_k \sum_j d_{j,k} \psi_{j,k}(x) \qquad (1)$$

where c_k – approximation coefficients, $d_{j,k}$ – detailed coefficients, k – location index, j – scale parameter.

A wavelet representation of a function consists of a coarse overall approximation together with detail coefficients that influence the function at various scales. For our experiment 2-D Haar DWT of the grayscale images was selected. Given the resolution of images, we decide to stop at second level of decomposition.

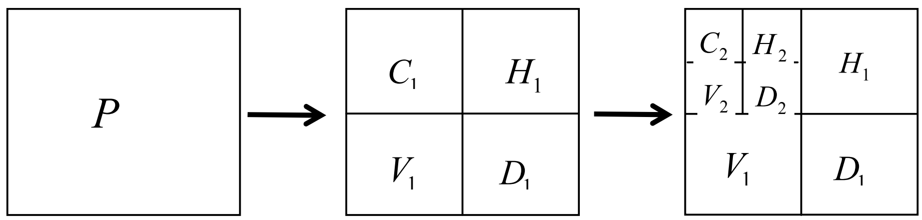


Fig.3. Two-level 2-D Wavelet decomposition

2-level 2-D Wavelet decomposition of input image P_i, and C_j – the approximation coefficients, H_j, V_j, D_j – respectively the horizontal, vertical and diagonal detailed coefficients in *j*-th level of the wavelet decomposition are shown in Figure 3. Transform going to higher level decompositions can significantly reduce the decomposed image's size, but higher level of decompositions discard important features in images, like edges and high frequency patterns useful for comparing of images in our experiment.

The Haar Wavelet is chosen as a wavelet type because of it is very effective in detecting the exact locations when a signal changes: image discontinuity is one of most important features chosen in image-based localization. The detail coefficients are the intensity variations along rows, columns and diagonals. Thus detailed coefficients can be used to detect and highlight image shapes. The higher the absolute values of coefficients are the higher probabilities that it encodes salient image features.

2 sets of 4 matrices of the horizontal, vertical and diagonal detailed coefficients with dimensions $a \times b = w/2^{wavelet\ ....level} \times h/2^{wavelet\ ....level}$ where w is the image's width and h is the image's height were obtained by using the functions from Matlab to calculate the coefficients of 2-level 2-D Haar DWT for each pair of pattern-result sub-images.

A metric that compares how many significant wavelet coefficients the query has in common with potential targets is used to compute image similarity usually [Pretto 2010]. But in our case, such a metric is not suitable, especially for the analysis of the repeated handwriting words. Therefore, considering the obtained matrix with coefficients as coordinates of multidimensional space, calculate the length of each vector $\rho(C_2), \rho(H_2), \rho(V_2), \rho(D_2)$ by the formula calculating the Euclidean distance between the coordinates of wavelet coefficients space for pattern and result sub-images with corresponding word:

$$\rho(C_2)=\sqrt{\sum_{l=1}^{a}\left(C_2^{sh}(l)-C_2^{t}(l)\right)^2+\sum_{c=1}^{b}\left(C_2^{sh}(c)-C_2^{t}(c)\right)^2},$$
$$\rho(H_2)=\sqrt{\sum_{l=1}^{a}\left(H_2^{sh}(l)-H_2^{t}(l)\right)^2+\sum_{c=1}^{b}\left(H_2^{sh}(c)-H_2^{t}(c)\right)^2},$$
$$\rho(V_2)=\sqrt{\sum_{l=1}^{a}\left(V_2^{sh}(l)-V_2^{t}(l)\right)^2+\sum_{c=1}^{b}\left(V_2^{sh}(c)-V_2^{t}(c)\right)^2},$$
$$\rho(D_2)=\sqrt{\sum_{l=1}^{a}\left(D_2^{sh}(l)-D_2^{t}(l)\right)^2+\sum_{c=1}^{b}\left(D_2^{sh}(c)-D_2^{t}(c)\right)^2} \quad (2)$$

where *sh* – index of matrixes' elements for approximation coefficients, for horizontal, diagonal and vertical detailed coefficients of second level wavelet decomposition of pattern sub-image, respectively,

t – index of matrixes elements for approximation coefficients, for horizontal, diagonal and vertical detailed coefficients of second level wavelet decomposition of result sub-image, respectively.

Distance between wavelet coefficients:

Shablon	Test	aproximate	horizont	vertical	diagonal
боти0.jpg	1.jpg	8615.42	3745.94	3687.62	1949.69
боти0.jpg	2.jpg	10306.16	3568.57	3764.89	1822.26
боти0.jpg	3.jpg	11057.46	3169.46	3623.40	1728.42
боти0.jpg	сам4.jpg	11741.59	3632.82	3856.86	1763.04
біб0.jpg	5.jpg	5218.48	2342.20	2476.77	1496.37
біб0.jpg	6.jpg	8495.67	2824.71	3037.32	1378.82
біб0.jpg	7.jpg	7404.42	2560.70	3020.47	1433.54
біб0.jpg	сам8.jpg	9721.91	2674.05	3229.37	1413.65
біб0.jpg	сам9.jpg	9209.23	2731.95	3241.18	1387.16
біб0.jpg	сам10.jpg	8793.02	2749.99	3118.10	1374.74
біб0.jpg	сам11.jpg	9157.25	2725.97	3187.95	1396.32
бік0.jpg	12.jpg	5378.52	2376.33	2095.80	1232.37
бік1.jpg	13.jpg	6569.02	2018.29	2085.77	1003.36
бік1.jpg	сам14.jpg	7698.29	2089.16	2028.22	926.28
бобер0.jpg	16.jpg	10772.04	4211.39	3803.16	1954.04
брат1.jpg	17.jpg	8802.34	3481.12	3227.12	1699.19
банка1.jpg	18.jpg	9824.43	3708.78	3282.47	1605.26
банка0.jpg	сам19.jpg	11539.41	3277.88	2626.58	1296.30
любити1.jpg	20.jpg	14250.36	4029.49	4162.91	1942.86
любити1.jpg	21.jpg	12506.20	3970.34	3802.85	1778.67
любити1.jpg	22.jpg	13528.34	4336.57	3716.31	1802.50
билка0.jpg	23.jpg	9152.23	3448.45	2993.95	1483.22
билка0.jpg	24.jpg	7851.68	3175.68	2715.22	1428.23
билка0.jpg	25.jpg	10422.15	3296.10	2871.13	1424.02
билка0.jpg	сам26.jpg	9959.44	2989.20	2524.55	1253.20
брюки0.jpg	27.jpg	14062.00	4311.65	4905.77	2171.07
брюки1.jpg	сам28.jpg	14645.20	3742.15	4724.41	2033.23
борода0.jpg	29.jpg	9624.16	3187.45	4090.36	1817.70

Fig.4. Example of part of results sub-images processing for one student

An example of the file with obtained according to (2) distances $\rho(C_2), \rho(H_2), \rho(V_2), \rho(D_2)$ between all pattern and result sub-images accordingly for one student are shown on Figure 4. Left two columns contain the names of files with compared sub-images and the next 4 columns contain $\rho(C_2), \rho(H_2), \rho(V_2), \rho(D_2)$ respectively. Obtained distances were grouped in 3 samples, depending on the length of the word-pattern. For each sample were calculated standard deviations.

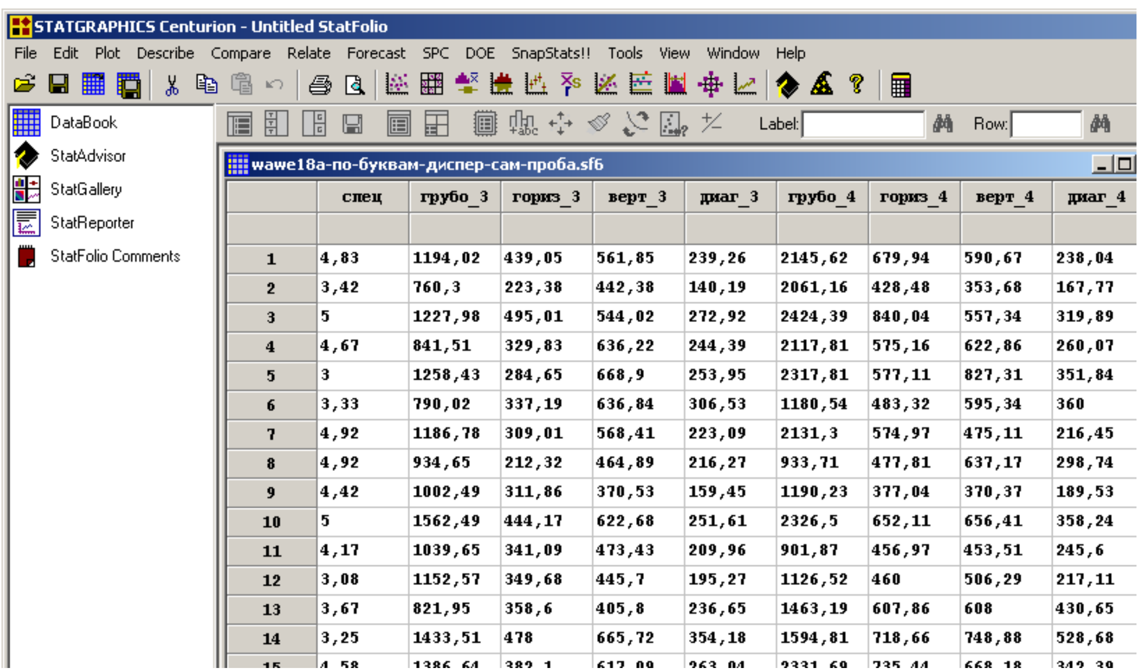
Standard deviation of the distance between the corresponding wavelet coefficients of pattern and result sub-images was calculated for each length of word. Correlation coefficients between the average score of general and special subjects groups disjoint and standard deviations of distances between 2D Haar DWT coefficients for every lengths of word were calculated in Statgraphics Centurion package.

Each line in Figure 5 contains data of one student in the following sequence: average scores on special academic subjects, standard deviations of $\rho(C_2), \rho(H_2), \rho(V_2), \rho(D_2)$ for three-symbol words and standard deviations of $\rho(C_2), \rho(H_2), \rho(V_2), \rho(D_2)$ for four-symbol words.

Pearson product moment correlations between each pair of variables x and y calculated by:

$$r(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad (3)$$

are represented in Table 2. P-values below 0,05 indicate statistically significant non-zero correlations at the 95,0% confidence level.



	спец	грубо_3	гориз_3	верт_3	диаг_3	грубо_4	гориз_4	верт_4	диаг_4
1	4,83	1194,02	439,05	561,85	239,26	2145,62	679,94	590,67	238,04
2	3,42	760,3	223,38	442,38	140,19	2061,16	428,48	353,68	167,77
3	5	1227,98	495,01	544,02	272,92	2424,39	840,04	557,34	319,89
4	4,67	841,51	329,83	636,22	244,39	2117,81	575,16	622,86	260,07
5	3	1258,43	284,65	668,9	253,95	2317,81	577,11	827,31	351,84
6	3,33	790,02	337,19	636,84	306,53	1180,54	483,32	595,34	360
7	4,92	1186,78	309,01	568,41	223,09	2131,3	574,97	475,11	216,45
8	4,92	934,65	212,32	464,89	216,27	933,71	477,81	637,17	298,74
9	4,42	1002,49	311,86	370,53	159,45	1190,23	377,04	370,37	189,53
10	5	1562,49	444,17	622,68	251,61	2326,5	652,11	656,41	358,24
11	4,17	1039,65	341,09	473,43	209,96	901,87	456,97	453,51	245,6
12	3,08	1152,57	349,68	445,7	195,27	1126,52	460	506,29	217,11
13	3,67	821,95	358,6	405,8	236,65	1463,19	607,86	608	430,65
14	3,25	1433,51	478	665,72	354,18	1594,81	718,66	748,88	528,68

Fig.5. Initial data for calculations of correlation coefficients

Table 2. Correlation coefficients between variables x and y

Number of symbols into word	Correlation coefficients $r(x,y)$				
	x	Group of academic subjects or educability type			
	Type of vector	y special	P-value	y general	P-value
3	$\rho(C_2)$	0,4752	0,0539	0,3952	0,1164
	$\rho(H_2)+\rho(V_2)+\rho(D_2)$	0,1459	0,5764	0,0382	0,8842
4	$\rho(C_2)$	0,5552	0,0207	0,6807	0,0026
	$\rho(H_2)$	0,5318	0,0280	0,4174	0,0955
5	$\rho(C_2)$	0,6905	0,0022	0,6228	0,0076
	$\rho(H_2)+\rho(V_2)+\rho(D_2)$	0,3418	0,1793	0,5378	0,0260

4 Interpretation of results and future works

Considering the Haar Wavelet, the approximation coefficient C_2 represent only the mean of the intensity of the pixels composing the macro-squares (4x4 pixels in our case) or general characteristic of a word. On the other hand, the difference between values of this coefficient for pattern and result sub-images is proportional to the number of pixels which involved to writing of analog of the pattern word but have distinct intensities of pixels at the pattern word or to general recognition, the readability of the word.

The horizontal, vertical and diagonal detailed coefficients H_2, V_2, D_2 are shown preferential direction of inaccuracies in the writing of words. Since the maximum value correlation coefficient between $\rho(H_2)$ and the average score in special academic subjects group for 4-synbols words (r = 0.5318) with no statistically significant correlation with $\rho(V_2)$ (r = 0.2389 and P-value = 0.3558) demonstrates that the loss of attention during process of writing leads to more extreme variations in the width of written word than variations of its height .

Absence statistically significant correlation coefficients between the mean scores for both groups of subjects and three-symbols words requires further study and may be explained by the fact that when writing of short words "truncated" motor skills are predominated in nervous activities of students, similar to the graphical skills of alone symbol writing. Movements of hand in such cases are minimum also and do not require a long and precise control of motor skills. Besides shorter words easier to remember, that is why smaller loss of attention occurs during process of writing short words. This gives rise to the formation of hypotheses about the higher values of correlation coefficients between special educability and measures of patterns-results sub-images similarity for seven-, eight- or nine-symbols words. Verifying this assumption, as well as the formulation of regression models of students' educability relating to the level of graphical skills automation of students will be the subject of future studies.

Conclusion

Thus, described experiments revealed the presence of statistic significant correlation between average score in special academic subjects group or in other words between special educability and level of graphic writing skills automation for high school students. Further study about abilities of using 2-D Discrete Wavelet Transform to quantify the level of graphic skills automation of writing process should give base for models of an objective assessment of educability.

Bibliography

[Kornev 1999] A. Kornev, Methodology to evaluate of automation of writing skills, Problems of pathology of development and disintegration of speech function. St. Petersburg: Pub. of St.-Petersburg University (1999), pp.54–59 (in Russian).

[Bernstein 1990] N. Bernstein, Physiology of movements and activity, Moscow: Nauka, (1990), 494p (in Russian).

[Podlasiy 2003] I. Podlasiy, Effective Pedagogy: the book for teachers, Moscow: Public education, (2003), 496p (in Russian).

[Bebis 2006] G. Bebis, A. Gyaourova, S. Singh, I. Pavlidis, Face Recognition by Fusing Thermal Infrared and Visible Imagery, Image and Vision Computing, vol. 24, no. 7 (2006), pp. 727-742.

[Pretto 2010] A. Pretto, E. Menegatti, Y. Jitsukawa, R. Ueda, T. Arai, Image similarity based on Discrete Wavelet Transform for robots with low-computational resources, Robotics and Autonomous Systems, 58 (2010) 879–888pp.

[Rakhmankulov 2008] V. Rakhmankulov, A. Akhrem, V. Gerasimova, A. Barsegov, Multidimensional wavelet analysis of images, Proceedings of ISA RAS T., 38 (2008), pp.278-288 (in Russian).

[Burrus 1998] S. Burrus, R. Gopinath, H. Guo, Introduction to Wavelets and Wavelet Transforms: A Primer, New Jersey: Prentice Hall, (1998), 282 pp.

[Iyengar 1997] S. Iyengar, L. Prasad, Wavelet Analysis with an Application to Image Processing, Chapman and Hall/CRC Press, (1997), 279p.

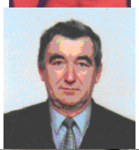
[Chui 1992] C. Chui, An Introduction to Wavelets, London: Academic Press, (1992), 266p.

[Mallat 1998] S. Mallat, A Wavelet Tour of Signal Processing, New York: Academic Press, (1998), 637p.

Authors' Information

Olga Kozina – Ph.D., lecturer of Computers and Programming Department of National Technical University ‘KPI’, Kharkiv, Ukraine; e-mail: hpi-kozina@rambler.ru
Major Fields of Scientific Research: medical information systems, segmentation and recognition ,system modeling

Nataliya Bilous – Ph.D., head of ITLCVS laboratory, professor of Kharkov National University of Radio Electronics, Kharkov, Ukraine; e-mail: belous@kture.kharkov.ua
Major Fields of Scientific Research: image processing, segmentation and recognition, building of computer vision system, computer testing and learning systems



Nikolaj Zapolovskij – Ph.D., Cheaf of Computer and Information Technologies Faculty of National Technical University ‘KPI’, professor of National Technical University ‘KPI’, Kharkiv, Ukraine
Major Fields of Scientific Research:system modeling, computer testing and learning systems

Vladimir Panchenko – teacher of Computers and Programming Department of National Technical University ‘KPI’, Kharkiv, Ukraine
Major Fields of Scientific Research: synthesis of tests for learning systems, artificial intelligence systems

CONCEPTUAL KNOWLEDGE MODELING ON THE BASIS OF NATURAL CLASSIFICATION

Mikhail Bondarenko, Kateryna Solovyova, Andrey Danilov

Abstract: It is difficult to exaggerate the importance, the urgency and complexity of “good” classifications creation, especially in knowledge management, artificial intelligence, decision making. To what extend it is possible within a short paper, the peculiarities and advantages of the new system method of the systemological classification analysis for the classifications of concepts creation were discussed. It is noted that the applying of the natural classification criteria improves considerably the quality and the power of the classification knowledge models and ontologies, allows taking into account the deep knowledge of any, including ill-structured, domains. In the process of the research conduction the system models of the domain fragment of the ontologies on the basis of the parametric classification were created. Some results of the actual domain “Social Networks in Internet” analysis and modelling and the ontology fragments, realized in the ontologies engineering tool Protégé 3.2, are also considered. The systemological classification analysis application has allowed proving the obtained classifications of social networks functions, taking into account the objects essential properties. It has also successfully recommended itself for deep knowledge acquisition; the basic hierarchy of classes, “good” classifications and ontologies creation; possesses predictive power, simple logically relevant structure, ensures the possibility of the correct inference on knowledge.

Keywords: conceptual knowledge, knowledge systematization, natural classification, ontology, systemological classification analysis, social network, hierarchy, systemology, artificial intelligence.

ACM Classification Keywords: 1.2 Artificial Intelligence – 1.2.6 Learning: Knowledge Acquisition

Introduction

The development of knowledge management, artificial intelligence, decision making and many other actual scientific and practical directions is determined by knowledge and its quality. As we know, knowledge, intellectual capital is the main competitive advantage, the foundation of modern organizations, enterprises, society, human and nations’ welfare and important component of decision making support systems.

In different spheres of knowledge acquisition and application conceptual models of subject domains play a leading role. "Historically," the species of domain models are: dictionaries, thesauri (in linguistics), conceptual models (infological, semantic models - in databases), UML diagrams (of classes, of use cases, ... - in object-oriented analysis and modeling), models of knowledge (semantic nets, frames, ... - in artificial intelligence), ontologies (from the viewpoint of the realization and application one of the most modern kind of a domain model, aimed primarily at the knowledge application in Internet).

The basis of such models is the relationships of the hierarchy between concepts (concepts classification), in the first place, the relations *genus-species* and *part-whole*, about two millennia known in formal logic. These relations in the theory of classification are called the relations of *taxonomy* and *meronomy*, in artificial intelligence – *genus-species: Isa* (class - class), *Instance-of* (class - element) and *part-whole: Part-of*; in object-oriented analysis and modeling – *generalization / specialization* and aggregation (in some cases, *composition*), respectively, etc. In systemology to these relations corresponds one *relation of the functional ability of the whole support*, respectively, for system-classes and concrete systems (which are reflected in general and single concepts).

How effective are the methods of the concepts classification creation - the basis of modern models of knowledge of domains? The analysis shows that in most domains the classifications are subjective; many of them do not meet even the requirements of formal logic. That is why it is proposed to apply a new unique method of the systemological classification analysis based on the natural classification [E. A. Solovyova, 1999; E. A. Solovyova, 1991; E. A. Solovyova, 2000], which has successfully recommended itself for deep knowledge acquisition, the basic hierarchy of classes, "good" classifications and ontologies creation in all, including ill-structured domains.

Introduction to the Natural Classification as the Conceptual Knowledge Systematics

As noted, this work is not about data classification into existing classes. We work with deep knowledge classifications and besides with the conceptual deep knowledge, on the conceptual level, determine classes (entities), properties and relations, and besides in accordance with their position in the domain, in the reality, in accordance with the systemic of the reality. Naturalists and other scientists interested for many centuries in the problem of "good" classification creation, *the position of objects in which reflects the reality (the domain), is determined by essential properties and relations of objects* and therefore possessing predictive power. This "good" classification was called systematics, or the natural classification, the first meaningful criteria of which were introduced by the Englishman Wavell more than 150 years ago; then by A. A. Liubishchev, Y. A. Schrayder and other scientists, for example, the natural classification - *is a form of the laws of nature presentation..., expresses the law of the systems of reality relationship, allows to reach the maximum number of goals, because it takes into account the essential properties, etc.* Such criteria are useful for fundamental science, but are not constructive for computer modeling, application in knowledge models and ontologies. That is why in Knowledge Acquisition Laboratory, at the Social Informatics Department and Scientific-Educational Knowledge Management Center for more than 20 years the systemic research of conceptual knowledge and natural classification has been conducted. For the first time the constructive criteria of the natural classification and a new method of systemological classification analysis which allow to take into account deep knowledge, objects essential properties and relations in domain models in the most objective way, have been obtained [E. A. Solovyova, 1999; E. A. Solovyova, 1991; E. A. Solovyova et al; 2000, E. A. Solovyova, 2000, etc.]. This method for the first time synthesizes system and classification analysis. The natural classification criteria correspond completely to the formal-logical criteria and also deepen and generalize them.

These fundamental results have not only theoretical but also an important practical value. They allow creating knowledge models and ontologies which take into account essential properties and causal-investigative relations, possess predictive power, simple logically relevant structure, allow generalization and unlimited knowledge refinement without redesigning classification, ensure the possibility of the correct inference on knowledge, recommendations and decisions making support, interface with the concepts of natural language application.

It is proved mathematically and systemologically and (with the use of the category theory and the categorical-functorial model of the natural classification obtaining) that the natural classification is the parametric one (including properties of all its elements), in which the properties classification determines (isomorphic) the objects classification, the properties properties classification – deep layer poperties – the properties classification, etc.). In practice, the consideration of one level of properties (their genus - species classification) allows making the classification model founded and really effective for solving on its basis the various tasks that require knowledge application.

Functional systemology - the systemic approach of the noospheric stage of science development – was created for and is aimed at complex, qualitative, ill-structured problems solving, it differs profitably from the traditional systemic approaches and for the first time really takes into account the systemic effect. Systemology, taking into account the principles of systemic, integrity and diversity, considers all objects, processes and phenomena as systems functioning to support the supersystem functional abilities. Systemology as modern system methodology does not regard system as a set but as a functional object which function is assigned by supersystem. Systemology in particular allows overcoming problems of traditional methods of system analysis at the expense of using conceptual knowledge as well as formalizing procedures of analysis and synthesis of complex systems and creating knowledge-oriented software tools for their simulation. The development of the concrete (internal) systems systemology of G. P. Melnikov for the system of classes allows deep knowledge getting and modeling for all, including ill-structured, domains [Bondarenko et al, 1998; E. A. Solovyova, 1999].

Social Networks Functions Classifications

The conceptual knowledge modelling will be accomplished on the example of the actual domain of social networks, including the ontology creation. Nowadays the need to solve complex problems requiring the knowledge of the domain specialists appears increasingly. To train highly qualified professionals progressive companies propose to use the conception of learning organizations. A learning organization as a tool for solving problems related to the company professional level improving. To create and acquire knowledge the company needs to be constantly in the process of self-improvement. One of the advanced methods of the organization development is the social networks use. The social networks in Internet functions research will allow understanding better the expediency of their use, to use the social networks more effectively in decision making, for further knowledge systematization in the social networks domain.

Resulting from the research the developped social networks classifications were not found. There are several articles where the social networks in Internet functions but not their classifications are mentioned. For example, the following main functions of social networks in Internet are allocated:

- profiles, communities, blogs dogear, activities [Byelenkiy, 2008];

- functions of personal profiles creation, of users interactions, of common goal achieving by means of the cooperation, of resources interaction, needs satisfaction due to the resources accumulation [Kuzmenko, 2009].

The functions research was conducted through the practical use of the network with identifying semantics the functions during the direct work with the network. Due to the weak structuring of the domain and the distinguished functions names in different networks, which in practice often have the same functional destination in most networks, it was necessary to undergo the registration process (or just to come into the network workspace) and to use practically the social network functions to determine their real functional destination. Also during the analysis process, the difficulties have appeared due to not clearly defined functions names, which haw required the additional research and the functions comparative analysis conduction [Danilov, 2010].

The analysis shows that in the first division base, for example, the communication (messages interchange) functions class is absent. In the second division base the search functions are absent and it is also not clear what is meant by the functions of common goal achieving by means of the cooperation. The authors of the given divisions do not exemplify the functions which refer to the classes of these divisions.

Thus, the knowledge systematization in the domain of social networks is needed. Subsequently it will allow not only to obtain the social networks ontology but also to improve the considered nets from the functional viewpoint, to expand the set of their functions, to improve the meaningful placement of the menu functions in concrete social networks. The results of the social networks systematization may be applied for a new social network creation taking into account the advantages and disadvantages of the existing social networks.

In this case we consider the classification creation by the functionality as the knowledge systematization in the given domain [Solovyova, 1999].

Let us consider the comparative analysis of the existing classification functions and the proposed network supposed systematics.

As an example of the hierarcy first level of the social network "В Контакте.ru" functions classification, implemented in the software tool Protégé 3.2. (see Figure 1) is presented.

The created classifcations are implemented in the software tool Protégé 3.2. The choice of this software tool is grounded by the fact that it is a free, open-source ontology editor. Protégé has an open, easy spread architecture at the expense of the functionality extension modules support which are freely available on the Protégé official site. The knowledge model is an OKBC-compatible (Open Knowledge Base Connectivity – it is the application programming interface for the access to knowledge bases), this allows applying in Protégé the one customized interface for different semantic markup languages processing. An example of such language is OWL (Web Ontology Language). All the listed above possibilities of Protégé, as well as the visibility of the obtained classifications was the reason for using namely this software tool.

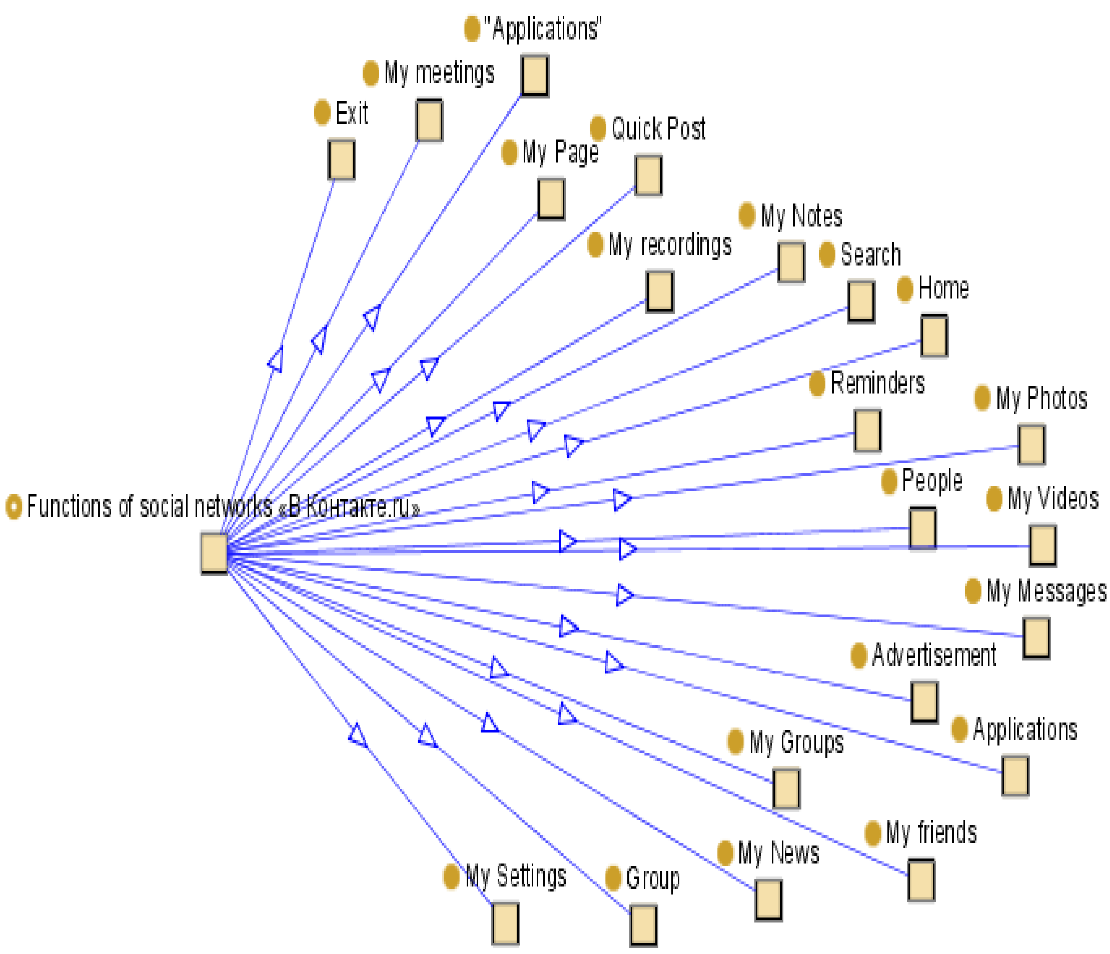


Figure 1. The social network «В Конmакmе.ru» functions of the first classification level.

The recommended informative placement of the functions of the first level of the hierarchy, taking into account the systemological classification analysis use and the natural classification criteria, is shown in Figure 2. The functions placement is understood as their placement in the networks workspace. The informative placement is understood as the functions structure their hierarchy in the social network menu. Our informative placement displays the functions relationship taking into account the knowledge systematization and the relations semantics between them in the best possible way.

Resulting from the analysis it was proposed to divide all the functions of the first level of the hierarchy on such groups: user information, my data, my messages, search and the function of input/output from the network workspace. Such functions placement will allow reducing the load on the user and speeding up the process of the needed function search, as it reflects the functioning of the whole support ability relation. The substantial functions placement in the social networks proposed, as the result of the conducted functions system analysis, will facilitate the functions search for the social network new users, will reduce the sense loading on the user when working with the network, by reducing the number of functions on the same level of the hierarchy.

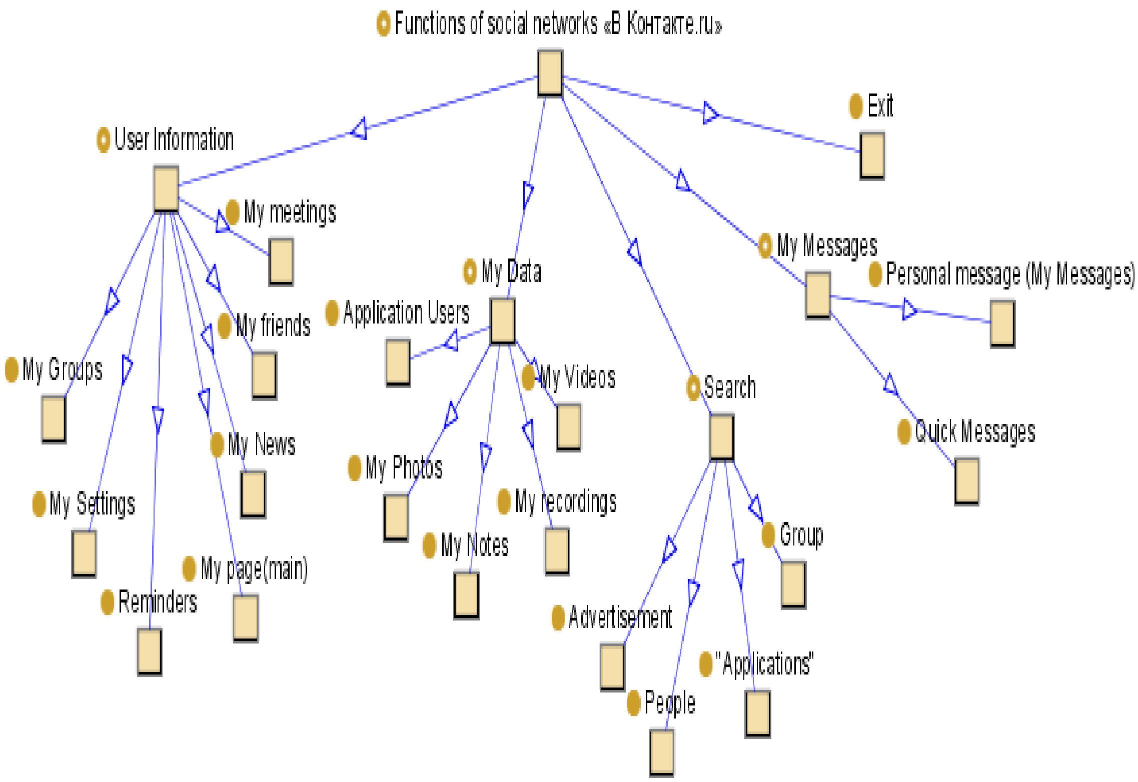


Figure 2. The social network "В Контакте.ru" systemological informative functions placement.

Also the knowledge systematization in the social networks specific functions will help to systematize the knowledge in the domain of the considered social networks in the Internet. After analyzing the social networks functions classification, created in the process of work, it was found that most social networks have a number of similar functions. In most cases, the functional destination coincided fully, and the functions differed significantly, for example, the function "Work with groups" appeared under the name "Groups", "Communities", "Groups and discussions".

The created functions classification reflects the semantics of the functions and of the relations between them, as well as the results of the knowledge systematization in the social networks in the Internet domain. Increasing the levels of the hierarchy in the functions classification, shown in Figure 2, allows finding the needed function faster, increasing the user convenience, accelerating the further functions menu development. Such analysis of the considered social networks functions allows creating the recommended functions classifications for each of them.

The advantage of the proposed classification of the social networks in Internet functions is that it includes the functions considered in popular social networks «В Контакте.ru» (http://vkontakte.ru), «Викепедия» (http://ru.wikipedia.org/wiki), «Мой Мир» (http://my.mail.ru/mail), «Connect.ua», «МойКруг» (http://moikrug.ru), «Science-community.org».

For these networks the functions classifications *by the relation of the functional ability of the whole support* were created that has given the possibility to develop the recommendations or the meaningful placement of the menu functions of the social networks according to the requirements of systemology and formal logic. As an example, in Figure 3 our recommended classification of functions of the first level of hierarchy by the relation "part-whole" for the social network of scientists «Science-community.org», implemented in a software tool Protégé 3.2 is shown.

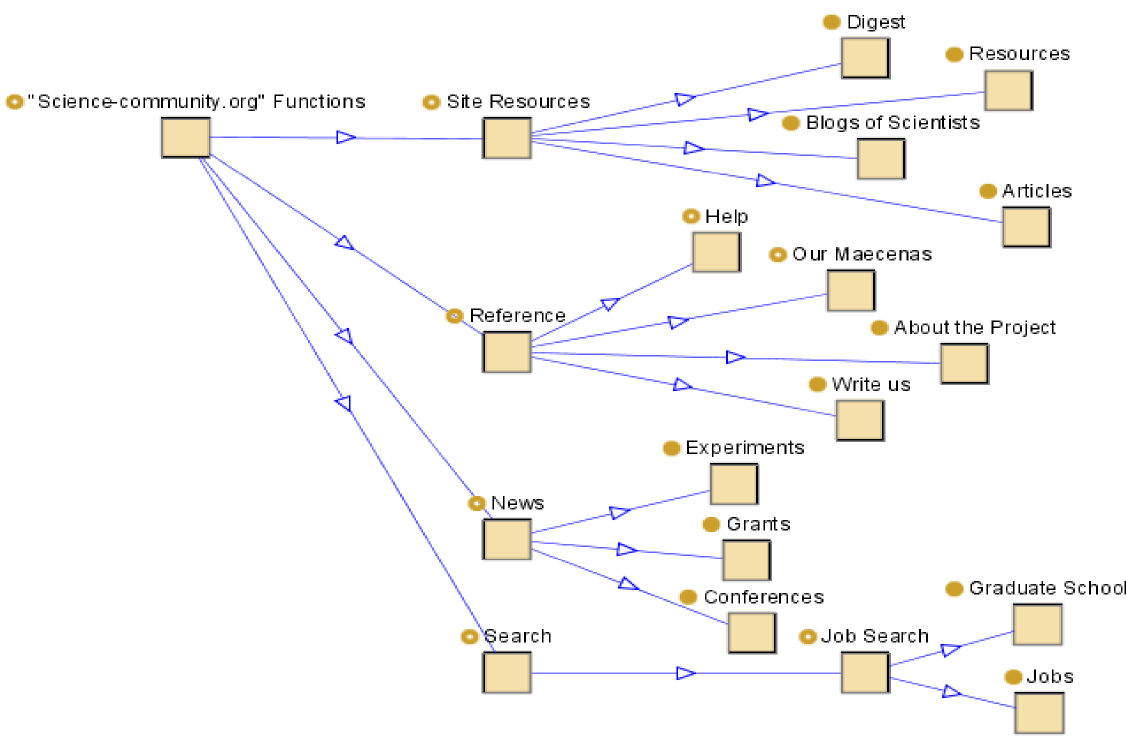


Figure 3. Recommended placement of the functions for the social network «Science-community.org»

The systemological classification analysis application has allowed justifying the obtained classifications of social networks in Internet, to take into account the objects essential properties of them. This classification gives the possibility to detect and predict the objects properties by their position in the classification, i.e. from the viewpoint of the possibility to apply the classification not only as an effective practical tool but also as a tool of the theoretical analysis in the correspondent domain.

The use of the systemological classification analysis allows formulating recommendations for the hierarchical structure of functions implementation in the social network, for their meaningful placement in the menu in accordance with the created classification. Such natural placement will allow to reduce significantly the load on the user, will improve his work, networks and the principles of their functioning mastering.

The obtained classifications of the social networks in Internet functions allow to determine easy which class this or that concrete function of social networks refers to with which the user may meet while working with social networks in Internet. The greatest number of functions refers to the functions of "search" and "work with network resources," the functions of "communication" are also important. This classification of the social networks in Internet functions can be viewed as a parametric (including the classification of properties) one, because the classes functionality are seen from their names. Resulting from the functions of various social networks research the functions classification fragment, shown in Figure 4 was built. The created classification fragment allows determining to which class refer the functions of the first level of the hierarchy of the social networks: «В Контакте.ru» (http://vkontakte.ru), «Википедия» (http://ru.wikipedia.org/wiki), «Мой Мир» (http://my.mail.ru/mail), «Connect.ua», «МойКруг» (http://moikrug.ru), «Science-community.org». The functions search was done by means of the practical use of a concrete function to verify its functionality. First the functionality for each concrete function was determined, and then the function appurtenance to the concrete class was determined. The obtained fragment of the classification “social networks functions” was realized in the software tool Protégé 3.2 is shown in Figure 4. This software tool was chosen due to a number of advantages [Shcherbak, 2008, etc.].

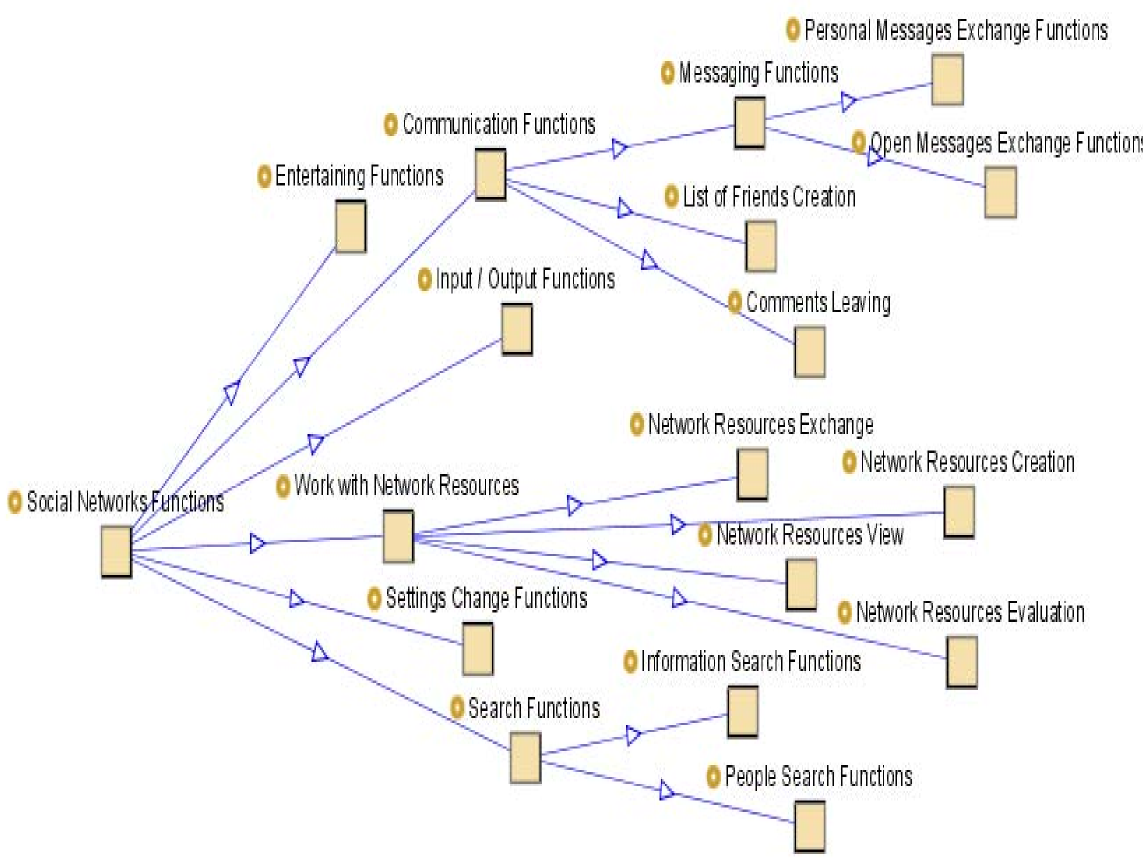


Figure 4. The social networks functions fragment classification by the relation of the functional ability of the whole support.

The obtained fragment of the social networks in Internet functions classification will allow becoming faster familiar with the functions of social networks in Internet, to choose more effectively the social network for registration, taking into account the functionality of the social network. The obtained results should be used for further knowledge systematization in the field of social networks in Internet.

The use of the method of systemological classification analysis allows obtaining new deep knowledge and systematizing knowledge in any domain in the most adequate and objective way taking into account the substantial properties and relations. The use of the systemological classification analysis allows evaluating the validity of any knowledge classification, the objects essential properties reflection in it; predicting new objects based on their properties.

Systemological research of social networks will allow systematizing knowledge in the social netwoks in the Internet domain and defining the appropriateness of various functions use in this or that social network, in a concrete organization.

Systemological Classification Analysis Application in the Social Networks Construction

Increase of the social networks in Internet influence of on the society has convinced many people to use social networks in business. Large corporations can afford to order a strong social network from firms of developers, but creation of such a network will require a lot of money. The enterprises (low-budget organizations) with a small income have not such a possibility, they may or attempt to use already functioning network or to attempt to create a social network by themselves. The latter variant is more advantageous, as the company itself regulates who will be the participant of the network, what tasks the social network must solve within the organization, etc. To create a social network in Internet it is necessary to use software for social networks creating.

Nowadays Internet is filled with a variety of software for the own social network creation. Many of them paid and (or) require deep knowledge in programming. There is also a number of software proposing to create a social network for free. This software proposes some free set of functions for a simple social network creating, there is also the possibility to use the supplement paid services.

The analysis of the software «Socialtext», «IBM Lotus Connections», «Jive SBS», «СвояСеть», «Connectbeam», «Ning», «Taba.ru» allows to make the conclusion that «Ning» (http://www.ning.com/), «Taba.ru» (http://taba.ru/), «СвояСеть» (http://svoyaset.ru/getform.html#) are the most acceptable for writing the recommendations to the social networks creation. They are conditionally free and do not require deep knowledge in programming. The disadvantage of the program service «Ning» is the absence of the interface in Russian. This disadvantage is significant for the recommendations to the social networks creation. In connection with it the software «Taba.ru», and «СвояСеть» were chosen. While creating the social network in «Taba.ru» it is recommended to use the social networks in Internet classification fragment shown in Figure 2.

In the process of writing recommendations the alternative menu creation of the social network has been tested using the systematological classification analysis. The social networks functions alternative menu created taking into account the results mentioned above was maximally approximated to the menu corresponding to the formal logic and systemological classification analysis. Unfortunately, the considered designers have the limited functionality and do not allow applying fully the results of the conducted research. In the process of work guidelines and recommendations to social networks creation in Internet in the software «Taba.ru», «СвояСеть» have been developed, the shortcomings and benefits of a social network creation in the selected designers have been revealed, as examples the demoversions of social networks in each of the designers have been created.

Model Creation of Knowledge Dissemination in an Organization with a Help of a Social Network

Despite the widespread of social networks in Internet, the models of knowledge dissemination in an organization by means of social networks sites have not been found. There is a number of articles describing the use of social networks for the social capital creating and using but there are no models of social networks implementation in an organization for knowledge dissemination and the employees' intellectual capital enhancing. The model creation of knowledge implementation and dissemination in an organization will allow increasing the organization competitiveness and solving such important practical tasks as the social networks implementation process acceleration, improving their functioning effectiveness and facilitating the process of knowledge acquisition and dissemination in the social network space by the employees.

When choosing a social network it is necessary to take into account several factors, such as the creation goals, the project budget, the tasks which will be solved by means of the social network, the expected users' range. The informational business model of a social network choice, aimed at the concrete organization problems solving, will help to facilitate the process of choosing a social network.

Systemology and the systemological classification analysis on the basis of Natural Classificaton application in social networks will allow increasing the functioning effectiveness of the fuctional menu and the functioning effectiveness of social networks, the networks implementation, facilitating the new functions implementation. Using the knowledge obtained during the social networks functions classification creation, the informational business model(Figure 5) describing the process of a social network in the Internet choice and development for increasing the organization competitiveness. This model realised in BPWin and describes the main processes in the organization when choosing and creating a social network (the definition of the goals and tasks solved by the social network; the means and the software tool for creating the social network choice; a brief description of the processes associated with the immediate introduction of a social network in operation, of the ways of promoting the social network use by the employees).

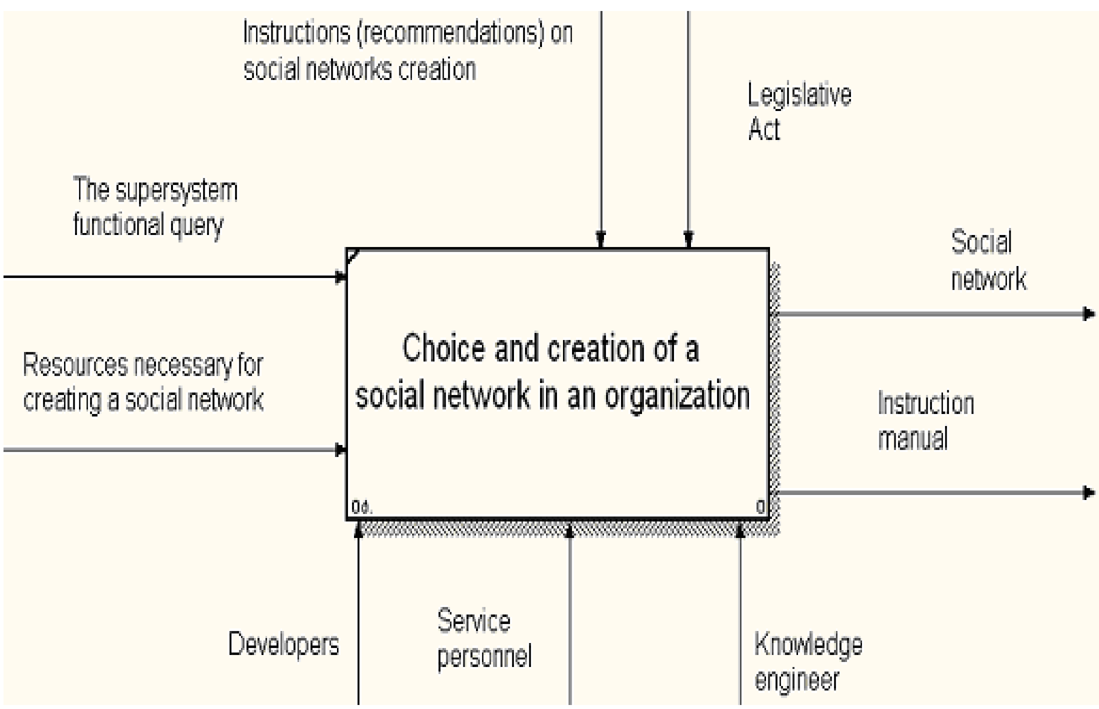


Figure 5. Context diagram of the informational model of the social network selection and creation in the organization.

In the future it is planned to develop the business model of social networks application in the Internet for knowledge management. The model will include the methods of exchanging both explicit and tacit knowledge of knowledge that will allow to increase the effectiveness of the social networks sites application in the organization for knowledge management.

The use of the created informational business model will allow facilitating and accelerating the Internet social networks choice and implementation process in the company and minimizing changes necessary for the effective functioning of the social network in the Internet; will allow reducing costs during the social network in the Internet creation and use.

The results of the work can be used as recommendations for the construction or choice of a social network by the organization for increasing its competitiveness, strengthening the relationships between the employees (increasing the social capital), increasing the intellectual capital of the employees and of the company on the whole.

The proposed results of the social networks may be used in the process of a learning organization creation, for decision making, intelligence technologies and artificial intelligence development.

Conclusion

The classifications of concepts are the basis of each science and are applied for solving various scientific-practical tasks. Now the classifications has got "the second birth" and are an main element of ontologies, computer models of knowledge, object-oriented analysis and modeling, intelligence technologies, knowledge management, decision making support and artificial intelligence, etc. That is why the role and the necessity of "good" classifications of concepts have increased now even more. Systemology application has allowed synthesizing system and classification analysis, discovering new criteria of systematics (natural classification) and their applying for knowledge systematization in any domain.

The Natural Classification criterion has been successfully used during the new method of the systemological classification analysis application. The results of the systemological research partially included in the paper may be used for the further knowledge systematization, creation of more effective alternative menus, natural language processing, etc.

Acknowledgements

The paper is partially financed by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA and the Consortium FOI Bulgaria (www.ithea.org, www.foibg.com).

Bibliography

[[E. A. Solovyova, 1991] E. A. Solovyova. Mathematical Modeling of Conceptual System: a Method and Criteria of a Natural Classification (Article). New York: Allertion Press, Inc., V. 25, No. 2., 1991.

[E. A. Solovyova, 1999]. E. A. Solovyova. Natural Classification: Systemological Bases [In Russian], Kharkov: KhNURE, 1999.

[E. A.. Solovyova, 2000]. Mathematical and Systemological Foundations of Natural Classification (Article). New York: Allerton Press, Inc., V. 33, No. 4., 2000.

[E. A. Solovyova, et al, 2000] D.B. Elchaninov, S.I. Matorin]. Application Of Categories Theory To Research and To Modeling Of Natural Classification (Article). New York: Allerton Press, Inc., V. 33, No. 2., 2000.

[Bondarenko et al, 1998] M. F. Bondarenko, E. A. Solovyova, S. I. Matorin. Foundations of Systemolgy, [In Russian], Kharkov : KhTURE, 1998.

[M. Auxilio, M. Nieto, 2003]. M. Auxilio. An Overview of Ontologies: Technical Report, Mexico: Center for Research in Information and Automation Technologies, 2003.

[Byelyenkiy, 2008]. A. Byelenkiy Business Perspectives of Social Networks // http://www.compress.ru/article.aspx?id=18650&iid=865 [In Russian].

[Kuzmenko, 2009] Kuzmenko. Social Network // http://www.itpedia.ru/index.php/Социальная_сеть [In Russian].

[Shcherbak, 2008] S. S. Shcherbak. A Few Words about the Protocol Open Knowledge Base Connectivity (OCBC) and about the ontologies redactor Protégé // http://semanticfuture.net/index.php?title [In Russian].

[Danilov, 2010] A. Danilov To the question of the knowledge systematization in the social networks domain, Newsletter of National Technical University "KhPI", [In Russian], Kharkov: NTU "KhPI", 2010. - №67.-199 p.

Authors' Information

Mikhail Bondarenko *– Rector of Kharkov National University of Radio Electronics, Corresponding Member of the National Academy of Sciences of Ukraine, Lenin Ave., 14, Kharkov, 61166, Ukraine; e-mail: rector@kture.kharkov.ua*

Major Fields of Scientific Research: Intelligence Technologies, Information and Knowledge Management, System Analysis, Artificial Intelligence, Decision Making, Knowledge Research and Application, Natural Language Processing, Knowledge and Natural Language Modeling, Business Intelligence, Competitive Intelligence, Modern (e-) Learning, Knowledge-Based Systems and Technologies, Systemological Analysis.

Kateryna Solovyova *- Chief of Social Informatics Departmen and Knowledge Management Center, Professor, Doctor of Technical Sciences, Kharkov National University of Radio Electronics, Lenin Ave., 14, Kharkov, 61166, Ukraine; e-mail: si@kture.kharkov.ua*

Major Fields of Scientific Research: Knowledge Classification, Systematization, Elicitation, Acquisition and Modeling, Knowledge Management, Ontological Engineering, Systemological Analysis, Knowledge Research and Application, Decision Making, Knowledge-Based Systems and Technologies, Artificial Intelligence, Business Intelligence, Modern (e-) Learning, Competitive Intelligence, Cognitive Modeling, Intellectual Capital .

Andrey Danilov *– Social Informatics Department, Kharkov National University of Radio Electronics, Engineer, Lenin Ave., 14, Kharkov, 61166, Ukraine; e-mail: si@kture.kharkov.ua, Skil06@ukr.net*

Major Fields of Scientific Research: Social Networks, Ontological Engineering, Competitive Intelligence, Decision Making, Intelligence Technologies, Knowledge Research and Application, Knowledge Management, (e-) Learning, Artificial Intelligence, Systemological Analysis, Social Capital.

SYSTEM OF INTELLIGENT SEARCH, CLASSIFICATION AND DOCUMENT SUMMARISATION FOR INTERNET PORTAL

Vyacheslav Lanin, Dmitriy Tsydvintsev

Abstract: *The article presents a description of alleged approaches to the implementation of data processing subsystem on Internet portal. Main problems are connected with exponential growth in number of documents, lack of semantic indexing and unstructured nature of information. In proposed approach, user receives an effective intelligent means of finding electronic documents on the basis of semantic indexing, automatic classification and cataloging of documents with construction of semantic links between them and automatic summarization of documents with the use of knowledge. The proposal is to increase the effectiveness of working with electronic documents with the help of intelligent analysis, for which agent-based and ontological approaches are used. In accordance with the proposed approach, ontology is used to describe data semantics of the document and its structure. Ontology is a central concept in process of document analysis. Through the use of ontologies the required data can be obtained; we know where to find information and how it can be interpreted. Ontology Repository contains three levels of ontologies. At the first level there are ontologies describing objects which are used in a particular system and which take into account system features. At the second level there are objects described in terms of the first level that are invariant to the domain. Objects of the third level describe the most general concepts and axioms, by which the lower levels objects are described. The third and second levels can be divided into two parts: a description of structures and description of the documents themselves.*

Keywords: *ontology, agent, multi-agent systems, intelligent search, semantic indexing, document analysis, adaptive information systems, CASE-technology.*

ACM Classification Keywords: *H.2. Database Management: H.2.3. Languages – Report writers; H.3.3. Information Search and Retrieval – Query formulation.*

Introduction

An exponential growth in the number of electronic documents is currently underway, and it clearly shows that traditional mechanisms for processing of electronic documents cannot cope with needs of user. This trend is evident both on Internet and in corporate networks. Currently, so-called information portals (thematic and corporate) become more and more popular, and their main objective is a consolidation of information and knowledge.

One of these solutions is a research portal – information-analytical system for the collection and analysis of data about innovation activity of regions to support effective management decisions (Research portal "Innovative development of regions"). Data for analysis is extracted from heterogeneous unstructured or semistructured data sources, in particular, Internet resources, as well as operational databases. According to the plan, the system must provide integration, coordination, aggregation and maintenance of previously disconnected data. Also it should support the various forms of data visualization and analysis, customized to the needs of users. From this it follows that the search and processing of unstructured text data from different sources in different formats, is becoming one of the main functions of the system under development.

Thus, the relevance of the problem is caused by the following reasons:

- Exponential growth in the number of documents that make it impossible to process data by traditional methods without loss of quality;
- Lack of semantic indexing, which does not allow for intelligent document processing in full;
- Unstructured nature of the information; the traditional mechanisms of its processing and analysis can't be used.

Consider these problems in more detail.

The exponential growth of information contained in the Internet is the reason for continuing increase in difficulty of finding relevant documents (Fig.1) and organizing them into a structured within the meaning of storage [6]. It becomes more and more difficult for user to find the necessary information; traditional search engines become ineffective.

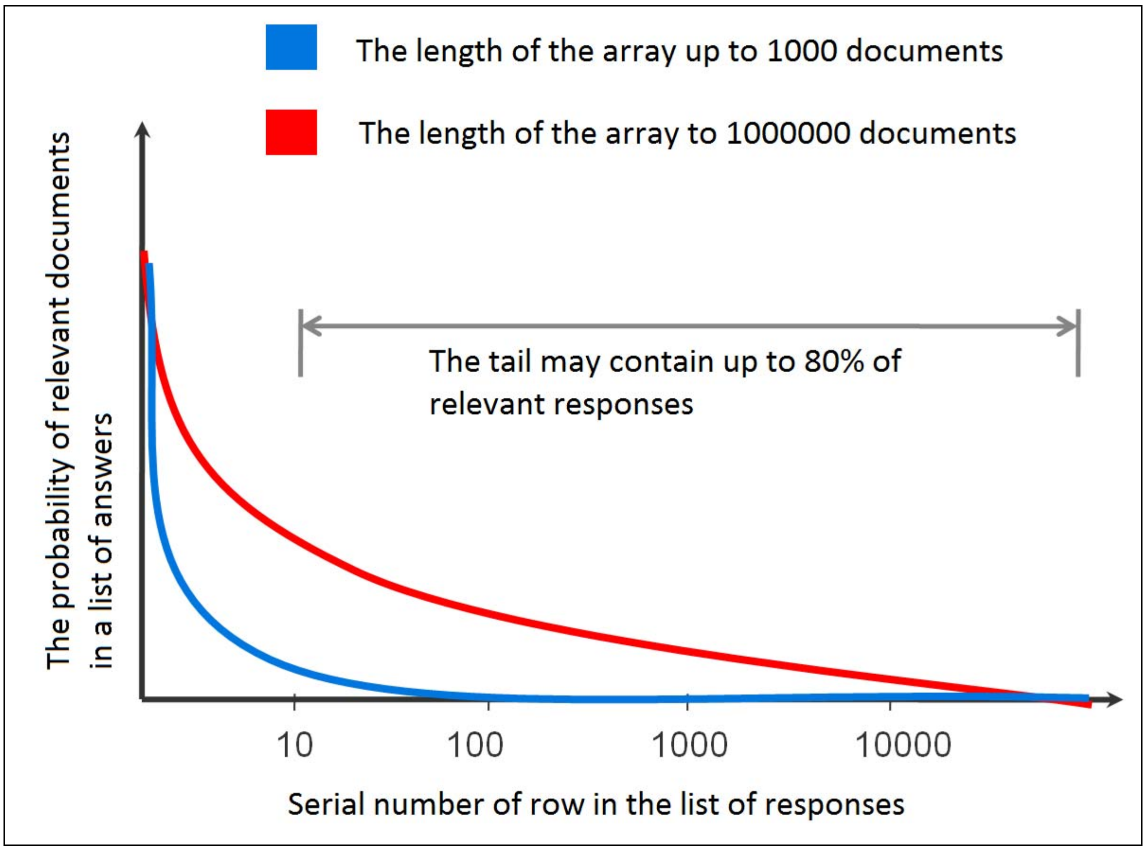


Fig.1. A problem of information retrieval with an increase in the number of documents

Most technologies of working with documents focus on the organization of effective work with information for the person. But often ways to work with electronic information simply copy methods of working with paper-based information. In a text editor, there are wide range of different types of text formatting (presentation in human readable form), but little or no ability to transfer the semantic content of the text, i.e. *no semantic indexing*. To effectively address the search problem we need to expand our notion of a traditional document: *the document should be linked with knowledge to interpret and process the data stored in the document*.

Unstructured information constitutes a significant part of modern electronic documents (Fig. 2). Data Mining systems work with structured data. Unstructured content requires using Text Mining systems. In fact, they solve the same problem for different types of data, so it is assumed that these systems will converge in a "single point".

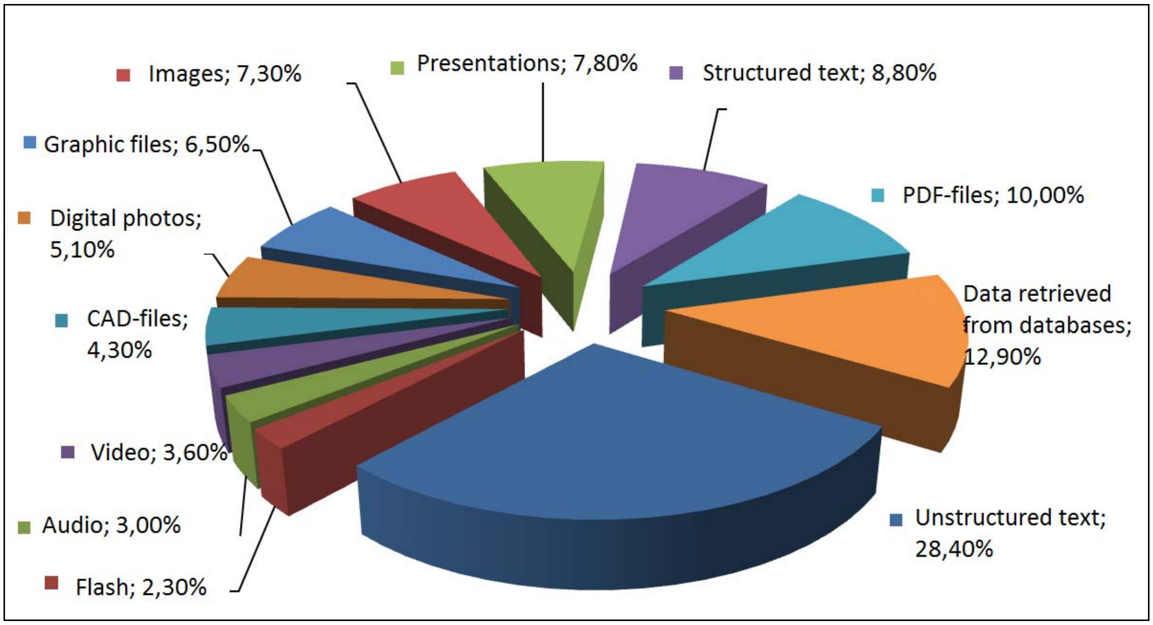


Fig. 2. Distribution of documents' categories

Text Mining allows to identify previously unknown relationships and correlations in the existing text data [5]. An important task of Text Mining technology is to derive from the text its characteristic elements or properties that can be used as document metadata, keywords or annotations. Another important task consists in assigning the document to certain categories of a given scheme of systematization. Text Mining also provides a new level of semantic search of documents. The possibilities of modern Text Mining systems can be applied in knowledge management to identify patterns in the text, to automatically "push" or post information on profiles of user interest, to create surveys of documents.

Tools and approaches of Text Mining will help to implement the intellectual capabilities of the portal when working with electronic documents.

The Approach to Semantic Indexing

Human factor has a great influence on the effectiveness of search process. User often is not ready for a long waiting of search results, for viewing and analyzing large amounts of resulting sample. In addition, most users ineffectively use search software, and usually, they ignore the advanced search capabilities, and make only short types of queries. Improvement of the electronic documents processing requires the availability of metadata describing the structure and semantics of documents. One of possible approaches to the description of information embodied in the document is an approach based on ontologies. Ontology is a knowledge base of a special type, which can be "read" and understood, alienated from the developer and / or physically separated by its users [4].

As an approach to semantic indexing has been chosen ontological approach [1], in which the ontology can describe both structure and content of the document, i.e. ontology is used to describe the semantics of the document data and its structure. Given the nature of solved targets in this paper we will concretize the notion of ontology. We assume that *ontology is a specification of certain domain*, which includes a glossary of terms (concepts) domain and a set of connections between them which describe how these terms relate to each other in a particular subject area. In fact, in this context *ontology is a hierarchical conceptual framework of the subject area*.

Ontology of document is used to analyze the document; due to it the required information can be obtained from the document: we know where to search for data and how it can be interpreted. If you represent documents using ontologies, the problem of matching ontologies and existing document is reduced to problem of search ontology terms in the document. As a result, the system needs to answer the question: Does this ontology describe the document or not. The latter question can be answered in the affirmative, if in the comparison process all the concepts included in ontology are found in document. Thus, the initial problem reduces to problem of finding general concepts in the text on the basis of formal descriptions.

Ontology repository contains *three levels of ontologies*. At the first level there are ontologies describing the objects used in a specific system and taking into account its peculiarities. The second level describes the objects that are invariant to the subject area. Objects of the second level are described in terms of objects of the first level. This is reflected in the relations of inheritance and metonymy. The objects of the third level describe the most general concepts and axioms, by which objects at the lower levels are described. The third and second levels can be divided into two parts: the description of structures and description of the documents themselves, with the documents described in terms of structures.

To address the problem of allocation of general concepts an agent-based approach is proposed on the basis of formal description [2]. This approach will satisfy the requirements of the search process, if all the advantages of multi-agent systems are realized in process of construction of the system.

When using this approach, for each node of the ontology, which contains a general concept, an agent is created which looks for this particular concept. In this approach, the agent is considered as a system aimed at achieving a particular purpose, capable of interaction with the environment and other agents. To be intelligent, the agent should have a knowledge base. Thus, to identify active agents in the system, you must choose a way to describe the knowledge base, the nature of interaction with the environment and cooperation.

Knowledge base of agent for finding find common concepts of the ontology can be also conveniently presented in the form of ontology. To enable the user to add new templates it is necessary to select the basic concepts for the formation of general ones.

One of the most important properties of agents is *sociality*, or the ability to interact [2]. As mentioned above, the agent is created for each node of the ontology, which contains a general concept. According to the accepted classification of agents it is *intentional agent*.

This agent is designed to address two problems:

1. It breaks the entire list of available templates concepts into separate components and runs simple search agents for searching of derived components.
2. Assembles results from all the lists submitted by agents of the lower level.

The agents at a lower level mentioned above are called *reflex agents*. They get a template, and their goal becomes finding phrases in the text covered by this template. Search results for agents of all levels shall be recorded on the "bulletin board".

At this point in other systems the instruments of ontological nature are used in the following areas:

- WordNet in conjunction with the vector and Boolean models of information retrieval;
- Traditional information retrieval thesauri in combination with various statistical models;
- Thesaurus for automatic indexing in Boolean models of documents searching, in problem of automatic headings and automatic annotation.

Ontologies will form a core of portal metadata when working with electronic documents. Clearly defined subject area allows creating sufficiently detailed ontologies, which can be used by all its subsystems.

Automatic Abstracting

Currently, there are two approaches used for automatic summarization. A traditional approach (quasi abstracting), which is used by such systems as Microsoft Office, IBM Intelligent Text Miner, Oracle Context, is based on allocation and selection of text fragments from the source document and connection them in a short text. On the other side, there is approach based on knowledge, which involves preparation of summaries and transfers the basic idea of the text, perhaps even in other words.

Quasi abstracting is based on the allocation of specific fragments (usually sentences). For this purpose, a method of comparison of phrasal templates chooses blocks with the greatest lexical and statistical relevance. There are a model of linear weights is used in most implementations of the method. The analytical phase of this model is a procedure of appointing the weighting coefficients for each block of text in accordance with such characteristics as location of this block in the original, frequency of appearance in the text, frequency of use in key proposals, as well as indicators of statistical significance. So, there are three main directions, often used in combination: statistical methods, positional methods and indicator methods.

The main advantage of this model lies in the simplicity of its implementation. However, the selection of sentences or paragraphs, not taking into account the relationship between them, leads to the formation of disconnected essays. Some proposals may be omitted, or there can be "hanging" words or phrases in them.

To implement the second method, some ontological reference is needed, reflecting the views of common sense and the concepts of targeted subject area, to make decisions during the analysis and to determine the most important information.

Method of forming a summary suggests two basic approaches.

The first approach relies on the traditional linguistic method of parsing sentences. This method is also uses semantic information to annotate parse trees. Comparing procedures directly manipulate the trees to remove and rearrange the parts, for example, by reducing the branches on the basis of certain structural criteria, such as brackets or embedded conditional or subordinate sentences. After this procedure, the parse tree is greatly simplified, becoming, in essence, a structural "squeeze" of the original text.

The second approach for compiling a summary roots in artificial intelligence systems and relies on natural language understanding. Parsing is also part of such a method of analysis, but the parse tree in this case is not generated. On the contrary, there are conceptual representative structures of all the initial information are formed, which accumulate in the text knowledge base. As the structures, formulas of predicate logic or such representations as a semantic network or frame set can be used.

Automatic summarization is necessary for the developed portal. When user is searching, it is necessary to present him a document annotation, by which he can decide on the usefulness of this document.

Classification and Cataloging of Documents

The task of automatic classification and cataloging of documents is the task of partitioning the incoming stream of text into thematic substreams according to predetermined headings. Automatic cataloging of electronic documents, and documents posted on Internet in particular, is complicated because of the following reasons [8]:

– A large array of documents;
– An absence of special structures for tracking the emergence of new documents;
– Optionality of the author's classification of electronic documents (as opposed to print publications) through annotation, attribution of the qualifier codes, etc.;
– A problem of tracking changes in documents.

As for automatic abstracting, there are two opposite approaches to cataloging. *The methods based on knowledge* are the most effective, but it's difficult to implement them. When cataloging the texts on the basis of knowledge preformed knowledge bases are used. They describe language expressions, corresponding to a particular category, and rules for the selection of headings [5]. Another class of methods for automatic categorization of texts is *the methods of machine learning*, which can use manually pre-cataloged texts as training examples.

When implementing a system of automatic cataloging of the portal, it is necessary to solve two problems:

– *Establishment of a mechanism for introduction and description of categories*, as some expression on the basis of words and terms in documents. The problem can be solved on the basis of expert descriptions of categories or on the basis of machine learning methods with the help of pre-cataloged collections of documents.
– *Analysis of linguistic material and context of words' using*. It requires an extensive knowledge of the language and subject area.

Conclusion

The above approaches are used in the development of an electronic document management subsystem of research portal. Its distinctive feature is focus on the explicit knowledge representation by using ontologies. This approach will allow us to realize intelligent services for searching and processing of electronic documents related to the portal and gathered from different sources.

As a result, the following tasks will be solved by creating the research portal:

– Semantic indexing of documents and intelligent retrieval of data corresponding to users' queries and the specific subject area;
– Extracting information from unstructured documents;
– Intellectual classification and cataloging and automatic summarization of retrieved documents;
– Maintaining a history of electronic documents.

The implementation of the subsystem will significantly reduce complexity to find useful information, its analysis and possible use in research.

References

[1] Lanin V. Intelligent management of documents as the basis for the technology of adaptive information systems // Proceedings of the International Scientific-Technical Conference «Intelligent systems» (AIS'07). V. 2 / M.: Fizmatlit, 2007. P. 334-339.

[2] Tarasov V. From multi-agent systems to intelligent organizations: philosophy, psychology, computer science. M.: Editorial, URSS, 2002.

[4] Khoroshevskii V., Gavrilova T. Knowledge Base intelligent systems. Petersburg.: Peter, 2001.

[5] Lande D. Knowledge Search on the Internet. Professional work. M.: Publishing house "Williams", 2005.

[6] Efremov V. Search 2.0: fire on the "tail" // Open systems. DBMS № 08 (134), 2007.

[7] Chernyak L. Enterprise Search 2.0 // Database. - 2007. - № 07 (133).

[8] Fedotov A, Barakhnin V. Internet Resources as an object of scientific research [electronic resource. - 2007. - Mode of access: http://www.rfbr.ru/pics/28320ref/file.pdf.

[9] Weal M.J., Kim S., Lewis P.H., Millard D.E., Sinclair P.A.S., De Roure D.C., Nigel R. Ontologies as facilitators for repurposing web documents / Shadbolt. Southampton, 2007.

Authors Information

Vyacheslav Lanin – Perm State University, Department of Computer Science; Russia, Perm, 614990, Bukirev St.,15; e-mail: lanin@psu.ru.

Dmitriy Tsydvintsev – Perm State University, Department of Computer Science; Russia, Perm, 614990, Bukirev St.,15; e-mail: akinokinos@yandex.ru.

30 years anniversary from the firs publication

GROWING PYRAMIDAL NETWORKS

Victor Gladun, Vitalii Velychko

Abstract: *The article presents an universal approach to analysis of attributive information: growing hieratical network structures for memory organization called Growing Pyramidal Neural Networks (GPN) and Program complex CONFOR (abbreviation of CONcept FORmation) to be used in problems solving.*

Keywords: *ontology, agent,*

ACM Classification Keywords: *H.2. Database Management: H.2.3. Languages – Report writers; H.3.3. Information Search and Retrieval – Query formulation.*

Introduction

The key enabler of increase of search operations efficiency is use of network structures for modeling environments in which problems solving. Orientation to real applied environments essentially raises a level of requirements to network models. The real environments, in which the problem-solving processes operate, have some typical features, such as multicoupling, heterogeneity, hierarchiness, and dynamism. In order to give proper representation of the examined processes, the network structures, used for representing the environment, must take into account these specifics. By their construction, Pyramidal Growing Networks, proposed by Prof. Victor Gladun, were created as answer of these requirements.

Methods for solution of regularities discovery tasks based on pyramidal networks, and methods of using of the retrieved regularities for decision-making are implemented in program complex CONFOR (Abbreviation of CONcept FORmation). In the case of decision-making in risk management, the described objects are assigned to specific disasters and/or emergency situations. This makes it possible to apply universal approach of growing pyramidal networks to analysis of attributive risk management and disaster emergencies.
This chapter is written by.

Intelligent systems memory structuring

Formation of intelligent systems memory structure needs to be done simultaneously with perception of information and under the impact of the information perceived and already stored. The memory structure reflects the information perceived. Information structuring is an indispensable function of memory. [Gladun, 2003]

The main processes of structuring include formation of associative links by means of identifying the intersections of attributive representations of objects, hierarchic regulation, classification, forming up generalized logical attributive models of classes, i.e. concepts.

Under real conditions of information perception there is often no possibility to get whole information about an object at once (for example, because of faulty foreshortening or lighting during the reception of visual information). That is why the processes of memory formation should allow for the possibility of "portioned" construction of objects models and class models by parts.

In different processes of information processing, objects are represented by one of the two means: by a name (convergent representation) or by a set of meanings of attributes (displayed representation). The structure of memory should provide convenient transition from one representation to another.

Systems, in which the perception of new information is accompanied by simultaneous structuring of the information stored in memory, are called self-structured [Gladun et al, 2008]. Self-structuring provides a possibility of changing the structure of stored in memory data during the process of the functioning because of interaction between the received and already stored information.

The building of self-structured artificial systems had been proposed to be realized on the basis of networks with hierarchical structures, named as "*growing pyramidal networks*" (GPN) [Gladun et al, 2008]. The theory as well as practical application of GPN was expounded in a number of publications [Gladun, 1987], [Gladun, 1994], [Gladun, 2000], [Gladun and Vashchenko, 2000].

Pyramidal network is a network memory, automatically tuned into the structure of incoming information.

Unlike the neuron networks, the adaptation effect is attained without introduction of a priori network excess. Pyramidal networks are convenient for performing different operations of associative search. Hierarchical structure of the networks, which allows them to reflect the structure of composing objects and gender-species bonds naturally, is an important property of pyramidal networks. The concept of GPN is a generalized logical attributive model of objects' class, and represents the belonging of objects to the target class in accordance with some specific combinations of attributes. By classification manner, GPN is closest to the known methods of data mining as decision trees and propositional rule learning.

GPN realization has following stages:

- building the structure of a network for some initial set of objects, assigned by attributive descriptions;
- training the structure, with a purpose to allocate its elements, allowing classifying all objects of the initial set;
- recognizing the belonging to some class of objects of certain object, which not belongs to initial set of objects.

The growing pyramidal networks respond to the main requirements to memory structuring in the artificial intelligent systems [Gladun, 2003]:

- in artificial intelligent systems, the knowledge of different types should be united into net-like structure, designed according to principles common for all types of knowledge;
- the network should reflect hierarchic character of real media and in this connection should be convenient for representation of gender-type bonds and structures of composite objects;
- obligatory functions of the memory should be formation of association bonds by revealing intersections of attributive object representations, hierarchic structuring, classification, concept formation;
- within the network there should be provided a two-way transition between convergent and displayed presentations of objects.

The research done on complex data of great scope showed high effectiveness of application of growing pyramidal networks for solving analytical problems. Such qualities as simplicity of change introduction the information; combining the processes of information introduction with processes of classification and generalization; high associability makes growing pyramid networks an important component of forecasting and diagnosing systems, especially in the area of GMES. The applied problems, for solving of which GPN were used are: forecasting new chemical compounds and materials with the indicated properties, forecasting in genetics, geology, medical and technical diagnostics, forecasting malfunction of complex machines and sun activity, etc. Special kind of applications is aimed to support intelligent data processing in GMES.

GPN theoretical foundations

The word "object" here and further is understood in a broad sense is there can be a real physical object, some process, a situation, etc.

The model of classes of the objects, used for the decision of tasks of classification, diagnostics and forecasting, should include all the most important attributes describing a class. The model also should display for this class the characteristic logical connections between essential attributes. Therefore, the basic attention concentrates on formation of the generalized logical multivariate models of objects classes. Such models, in fact, are the concepts that correspond to the classes of objects [Voyshvillo, 1967], [Gorskii, 1985].

The concept is usually defined in logic as an idea that reflects essence of objects. Most of used concepts are the result of generalization of attributes that characterizes the objects of the class.

The concept "attribute" can be used for characterizing the objects and can be used in such logic operations as extraction, recognition, identification, etc. It is necessary to note, that separation of attributes on essential and unessential is conditional and depends on problems for which decision they are used.

1 Concept

From a philosophical point of view, the concept consists of two parts – extensional and intensional:

– extensional part covers all instances belonging to this concept;

– intensional part includes all the properties that representative to these instances.

The connections between instances and their attributes play an important role in determining the hierarchical relationship between concepts and attributes.

The set of the instances, generalized in concept, constitute the volume of defined concept. In system of knowledge, the concepts play a role of base elements for composing propositions and other logic forms of thinking. Transition from a sensual step of cognition to abstract thinking, actually, means transition from reflection of the world in the form of perception and presentations to its reflection in concepts.

Classification, generalization, structuring of the perceived information, its inclusion in system of knowledge, is carried out based on an available set of concepts. In these processes two basic functions of concepts: recognition and production of the elements models of the world, in which the bearer of knowledge operates, are realized. Recognition processes became for a long time object of research and automation while production of models for the present is little-investigated problem. Production of models plays the important role in creative activity. Only by a concrete definition of concepts we can create (for example, to draw) images of concrete houses, trees, cars, etc. Production of elements models of the world underlies designing of the engineering objects.

Attributes belonging to concept by their role in realization of the basic functions of concept divide on two types – disjunctive and unified:

– disjunctive attributes are attributes, which do not occur or occur seldom in concepts volume. These attributes are most effective at realization of recognition functions;

– unified attributes are those attributes, which are inherent in all or many elements of concept volume refer to, but they can be widespread and outside of concept volume. Without these attributes, the production of elements models of the world is impossible.

For example, for all birches such attributes, as presence of a trunk, roots, and crown are characteristic. It their unifying attributes which are inherent in all trees. A well-known disjunctive attribute of birches is white color of a bark.

The degree of detailed elaboration of model created based on concept depends on the purpose of a task. The model of the bridge created at the decision of a task "to draw a bridge", essentially differs on a degree of detailed elaboration from the model of the bridge created at the decision of a task "to design the bridge".

The success of the decision of the problems including production of models depends on that, how much used concepts correctly and full characterize corresponding classes of elements of the world.

Now, it is possible to give more constructive definition of concept, more suitable by consideration of information-technical aspects of problems of formation and processing of concepts.

Concept – an element of the knowledge system, representing the generalized model of some class of instances.

In processes of recognition and production of models, the concept is used as logic function of the attributes, having the value "true" for instances from volume of concept and value "false" in other cases.

2 System of concepts

The set of concepts included in the system of knowledge, will be called the system of concepts of knowledge bearer.

Systems of concepts are hierarchical, as a rule. Volumes of concepts of all levels of hierarchy, except for bottom, are formed by consolidation of volumes of some concepts of lower levels. For example, the volume of concept "fruit" unites volumes of concepts "apple", "pear", etc.

Systems of concepts are dynamical. The structure of concepts varies because of interaction of their bearers with an environment, and during the decision of problems.

At each moment of time, the state of system of concepts reflects individual experience of its bearer. Therefore, separate concepts and systems of concepts in the whole are subjective.

Any system of concepts by virtue of the discreteness, limitation of structure of concepts, imperfection of separate concepts cannot reflect variety and a continuity of the real world. Volumes of the concepts which have been not introduced "by definition", as a rule, have no precise dividing boundaries. There are many of the transitive forms that complicate carrying out of conditional boundaries between volumes of concepts. There are many transitional forms that complicate definition of conditional boundaries between volumes of concepts.

Because of incompleteness of the world mapping in concepts system, and of concepts subjectivity, univocal identification of elements of the world based on concepts system often appears inconvenient or even impossible. Therefore, volumes of many concepts can be considered as fuzzy sets. Each bearers of concepts system possesses the membership function, which, thus, has subjective character.

3 *Inductive formation of concepts*

From logic structure point of view, the concepts are categorized as:

- conjunctive concepts, which can be described by conjunction of attributes;
- disjunctive concepts, which can be described by disjunction of conjunctions of attributes;
- concepts with the exclusive attributes, reflecting absence of some attributes in the instances, which belongs to concept volume.

The concepts, included in everyday practice, usually are conjunctive. More complex logic structure is characteristic for the concepts, formed in the research process.

In this case, the logical complexity of the concepts usually arises following circumstances:

- the attributes space is incorrectly chosen;
- training set is incomplete reflects specificity of concept volume;
- the volume of formed concept consists of instances that are vastly different from each other.

Consider a task of inductive formation of concepts for not intersected sets of objects $V_1, V_2, \ldots, V_n$, , each set represents some class of objects with known properties. Let L – the set of objects used as training set. All the objects of set L are represented by sets of attribute values. Relations $L \cap V_i \neq \varnothing$ and $V_i \not\subset L$ $\left(i = 1, 2, \ldots, n\right)$ take place. Each object from set L corresponds to one of set V_i. It is necessary to generate n concepts by analysis L. The amount of these concepts must be sufficient for correct recognition of belongings of anyone $l \in L$ to one of set V_i.

In forming the concept corresponding to set V_i, the objects of training set included in V_i, are considered as examples of set V_i, and the objects, not included in V_i, – as counterexamples of set V_i.

Each concept, generated on the basis of training set, is approximation to real concept. The proximity of concepts depends on representativeness of training set, i.e. on the detailization of the concept volume peculiarities.

Problem of inductive formation of concepts is similar to the problem of learning pattern recognition. And in both cases as a result of learning a model of a class of objects is constructed. At formation of concepts stronger requirements are made to this model (concept). It must provide not only recognition, but also the opportunity to generate models of concrete objects. In this regard, the model should be reflected attributive, structural, and logical characteristics of objects.

The training set usually has the tabular form. The rows of the table correspond to the set of objects properties, columns – to attributes. Names of classes are specified in a special column. The concept, which is formed because of the analysis of the training set, is usually described by a logical expression in which the variables are the names of the attributes values.

Known methods of formation of concepts [Gladun, 1987] [Bongard, 1967], [Vagin, 1988], [Gladun and Vashchenko, 1995], [Pospelov, 1986], [Gladun and Rabinovich, 1980], [Michalski et al, 1986], [Piatetsky-Shapiro and Frawley, 1991] as a matter of fact are methods of controlled choice of the attributes values that characterize the classes of objects. The choice can be simplified due to use of adequate representation of the analyzed information.

Requirements to the methods of concepts formation

During the work, following requirements to the methods of concepts formation are revealed:

- for increasing the reliability of the diagnosis or the forecast, it is necessary to consider dependence of the defined variable from combinations of known attributes, i.e. to consider joint simultaneous influence of attributes. The formed concept should reflect such dependences;
- depending on a choice of a method of training for concepts of the same class of objects various logic descriptions can be received. Naturally, there is a question on quality of logic models. The best results of application of concepts for classification, diagnostics, and forecasting, as a rule, correspond to more generalized concepts, i.e. concepts that are described by more simple logic expressions. Degree of complexity of logic expression can be estimated by the number of its variables. The method of concepts training should provide formation of as more as possible simple concepts;
- choice operations, such as a choice of values of properties, objects, combinations of values of signs, etc., prevail in processes of knowledge mining. It is a combinatory problem. The volume and time of choice operations quickly grows with increase amount of data. This effect of "information explosion" blocks practical application of many methods. In this connection, there is a necessity for use the network structures reducing amount of search operations at realization of processes of knowledge mining.
-

Requirements to the network structure used for the knowledge mining

The key enabler of increase of search operations efficiency is use of network structures for modeling environments in which problems solving. Orientation to real applied environments essentially raises a level of requirements to network models. We shall define the features of real environments rendering strong influence on processes of the problems solving:

- multicoupling. Real environments usually include many objects connected by a lot of relations;
- heterogeneity. For real environments the variety of objects and relations is characteristic;
- hierarchy. In real environments it is necessary to operate with the compound objects representing compositions of more simple objects;
- dynamism. Real environments are usually subject to frequent changes.

Given the above features of real environments, we formulate requirements to the network structure, representing the environment.

- the network should possess the developed associative properties, i.e. to provide effective performance of various search operations;
- the network should reflect the hierarchy of real environments and therefore should be convenient to represent genus-species relations and structures of composite objects;
- in a network the means limiting zones of search by time, spatial or substantial criteria should be stipulated, i.e. the network should provide selectivity of search on a time, spatial or meaningful context;
- at construction of a network the classes of objects and situations should be formed; input of the new information into the network must be accompanied by the classification process;
- the network should allow parallel execution of search operations.

Pyramidal networks

The set forth above requirements are answered to the full with pyramidal networks [Gladun et al, 2008].

A growing pyramidal network (GPN) is an acyclic oriented graph having no vertexes with a single incoming arc. Examples of the pyramidal networks are shown in figures below. Vertices having no incoming arcs are referred to as *receptors*. Other vertices are named *conceptors*. The *subgraph* of the pyramidal network that contains vertex a and from all the vertices that belong to subgraph there are paths to vertex a is named the pyramid of vertex a. The set of vertices contained in the pyramid of vertex a is referred to as the *subset* of vertex a. The set of vertices reachable by paths from vertex a is named the *superset* of vertex a. The set of vertices that are connected with paths to vertex a, is referred to its superset.

In subset and superset of the vertex, *0-subset* and *0-superset* are allocated, consisting of those vertices, which are connected to it directly.

When the network is building, the input information is represented by sets of attributes values describing some objects (materials, states of the equipment, a situation, illness etc.). Receptors correspond to values of attributes. In various tasks, they can be represented by names of properties, relations, states, actions, objects or classes of objects. Conceptors correspond to descriptions of objects in general and to crossings of descriptions and represent GPN vertices.

Building of GPN

Initially the network consists only of receptors. Conceptors are formed as a result of algorithm of construction of a network. After input of object attribute description, corresponding receptors switch to a state of excitation. The process of excitation propagates through the network. A conceptor switches into the state of excitation if all vertices of its 0-subset are excited. Receptors and conceptors retain their state of excitation during all operations of network building.

Let F_a be the subset of excited vertices of the 0-subset of vertex a; G be the set of excited vertices in the network that do not have other excited vertices in their supersets. New vertices are added to the network by the following two rules:

Rule A1. If vertex a, that is a conceptor, is not excited and the power of set F_a exceeds 1, then the arcs joining vertices of set F_a with the vertex a are liquidated and a new conceptor is added to the network which is joined with vertices of set F_a by incoming arcs and with the vertex a by an outgoing arc.

The new vertex is in the state of excitation. Rule A1 is illustrated in Figure 1(a,b). According to the Rule A1, the condition for adding a new vertex to the network is a situation, when certain network vertex is not completely excited (at least two vertices of 0-subset are excited). Figure 1(a) shows a fragment of network in some initial state. Receptors 4, 5 switch to a state of excitation, the network switches to state (b), and a new vertex appears – a new conceptor. Receptors 2, 3 switch to a state of excitation additionally. The network switches to state (c).

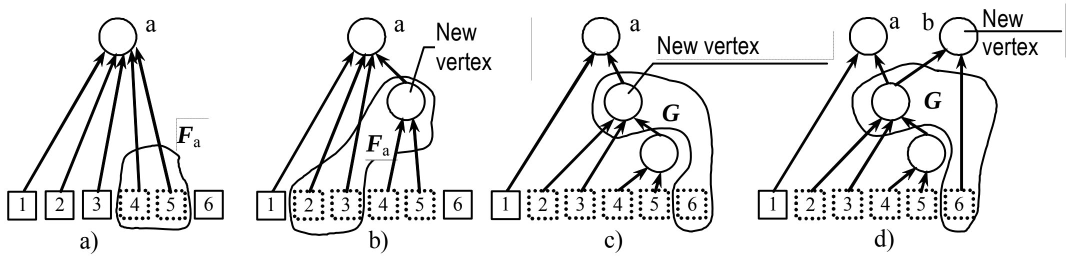


Figure 1. Pyramidal network building

New vertices are inserted in 0-subset of vertices, which are not completely excited. New vertices correspond to intersection of object descriptions, represented by incoming arches. Once new vertices have been introduced into all network sections where the condition of rule A1 is satisfied, rule A2 is applied to the obtained network fragment, concluding the object pyramid building.

Rule A2. If the power of set G exceeds 1 element, a new conceptor is added to the network, which is joined with all vertices of set G by incoming arcs.

The new vertex is in the state of excitation. Rule A2 is illustrated in Figure 1(c,d). The state (d) was obtained after the excitation of receptors 2-6.

In applying the Rule A1 the main cross-linking relation is a relation of intersection of receptor set, excited by input of the object description and other sets of receptors included into pyramid, recently formed by conceptors. Rule A2 concludes the building of pyramid, which represents complete description of the introduced object.

1 Properties of a pyramidal network

Let us note some properties of pyramidal networks.

Depending on applied area in which networks are used, the receptor can represent the attribute value, the elementary fact from the description of a situation, value of an economic parameter, a symptom of illness, the letter, a word, etc. Conceptors correspond to descriptions of objects, situations, realizations of processes or the phenomena, words, phrases, plans, and also crossings of descriptions.

The pyramidal network is the network memory, which is automatically customized on structure of the input information. Optimization of the information representation due to adaptation of network structure to structural features of entrance data is as a result reached. In addition, unlike neural networks, the effect of adaptation is reached without introduction of aprioristic redundancy of a network.

Pyramidal networks are convenient for execution of various operations of associative search. For example, it is possible to select all the objects that contain a given combination of attribute values by tracing the paths that outgo from the network vertex corresponding to this combination. To select all the objects whose descriptions intersect with the description of a given object it is necessary to trace the paths that outgo from vertices of its pyramid. Rules A1, A2 establishes associative proximity between objects having common combinations of attribute values.

Hierarchical organization is an important property of pyramidal networks. This provides a natural way for reflecting the structure of composite objects and generic-species interconnections.

The algorithm of a network building provides an automatic establishment of associative affinity between objects based on common elements and their descriptions. All the processes connected with a network building at processing one description are localized in rather small part of a network – a pyramid corresponding to this description. The important property of semantic networks of pyramidal structure is their hierarchy allowing naturally mapping structure of compound objects and genus-species relations.

Conceptors of the network correspond to combinations of attribute values that define separate objects and conjunctive classes of objects. By introducing the excited vertices into the object pyramid, the object is referred to classes, which descriptions are represented by these vertices. Thus, during the network building the conjunctive classes of objects are formed, the classification of objects is performed without a teacher. Classifying properties of pyramidal network are vital for modeling environments and situations.

Profitability is also the advantage of pyramidal networks, because identical combinations of attributes values several objects are represented in network by one common pyramid.

In a pyramidal network, the information is stored by its mapping in structure of a network. The information on objects and classes of objects is presented by vertex ensembles (pyramids) distributed on all network. Entering of the new information causes redistribution of links between vertexes, i.e. change of network structure.

Certainly, the full advantages of pyramidal networks are appeared at their physical realization supposing parallel distribution of signals on a network. The important property of a network as means of storage of the information is that the opportunity of parallel distribution of signals is combined with an opportunity of parallel reception of signals on receptors.

This property appears useful at applications of pyramidal networks in robotic systems, the automated systems of scientific researches, systems of the automated designing. Conversion from converged representation of objects (conceptors) to expanded (sets of receptors) is performed by scanning pyramids in top-down and down-top directions.

2 Concept formation in a pyramidal network

Training GPN consists in formation of the structures representing concepts, on a basis of attributive descriptions of the objects incorporated into classes with known properties.

Concept is an element of knowledge system, representing generalized logic attributive model of objects class. This model is used in processes of objects recognition. The set of objects generalized in concept is its volume.

The combinations of attributes allocated in ready-built pyramidal network, representing descriptions of objects of training set, are used as "a building material", a basis of further logic structure of concept.

Let L be the pyramidal network representing all of training set objects. For formation of concepts $A_1, A_2, \ldots, A_i, \ldots, A_n$ corresponding to sets $V_1, V_2, \ldots, V_i, \ldots, V_n$, pyramids of all objects of training set are scanned in order. The vertices of scanned pyramid during its scanning are considered excited. Special vertices in network are identified in order to recognize objects from the concept volume. They are referred to as check vertices of a certain concept. At performance of inductive generalization, it is natural that the most important attribute or the combination of attributes describing group of objects – concept A_i, those vertex from pyramid A_i, which meet in pyramid A_i more often. Such attributes (or their combinations) are necessary for noting as check vertices. Check vertices are used in the further at decision about belonging of a new object to the concept. If in a pyramid of concept A_i there are some vertices, which include into equal quantity of objects from the given concept volume, it is natural to choose from the given vertex such, which unites maximal quantity of attributes (receptors) from a concept pyramid. This vertex defines the most typical combination of attributes of the objects incorporated into concept. In selecting the check vertexes, two characteristics of network vertices are used:

– $\{m_1,\ldots,m_i,\ldots m_n\}$, where m_i $(i = 1,2,\ldots,n)$ is a number of objects of volume of concept A_i, which pyramids include the given vertex;
– k is the number of receptors in the pyramid of this vertex.

For receptors $k = 1$. While scanning, the pyramid is transformed by the following rules:

Rule B1. If in the pyramid of an object from concept volume A_i, the vertex, having the largest k among all the vertices with the largest m_i, is not a check vertex of concept A_i, then it is marked as a check vertex of the concept A_i.

The rule allows existence several vertexes among the excited vertexes with identical m_i, exceeding m_i of other excited vertexes. If in group of the vertexes having largest m_i, values k of all vertexes are equal, any of vertexes can be marked as check vertex of concept A_i.

The rule B1 is illustrated in Figure 2. In this situation, vertex 6 is selected as check vertex, because it has the largest k among vertices with the largest m_i (6, 13, 14). Values m_i are shown inside symbols of vertices.

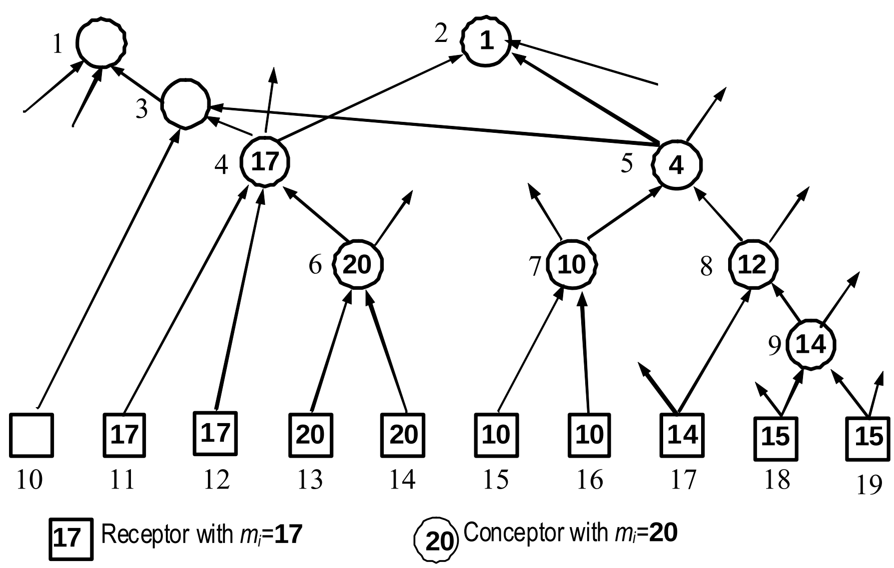


Figure 2. Forming of pyramidal growing network – rule B1

Rule B2. If the pyramid of an object from concept volume A_i contains check vertices of other concepts whose supersets do not contain excited check vertices of concept A_i, then in each of these supersets the vertex, having the largest k among all excited vertices with the largest m_i, is marked as a check vertex of concept A_i.

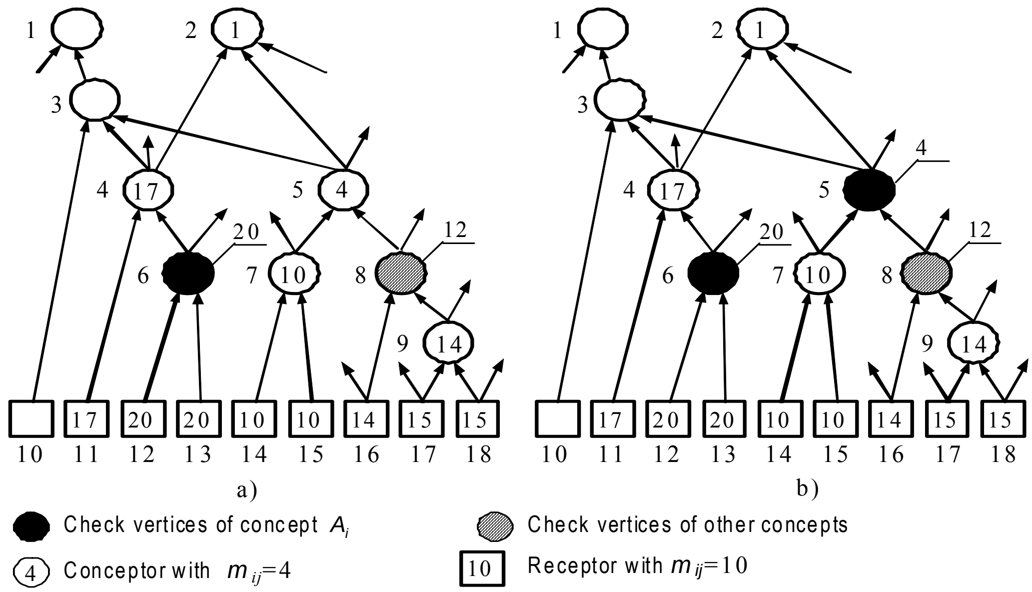


Figure 3. Forming of pyramidal growing network – rule B2

According to this rule the excitation of the pyramid of vertex 2 (Figure 3a) on the condition, that it represents an objects from concept volume A_i, results in choosing vertex 5 as the check vertex of concept A_i (Figure 3b). By check vertexes we select the most typical (having the largest m_i) combinations of attribute values, belonging to objects from concept volume. For example, selecting the vertex 8 as a check vertex means selection of combination of value attributes, corresponding to receptors 17, 18, 19.

If at least one new check vertex appears while scanning objects of the training set, i.e. conditions of Rules B1 or B2 have been performed once at least, the training set is rescanned. The algorithm stops if during the scanning of the training set no new check vertex appears.

Recognition on basis of GPN

The task of recognition is based on the following rule.

Certain object belongs to the concept volume A_i if its pyramid has check vertexes A_i and does not contain check vertices of any other concept not having excited check vertices of concept A_i concept in their supersets. If this condition does not hold for any of the concepts, the object is referred to as unrecognized.

The execution time of the above algorithm is always finite. If the volumes of the formed concepts $V_1, V_2, \ldots, V_i, \ldots V_n$ do not intersect, than after execution the algorithm the recognition rule completely divides the training set into subsets $L_i = V_i \cap L \left(i = 1 \ldots n \right)$

The formed concepts are represented in the network as ensembles of check vertices.

There is an algorithm [Gladun, 1987] of composing the logic descriptions of concepts, formed in the network as a result of the training process, described above. The formed logical expression contains logical relations, represented by allocation of check vertices, describing the concepts in the network, defining different classes of objects.

For example, the concept, presented on Figure 3b check vertices with numbers 5 and 6, by following expression is described:

$$(12\wedge13)\vee(14\wedge15)\wedge\neg(16\wedge17\wedge18).$$

The analytical tasks, such as diagnostics or forecasting, can be reduced to the task of classification, i.e. to belongings the research object to a class of objects, with a property characteristic or a set of properties significant for prognosis

Classification of new objects is performed by comparing the attribute descriptions with the concept, defining a class of predictable or diagnosing objects. Objects can be classified by evaluating the value of the logical expressions that represent corresponding concepts. The variables, corresponding to the attribute values of the recognized object, set 1, other variable set 0. If the entire expression possesses the value 1 which means the object is included into volume of concept.

In concept, which is formed by algorithm, the general essential attributes of objects from volume of concept and logic relations between attributes are reflected. Unifying attributes are allocated as a result of performance of rule B1. At performance of rule B2 disjunction attributes are allocated.

An important distinction of a method of concepts formation in growing pyramidal networks is the possibility to introduce in concepts the so-called excluding attributes which do not correspond to objects of a researched class. As a result, the formed concepts have more compact logic structure, which allows increasing the accuracy of diagnosis or forecasting. In logic expression the excluding attributes are presented by variables with negation.

All search operations in growing pyramidal network are limited to rather small fragment of a network, which includes an object pyramid and vertices directly linked to it. As a result, we have a possibility solve practical analytical problems based on large-scale data.

Program complex CONFOR

Methods for solution of regularities discovery tasks based on pyramidal networks, and methods of using of the retrieved regularities for decision making described in the previous section are implemented in program complex CONFOR (Abbreviation of CONcept FORmation). In the case of decision-making in risk management, the described objects are assigned to specific disasters and / or emergent situations. This makes it possible to apply universal approach of growing pyramidal networks to analysis of attributive risk management and disaster emergencies.

Let us consider briefly basic functions and structure of a program complex.

Architecture of CONFOR

The basic functions of program complex CONFOR are:

- discovery of regularities (knowledge) inherent to data;
- using of the retrieved regularities for object classification, diagnostics and prediction.

Main program unit and interrelations between subsystems are presented on Figure 4.

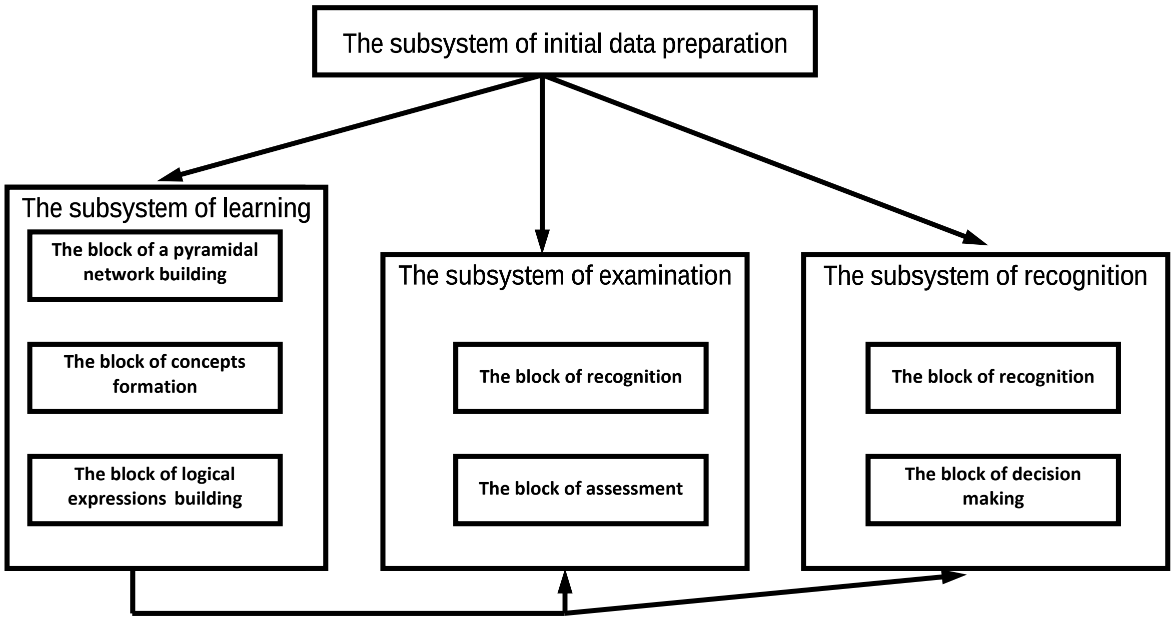


Figure 4. The structure of program system CONFOR

The program complex CONFOR includes following subsystems:

1. Subsystem of initial data preparation.
1. Learning subsystem.
2. Examination subsystem.
3. Recognition subsystem.

Initial data preparation subsystem

The subsystem of data preparation can be functionally extracted in the separate block though it is an integral part of subsystems of training, examination, and recognition. The subsystem realizes multitude of the operations, allowing entering attributive descriptions of situations directly from the screen or from the previously prepared text file.

The subsystem will transform the entered information in the form of training, examination, or recognition set, to internal representation, which is used for construction or the analysis of a pyramidal network.

Objects of the training, examination, or recognition set are represented by attribute descriptions, i.e. by sets of attribute values.

Before operating with the system, it is necessary to:

– choose the set of attributes that will be used for describing investigated objects;
– specify for every attribute a set of its values. Attribute values can be given in numerical, Boolean or nominal scales. Numerical attributes must be discretized, that is divided into subintervals;
– describe objects using the chosen attribute values, that is for every object specify its value for every attribute;
– specify for every object (of the training and examination set) the class name that the object belongs to.

The entered attribute descriptions of situations are displayed on the screen in the form of the table, owing to what it is convenient for supervising and editing.

Input attribute descriptions of situations from the screen is provided with a following set of functions: Add Column, Add Row, Delete Column, Delete Row, Rename Column.

Input attribute descriptions of situations from the previously prepared text file is carried out by means of function "Import". The test file can be prepared outside of a tool by means of the text editor in the form of a *.txt-file according to strictly certain syntax: the first line – heading; lines-descriptions of objects further follow; each line comes to an end with a symbol "line feed"; elements of a line are separated by a symbol "a comma (,)"; unknown values of attributes fall, but the divider (,) is put, i.e. empty value is designated as ","; for a designation of a fractional part of a quantitative attribute the "point (.)" is used.

After input of descriptions of situations (from the screen or from a file), the columns containing a name of object (situation) and a name of a class (type of situation) should be marked in appropriate way. With this purpose, next functions are used: to choose a column "Object"; to choose a column "Class".

The specified functions expand opportunities of a tool as research system, allowing investigating the same training set from the various points of view, easily changing an attribute, which serves as a name of a class.

Learning subsystem

Subsystem of learning realizes the discovery of the regularities that characterize the class of emergency. Input data for a subsystem of learning is served training set, which includes examples of the situations, described as various types (classes). Each class of situations should be presented by a quantity of examples; sufficient that based on them the regularity describing the given class has been allocated. Examples of situations are represented by set of attributive descriptions. Attributive description of a situation should include a name of a situation, a class to which the situation belongs, and a set of attributes values, which characterize a situation.

Output data for a subsystem of learning is served regularities, which characterize classes of disasters situations.

Process of learning consists of following stages:

– representation of initial data in the form of a pyramidal network;
– concepts formation on the basis of a pyramidal network;
– construction of the logic expressions corresponding generated concepts.

1 Block of pyramidal network building

The block of a pyramidal network building realizes the first stage of process of training when internal representation of training set objects will be transformed to a pyramidal network.

As the basic process at regularities extraction is search of combinations of the attributes values describing groups of similar objects, as a rule, it is necessary to look through repeatedly objects training set that with growth of objects number and of number of attributes values leads to "information explosion". Representing of data in the form of a pyramidal network allows to avoid this danger, first of all, due to features of algorithm of the network building providing allocation of the common combinations of attributes values during input of objects in a network, and also due to associativity and hierarchy of the network. Only two viewings of training set are necessary for full a network building.

The pyramidal network is dynamic structure, which is restructured depending on the incoming information. The network consists of two types vertices: receptors and conceptors. Receptors correspond to values of attributes. In various tasks, they can be represented by names of properties, relations, states of the equipment, a situation, actions, objects, or classes of objects. Receptors have no input connections.

Conceptors correspond to crossings of objects descriptions and to descriptions of objects in general and represent GPN vertices. Conceptors of the first type is named as intermediate vertices and have input and output connections. Conceptors of the second type – the main vertices, which can have output connections only in that case when the description of the object, corresponding the given main vertex, is a part of the description of some other object.

The main process of a pyramidal network building, at which in a network the common combinations of attributes values (all intermediate vertices are building) are fixed, is realized even at the first viewing of objects descriptions of training set. At the second viewing, the network only is corrected to provide unequivocal representation of each object in the form of a separate pyramid (with one main vertex). At network building objects of training set are entered into a network serially and thus the description of the next object is compared to those objects of training set which already have been entered into a network. Comparison is carried out by tracing output connections of receptors from the description of considered object in a direction to the main vertices of pyramids. Tracing is carried out by consecutive transition from the analysis in vertex of lower level of a network to the analysis of higher level.

During tracing it is formed and corrected TBC (the table of target communications) in which pairs analyzed units in the form of "the subordinated vertex – the subordinating vertex" are fixed. The "excited" fragment of a network, which is already constructed fragment of a network, which includes receptors from the description of considered object, is as a result allocated.

Thus are fixed both completely excited vertices of a network, and partially excited. Vertex at which all input connections conduct from receptors from the description of selected object is considered as completely excited. At partially excited vertex from receptors from the description of selected object, two input connections conduct, at least. Other input connections can conduct from receptors, which correspond to the values of attributes not inherent in selected object.

If on a way of tracing of receptors output connections there are completely excited vertex, it means, that these vertex should be included in a pyramid of selected object, and process of tracing proceeds. At detection of partially excited vertex process of tracing in this direction is stopped. Process of tracing finally stops, when partially excited vertices and the "highest" completely excited vertices are revealed and fixed all. "Highest" completely excited vertices are named such completely excited vertices from which do not conduct output connections to other completely excited vertices.

As to the beginning of the second viewing in a network intermediate and main vertices are already constructed all, at this stage also calculation m_i is carried out for each vertex (m_i is the number of excitation for each i-th class). Number of excitation of vertex for some class to equally number of objects of the given class in which pyramids there is this vertex.

2 Block of concepts formation

The block of concepts formation realizes process of regularities discovery proper. This block on the basis of the analysis of the constructed pyramidal network selects from the combinations of attributes values most essential to everyone class and determines connections between them. Vertices, which correspond to the selected combinations, were named control vertices.

The formed concepts are represented in the network structure as ensembles of check vertices, which belong to corresponding classes.

Process of concepts formation is carried out by consecutive updating formed concepts on each object of training set.

In spite of the fact that during updating training set is looked through some times, this process is carried out much more quickly, than process of a network building because each time the area of viewing is limited only to a pyramid of object. Pyramids are looked through "from top to down", i.e. in a direction from the main vertices to receptors. Process ends, when there are no conditions for occurrence of new check vertices, i.e. objects of training set are divided completely.

3 Block of logical expressions building

The block of logical expressions building makes it possible to represent the generated concepts in the form of logic expressions. Each logical expression, which corresponds to concept, has as operands of attributes value by means of which situations were described, and as operators – conjunction, disjunction, and negation. Representation of concept in the form of logical expression is evident, is well interpreted, and can be used by the expert for the analysis with the purpose of deeper understanding of regularities, which are inherent in a object domain. Logical expressions' building is carried out consistently for each class, which objects are presented in training set. Generated logic expressions are written in a text file, which can be screened by a special command, or is read by means of a text editor.

After adjustment for the next class all check vertices of the given class are analyzed. Check vertices of a considered class are named by "positive" check vertices, all check vertices of other classes are considered "negative" in relation to a considered class. Construction of logic expression begins with ordering positive check vertices of a considered class in decreasing order m – their excitation numbers for the given class. Each "positive" control unit is a basis for formation of a corresponding disjunctive member of the logic expression, representing concept of the considered class.

Formation of each disjunctive member begins with even not considered check vertices with the greatest m. First of all in a text file the is written number m which corresponds to number of objects of the training set belonging the given class and containing in its description receptors, corresponding the chosen check vertex. Then in a text file the receptors of the pyramid of the selected node are written out, for which the pyramid is scanned in the direction "from top to down". In a text file receptors are bound by symbol of conjunction (&). Such conjunction we shall name base conjunction for a formed disjunctive member.

Further so-called conjunction-exceptions form for what in superset of positive check vertices search of the nearest negative check vertices is carried out. Nearest negative checked vertices are the vertices on a path to which from positive check vertices, there are no other check vertices. As negative checked vertices the vertices belonging to any other class are considered.

If those are not present in a network, in a text file the symbol of a disjunction (V) enters and formation of the next disjunctive member begins. If in superset of positive check vertex there are negative check vertices, formation of conjunction-exception proceeds by tracing out the receptors entering into a pyramid of negative check vertex, but without taking into account receptors which have entered into a pyramid of positive check vertex.

Written out receptors are united by the symbol of conjunction (V), undertake in brackets and join with earlier generated part of logical expression through symbols of conjunction (&) and negation (¬). The analysis of negative check vertices and formation of corresponding conjunction-exceptions also is carried out in order decreasing m, describing a class corresponding negative check vertex. In this case formation of a disjunctive member comes to end after the analysis of all check vertices which are negative in relation to the given positive check vertex.

After all positive check vertices of considered class are analyzed; changeover to formation of logic expression for a next class is carried out.

Examination subsystem

The subsystem of examination is intended for testing quality of a tool training and quality of training set. The quality of tool training depends from:

- qualities of tools used for training;
- a material for training, i.e. structure of objects of training set and a manner of their description.

As the algorithm of concepts formation based on a pyramidal network provides 100% division of learning set, testing of quality of tool training is reduced to check quality of training set by recognition of examination set objects. Comparison of objects recognition results of examination set to the information on a real accessory of situations to classes allows to judge about quality of a complex training.

The subsystem of examination includes following blocks:

- block of recognition analyzing situations, which do not enter into training set but for which their accessory to one of investigated classes is known;
- block of an assessment, which gives out the information on quantity of correct, wrong and uncertain answers of a subsystem.

The important feature of realizable process of recognition is the opportunity to give out uncertain answers when recognizable objects contain in the description a combination of receptors, characteristic simultaneously for different classes, or when recognizable objects are not similar to objects of training sample. A large number of incorrect and uncertain answers of a subsystem demonstrate the necessity continuation of the learning by improving of training set.

Process of recognition can be carried out both based on the analysis of the trained pyramidal network, and by means of the constructed logic expressions.

The block of an assessment compares with the results received at recognition of objects of training set, with the information on a real accessory of objects to classes and gives out a percentage of correct, wrong, and uncertain answers of a subsystem.

Recognition subsystem

The subsystem of recognition realizes second of the basic functions of a complex, namely, use of the regularities allocated in a learning stage for classification of new situations and outputting of the control decision on elimination of an unforeseen contingency.

The subsystem of recognition consists of following blocks:

- block of recognition, which allows to classify a new situation to one or another class;

- block of decision making, offering to the operator the recommendation at the choice of the operating decision with the purpose of normalization the contingency.

For recognition in a subsystem is used the same block, as in a subsystem of examination. The block of decision-making, as well as the block of assessment of examination results, can give out both exact and uncertain answers. As each class of objects of training set represents the contingency, identification of a new contingency is unequivocally connected with sequence of control actions on its normalization. Identification of an accessory of a new contingency to some class is the operator prompting. Operator makes the final decision on a choice of actions on normalization of a situation.

In case of the uncertain answer the subsystem gives out the additional information on that, how much distinguished situation is similar to the situations corresponding different classes, or absolutely not similar to situations from training set.

For an estimation of a similarity degree the function of confidence is used. This function is calculated based on the analysis of conjunctions involved in the recognition of this situation. Function of confidence reflects a percentage parity of conjunctions informativity, describing the regularities of different classes of situations.

Conclusion

The main characteristic of the pyramidal networks is the possibility to change their structure according to structure of the incoming information. Unlike the neural networks, the adaptation effect is attained without introduction of a priori network excess. Pyramidal networks are convenient for performing different operations of associative search. Hierarchical structure of the networks, which allows them to reflect the structure of composing objects and gender-species bonds naturally, is an important property of pyramidal networks. The concept of GPN is a generalized logical attributive model of objects' class, and represents the belonging of objects to the target class in accordance with some specific combinations of attributes. By classification manner, GPN is closest to the known methods of data mining as decision trees and propositional rule learning.

Bibliography

[Bongard, 1967] Bongard M. The problems of knowledge formation. Nauka-Moscow, 1967, 320 pp. (in Russian)

[Gladun and Rabinovich, 1980] Gladun V., Rabinovich Z. Formation of the World Model in Artificial Intelligence Systems. Machine Intelligence, 9, Ellis Herwood Ltd., Chichester, 1980, pp.299-309.

[Gladun and Vashchenko, 1995] Gladun V., Vaschenko N. Local Statistical Methods of Knowledge Formation. Cybernetics and System Analysis, v.31, N2, 1995, pp.207-217 (in Russian).

[Gladun and Vashchenko, 2000] Gladun V.P., Vaschenko N.D. Analytical Processes in Pyramidal Networks. Int. Journal Information Theories and Applications, Vol.7, No.3, 2000, pp.103-109.

[Gladun et al, 2008] Gladun V., Velichko V., Ivaskiv Y. Selfstructurized Systems. International Journal Information Theories and Applications. FOI ITHEA, Sofia, Vol.15,N.1, 2008, pp.5-13.

[Gladun, 1987] Gladun V.P. Planning of Solutions. Kiev, Naukova Dumka, 1987, 168 p, (in Russian).

[Gladun, 1994] Gladun V.P. Processes of New Knowledge Formation. Sofia, SD Pedagog 6, 1994, 192 p, (in Russian).

[Gladun, 2000] Gladun V.P. Partnership with Computers.. Man-Computer Task-oriented Systems. Kiev, Port-Royal, 2000, 120 p, (in Russian).

[Gladun, 2003] Gladun V.P. Intelligent Systems Memory Structuring. Int. Journal Information Theories and Applications, Vol.10, No.1, 2003, pp.10-14.

[Gorskii, 1985] Gorskii D. Generalization and knowledge. Mysl-Moscow, 1985, 208 pp. (in Russian)

[Michalski et al, 1986] Michalski S., Carbonell G., Mitchell M.(eds) Machine Learning, an Artificial Intelligence Approach.-Morgan Kaufmann, San Mateo, California, v.1,2, 1986.

[Piatetsky-Shapiro and Frawley, 1991] Piatetsky-Shapiro G., Frawley W. (eds) Knowledge Discovery in Databases. AAAI Press, Menlo Park, California, 1991.

[Pospelov, 1986] Pospelov D. The situational control. Theory and practice. Nauka-Moscow, 1986, 278 pp. (in Russian)

[Vagin, 1988] Vagin V. Deduction and generalization in the decision-making systems. Nauka-Moscow, 1988, 383 p. (in Russian)

[Voyshvillo, 1967] Voyshvillo E. The Concept. MGU-Moscow, 1967, 285 p. (in Russian)

About:

ITHEA® International Scientific Society

ITHEA International Scientific Society (ITHEA ISS) is aimed to support growing collaboration between scientists from all over the world.

ITHEA ISS a successor of the international scientific co-operation organized within 1986-1992 by international workgroups (IWG) researching the problems of databases and artificial intelligence. As a result of tight relation between these problems in 1990 in Budapest appeared the international scientific group of Data Base Intellectualization (IWGDBI) integrating the possibilities of databases with the creative process support tools. The leaders of the IWGDBI were Prof. Victor Gladun (Ukraine) and Prof. Rumyana Kirkova (Bulgaria).

Starting from 1992 until now the international scientific co-operation has been organized by the Association of Developers and Users of Intellectualized Systems (ADUIS), Ukraine. It has played a significant role for uniting the scientific community working in the area of the artificial intelligence.

To extend the possibilities for international scientific collaboration in all directions of informatics by wide range of concrete activities, in 2002 year, the Institute for Information Theories and Applications FOI ITHEA (IITA FOI ITHEA) has been established as an international nongovernmental organization. IITA FOI ITHEA is aimed to support international scientific research through international scientific projects, workshops, conferences, journals, book series, etc. The achieved results are remarkable. IITA FOI ITHEA became worldwide known scientific organization. One of the main activities of the IITA FOI ITHEA is building the ITHEA International Scientific Society aimed to unite researches from all over the world who are working in the area of informatics.

Till now, the ITHEA International Scientific Society was joined by more than 2600 members from 48 countries all over the world: *Armenia, Azerbaijan, Belarus, Brazil, Belgium, Bulgaria, Canada, China, Czech Republic, Denmark, Egypt, Estonia, Finland, France, Germany, Greece, Hungary, India, Iran, Ireland, Israel, Italy, Japan, Jordan, Kyrgyz Republic, Latvia, Lithuania, Malaysia, Malta, Mexico, Moldova, Netherlands, Peru, Poland, Portugal, Romania, Russia, Scotland, Senegal, Serbia, Montenegro, Sweden, Spain, Sultanate of Oman, Turkey, UK, Ukraine, and USA.*

ITHEA Publishing House (ITHEA PH) is the official publisher of the works of the ITHEA ISS. The scope of the books of the ITHEA ISS covers the area of Informatics and Computer Science. ITHEA PH welcomes scientific papers and books connected with any information theory or its application.

ITHEA ISS has two International Journals, established as independent scientific printed and electronic media, published by ITHEA Publishing House:

- International Journal "Information Theories and Applications" (IJ ITA), since 1993
- International Journal "Information Technologies and Knowledge" (IJ ITK) since 2007

Every year, ITHEA ISS organizes many scientific events called "ITHEA ISS Joint International Events of Informatics" (ITA). For the first 10 years of the XXI century, ITHEA ISS has organized 64 International Conferences and workshops and has published 44 scientific books.

All journals and books of ITHEA ISS are edited by the Institute of Information Theories and Applications FOI ITHEA, in collaboration with the leading researchers from the: *Institute of Cybernetics "V.M.Glushkov", NASU* (Ukraine)*; Institute of Mathematics and Informatics, BAS* (Bulgaria)*; Institute for Informatics and Automation Problems, NAS of the Republic of Armenia; University of Hasselt* (Belgium)*; Natural Computing Group (NCG) of the Technical University of Madrid* (Spain)*; Astrakhan State Technical University* (Russia)*; Taras Shevchenko National University of Kiev* (Ukraine)*; University of Calgary* (Canada)*; VLSI Systems Centre, Ben-Gurion University* (Israel).

The great success of ITHEA International Journals, Books, and Conferences belongs to the whole of the ITHEA International Scientific Society.

More information may be obtained from www.ithea.org.

Table of contents of IJ ITA Vol. 17, No.: 3