# THE SYSTEM OF QUALITY PREDICTION ON THE BASIS OF A FUZZY DATA AND PSYCHOGRAPHY OF THE EXPERTS

Voloshin O.F., Panchenko M.V.

*Abstract: The system of development unstable processes prediction is given. It is based on a decision-tree method. The processing technique of the expert information is offered. It is indispensable for constructing and processing by a decision-tree method. In particular data is set in the fuzzy form. The original search algorithms of optimal paths of development of the forecast process are described. This one is oriented to processing of trees of large dimension with vector estimations of arcs.*

*Keywords: Method of a decision-tree, fuzzy expert data, search of optimal paths.*

## Introduction

Methods of quantitative prediction (the time series, regression the analysis, simulation modeling etc.) give poor outcomes at prediction of "unstable" processes. In their basis there is "the prolongation past". The unstable processes are characterized “by violation of monotonicity”. It means, that there are discontinuous changes, unrepresentative for development of the process in past. The problem consists in representation of the future, which one can not be interpreted as customary prolongation past. The future can accept the in essence new forms. In a basis of such prediction (“quality prediction”, “guessing”) lays idea of immediate usage of knowledge of the person (expert). Thus first of all it is necessary to allow for “fuzzy” of the expert information [T.Terano, 1993], which one depends on its professional and psychological characteristics (competence, independence, objectivity, realism, tendency to hazard etc.). In a basis of the system, which one is described, the method of a decision-tree lays. The expert information will be utilized both for constructing the tree, and for an estimation of its arcs on the basis of method of on - pair matching. The expert information can be set both in the determined sort, and in the indistinct form. At processing the expert information for finding collective estimations the algebraic method will be utilized, in which one the metrics Hamming and measure of an incongruity of ranks of objects is applied.

In closing section the description of the system, ways of representation and input data processing is given.

As by search of optimal paths of the process development (optimal path from the root to leaves of a tree with allowance for of vector estimations) there is a task of large dimension, the original methods based on generalization of known scalar methods with usage of the schemes of a sequential analysis of variants are tendered [Voloshin, 1989].

The introduced outcomes are development of works [Voloshin, 1999], [Voloshin, 2001].

## Decision-tree constructing

The group from n (n $\geq$ 1) jointly working experts selects and sub problems and creates a decision-tree by determining importance (“weight”) each unit of a tree. The processing of the expert information at all stages will be carried out with allowance for of “weights” of the experts and degree of coordination of their judgments [Makarov, 1982].

In leaves of a tree there are tops, for which one sub problems are not determined any more. After arrangement of weights of arcs in a decision-tree the weight (probability, importance) each leave, that is path from the root into this leave is evaluated.

The decision-tree is created by a commission of experts (or which is a making decision person - a MDP). For each of tops of a tree (except for leaves) they determine the subordinate tops (for each problem there are determined sub problems). A weight of arcs can be set both in scalar and in the vector form. Last can be interpreted as a vector estimation of an arc (for example, efficiency and cost), or as an indistinct estimation of the expert assigned to a vector of values of the function of an accessory.

Such variants of the decision-tree definition are possible:

- The tree creates a MDP. Then the operation of the experts is only to place of a weight in a tree;
- The tree is created by a commission of experts with application of a method of pair matching.

Thus, if the task of "measurement" (obtaining of a scalar estimation) is considered, each expert gives three estimations: $a_1^i$ - "optimistic", $a_2^i$ - "realistic" and $a_3^i$ - "pessimistic".

The resulting estimation is thus. An estimation of each expert $a_i = \frac{a_1^i * \gamma_1 + a_2^i * \gamma_2 + a_3^i * \gamma_3}{\gamma_1 + \gamma_2 + \gamma_3}$ at first averages, and then, with allowance for of weights of the experts, we compute a resulting estimation. The coefficients $\gamma_1$, $\gamma_2$, $\gamma_3$ are defined empirically. On one technique $\gamma_1 = \gamma_3 = 1$, $\gamma_2 = 4$ (expert - "realist"), on other - $\gamma_1 = 3$, $\gamma_2 = 0$, $\gamma_3 = 2$ (for "optimist") and $\gamma_1 = 2$, $\gamma_2 = 0$, $\gamma_3 = 3$ (for "pessimist").

For definition of a psychological type of the expert (pessimist, optimist, realist) by included in the system resources, will be carried out psychological testing of the experts and the coefficients of "realness" ($\frac{1}{3} \le \lambda \le \frac{2}{3}$ for the realist, $0 \le \lambda \le \frac{1}{3}$ for the pessimist, $\frac{2}{3} \le \lambda \le 1$ for the optimist) are further allowed.

The coefficient of competence $k_2^i$ is calculated on the basis of accuracy of the previous prognoses on a technique [Makarov, 1982]. The initial coefficients $k_2^i$ are selected equal $\frac{1}{n}$.

One consists of main reasons of unauthenticity of the expert information in not the registration of the fact of possible "dependence" of the experts (from an eventual result, from administrative subordination etc.) or their truthfulness (for definite reasons, including, irrelevant with competence). Sometimes person is realized or is not realized speaks a lie. Also, on the basis of the psychological tests the coefficients of "truthfulness" $k_3^i$, "independence" $k_4^i$ and "caution" $k_5^i$ of the experts are defined. In outcome are evaluated of a weight of the experts $\alpha_i \frac{\sum \mu_i * \kappa_i}{\sum \mu_i}$, where $\mu_i$, in turn, weight of the factors $\kappa_j^i$, $i = \overline{1,n}$, $j = \overline{1,s}$. For different problems their value can vary. The indistinct definition of parameters $\kappa_j^i$ is admitted.

## Pair matching method for weak ranging

*The definition of advantage in the determined form*. Experimentally is shown, that the complication for the expert represents constructing ranging on the basis of the simultaneous registration of several different properties, on which one the objects $o_i, i = \overline{1,n}$ are evaluated. In such cases the experts decide the tasks of pair matching.

Each expert makes $C_n^2$ of matching, comparing each object with each. The outcome of matching j-th of the expert is represented by a matrix $A^j$ by a size $n \times n$. A unit $a_{ik} = 1$ in only case when in judgment of the j-th expert the i-th object is more preferential then k - th. By an necessary and sufficient condition that it is possible to set advantages thus, there is the ratio acyclicity of the expert advantage [Makarov, 1982].

After the definition of matrixes $A^j$ the matrix $A = (a_{qt}) = \sum_{j=1}^{N} A^j$ is evaluated then one finds $a_s = \sum_{t=1}^{n} a_{is}$ (here $s = \overline{1,n}$). It also is estimations every variants. Such processing technique of the information is further offered:

The definite level L and is set, if $a_s < L$ ($s = \overline{1,n}$), the variants are discarded, if the estimations do not exceed a given level. For variants, which one have remained, the probabilities, proportional their estimations $a_s$ are evaluated;

Any of variants is not discarded and to each the appropriate probabilities are assigned.

*The definition of advantage in a fuzzy form.* The units of a matrix $A^j$ are vectors by dimension m (values of the accessory function). Each unit of such vector is real number from 0 up to 1. Here is $a_{tq}^j = | a_{qt}^j - 1 |$,

$a_{ii}^{j}=(\frac{1}{2},...,\frac{1}{2});$ (here 1 is a unit vector of dimension m). Accordingly, in a matrix A there is $a_{qt}=\frac{\sum_{j=1}^{N}a_{qt}^{j}}{N}$,

$a_s=\sum_{i=1}^{n}a_{is}$ $o_i > o_j$ in only case when $a_i > a_j$.

## Algebraic processing techniques of the expert information

The essence of an algebraic processing technique of the expert information consists in introduction of some distance between estimations and according to this comparison to the system of some ranging.

Let $\Omega$ - set of all weak ranging of objects. Then the resulting estimation is on one of the formulas:

$A_1 \in Arg\min_{A\in\Omega}\sum_{i=1}^{N}d(A,A^i)$ (Cameni-Snell median);

$A_2 \in Arg\min\sum d^2$ (average meaning);

$A_3 = Arg\min_{A\subseteq\Omega,i=1,N}d(A,A^i)$ (Compromis);

d is distance between ranging.

As distance between ranging the metrics of the Humming is used [Voloshin, 2001]:

$$d(A,B)=\frac{1}{2}*\sum_{i,j=1}|a_{ij}-b_{ij}| \quad (1)$$

Or there is used measure of an incongruity of object ranks

$$d(A,B)=\sum_{i}|a_i-b_i|, \quad (2)$$

Here $a_i,b_i$ - rank of i-th object in ranging, which one is set by matrixes A and B.

In case of the definition of advantage in a fuzzy form in the formulas (1), (2) units are set through values of the accessory function, and the complication is encompass bought a large information content, which one is necessary for analyzing.

## Search of optimal paths

There is a necessity for finding optimal paths in a decision-tree. It combines definite tops, with highest weight (that is most probable of variants of a development of events). For this purpose the method of a sequential analysis of variants will be used [ Voloshin, 1989]. It grounded on a method Dejkstra [Edward Minieka, 1981]. It will be utilized for search of the shortest path from top s in top t:

Step 1. Before a start of algorithm execution all tops and arcs are not colored. at algorithm fulfillment of the number $d(x)$ is assigned To each top, which one is equal to length of the most short path from s in x, which one includes only colored tops ( $d(s)=0$, $d(x)=\infty$ ).

Step 2. For each uncolored top x we enumerate the value $d(x)$:

$$d(x)=\min\{d(x),d(y)+a(y,x)\}$$

If there is $d(x)=\infty$ for all uncolored tops x: to complete execution of algorithm. It means that in the initial graph there are paths from top s in uncolored tops. Otherwise it is necessary to color that of top x that has the least value $d(x)$. Let's assume y=x.

Step 3. If y=t to complete the procedure. The most short path from s in t is retrieved (it is alone path from s in t, which one consists of colored arcs). Otherwise: to go to step 2.

In case of the arcs graph definition is fuzzy (It means the vectors of numbers $d(x)=(d_k(x)),k=\overline{1,n}$ are used) the following methods of the most probable paths finding are applied.

*Method of convolutions.* It is offered to substitute vector estimations by numerical one. For this purpose apply known convolutions. For example, it can be average value of a vector units or value computed by “Hodge - Leman method”:

$$a_i = \beta * \min a_{ij} + (1-\beta) * \sum_i a_{ij} * p_j$$

$a_{ij}$ - are units of a vector, $p_j$ are their weights, $\beta \in (0;1)$ - is a “collective caution” coefficient

$$\beta = \frac{1}{n}\sum_{i=1}^{n} k_s^i$$

After that the circumscribed above Dejkstra method is applied. Let's mark, that if the convolutions are additive, one of Pareto-optimal paths will be retrieved [Makarov, 1982].

*Modified Dejkstra algorithm.* The following generalization of Dejkstra algorithm is offered. It is necessary to find the shortest path from top s in top t.

Step 1. $d_i(s) = (0,...,0)$ and $d_i(x) = (\infty,...,\infty)$ for all $x \neq s$, i=0.

Step 2. For each uncolored top x as follows we enumerate the value $d_i(x)$:

$$d_i(x) = \min\{d_i(x), d_i(y) + a(y,x)\}$$

If the vectors $d_i(x)$ and $d_i(x) + a(y,x)$ are incomparable: to store them both.

So, for tops $x_i$ there are such characteristics:

$$d(x_i) = (d_1(x_i),..., d_k(x_i))$$

$d_i(x_i)$ - is a length of one of possible paths in top $x_i$.

After that we select dominating tops $x_i, i \in A$. A is some set, for which one is fulfilled such condition $\neg\exists x_j, j \notin A \neg\exists k : d_k(x_j) \geq d_p(x_i) \forall p$, and color them.

Step 3. If y=t, we complete the procedure. The shortest path from s in t is retrieved. Otherwise: go to step 2.

The best case, when all paths in a tree will be comparable, the algorithm works with speed of Dejkstra algorithm.

*Matrix method.* Will be utilized in that case, when the tree is divided into levels, and each top of a tree refers only to tops of a lower layer. Each level of a tree is set by a matrix $A^k = \{a_{ij}\}^k = \{a_{ij}^1,..., a_{ij}^m\}^k$, here $a_{ij}$ - is probability of transition from i-th top of a k - th level into j-th top of a к+1 level is possible. And, if for к-й of a matrix was i of columns, for к+1 of a matrix will be i of rows. To each matrix is added one row and one column. If the maiden matrix has only one row, in the last column the unit vector is written. If it has some rows, in the last column then the vector $v_i = (\frac{1}{c},....,\frac{1}{c})$ is written, where with - amount of tops at the maiden level (that is all tops of the maiden level are equivalent). A unit of the last row is $a_{l+1,i} = (\sum_{j=1}^{n} a_{ij}^1 * a_{jm+1}^1)$.

Thus, in last column of the last matrix we obtain of tops weights of the lowest level in a tree. Utilizing the obtained information of tops weights in a tree, it is possible to define optimal paths in a tree. In case of large information content it is expedient to apply approximate methods of peephole optimization. For example, known method of a drop down vector [Sergienko, 1985]:

It is necessary to find the shortest path from top s to top t.

Step 1. y=s. We define a neighborhood r (it can be both path length, and amount of tops on the way). In the given neighborhood we discover top x in a tree. It is a shortest route from top y (and which one, accordingly, smaller than r, or consists less than r of tops). We color path from y in x.

Step 2. Let's assume y=x. We repeat the procedure.

Step k. y=t. The path from s in t is retrieved.

In case of a large information content the splitting of a tree into the sub trees is carried out with the help of decomposition methods of a sequential analysis of variants [ Voloshin, 1989].

The burn-time of Dejkstra algorithm is $O(1.5N^2)$, N there is an amount of tops in a tree. In case of the fuzzy definition of arcs, it is necessary to produce the convolution operation, then the burn-time of algorithm will be equal $O(1.5N^2 + K)$, here K is an amount of arcs.

## The description of the system.

The decision-tree is set by a matrix of incidences. In each cell of a matrix there is a vector $a_{ij}$, which one sets a transition probability from top $i$ in top $j$. It consists of ten natural numbers $(a_1,...,a_{10}), 0 \leq a_i \leq 1$. The sum of units of each row is equal to a unit vector. The matrix is filled in by interrogation of the experts. There are functions: addition of rows and columns, backup of number, dictionary, saving of the table in the file, loading of the table from the file.

For expert interrogation it is necessary to take advantage of the form, which one will allow to set up to 10 matrixes of identical dimension. Each of such matrix is an outcome of matching by the expert of tops variants, which one can be included in a tree (matrix $A^j$). The analysis of these matrixes will further be carried out. In outcome the tops are determined, which one is included in a tree. Probabilities of transition in them from high level top are determined. These outcomes are recorded in a current row of a matrix of the table, which one describes a decision-tree level.

After the definition of an incidences matrix, its analysis is possible. For this purpose it is necessary to set two tops in a decision-tree and the shortest (least probability) and most lengthy (most probable) path connecting these tops will be retrieved. As well the count of total load of each top in a tree is possible.

If the decision-tree is divided on some sub trees, which one have identical leaves, probabilities of these leaves in each of the sub trees are evaluated at first, and then there are average probabilities for entire tree as a whole.

## Conclusion

The application of the given system is possible in such areas as medical diagnostic, the prediction of currency course etc. And system accuracy depends only on proficiency of the experts.

## Bibliography

[T.Terano, 1993] T.Terano, K.Asai, M.Sugeno. “Applied fuzzy systems” (in Russian) M.:Mir 1993 – P. 230-241.

[Волошин, 2001] Voloshin O.F., Panchenko M. V. Unstable processes prediction by decision-tree method is based on pair matching method for analyzing expert data. Proceedings of KDS-2001 conference. Saint Petersburg. 2001.-p. 50-53.

[Voloshin O.F., 1989] Voloshin O.F. “Method of localization of area of an optimum in tasks of mathematical programming” (in Ukrainian)// The reports of AS USSR t. 293 №3, 1989. – P.264-273.

[Voloshin O.F., 1999] Voloshin O.F., Panchenko M.V., Pihotnyk E.P. “Expert system of support of forecasting of a grivna rate” (in ukranian)// Artificial intelligence, 1999, №2, -P.354-359.

[Makarov V.I., 1983] Makarov V.I. “Bases of acceptance of the decisions and theory of a choice” (in Russian) M.Nauka ,1983.–P.185-189.

[Voloshin O.F., 2001] Voloshin O.F. Gnatienko G.M. “Construction collective ranging by the measure ranks of objects” (in Ukrainian)// Bulletin of Kiev University, Series: Cybernetic, № 4, 2001. – P. 155-158.

[Sergienko I.V., 1985] Sergienko I.V. “Mathematical models and methods of the decision of tasks of discrete optimization” (in Russian) Kiev: Nauka Dumka, 1985.-384 p.

[Edvard Minieka, 1981] Edward Minieka “Optimization Algorithm for Network and Graphs” (in Russian).M.:Mir,1981. p.40-50.

## Authors information

**Oleksii F. Voloshin,** Kyiv T. Shevchenko national University, the faculty of cybernetics, Kiev, Ukraine. Professor. E-mail: ovoloshin@unisyb.kiev.ua; smost@i.com.ua

**Maksim V. Panchenko** - Kyiv T. Shevchenko national University, the faculty of cybernetics, Kiev, Ukraine. Post-graduate. E-mail: panchenko@ukr.net