UDK 004:93

V. Peredery, E. Borchik, O. Ohnieva

INFORMATION TECHNOLOGY OF UPDATING OF DECISION SUPPORT ALTERNATIVES IN THE DISTRIBUTED DATABASE OF CRITICAL APPLICATION SYSTEMS

Annotation. *Information technology of generating and evaluating of relevant alternatives for making decisions for the distributed database upgrading in ergatic systems of critical application is proposed. Using the method of fuzzy classification for generating and evaluating of relevant alternatives for making decisions in conditions of inconclusive, inaccurate and contradictory information in real time is considered.*

Key words: *decision – making relevance, ergatic systems of critical application, real time mode, informational needs.*

Actuality of the research

This paper aims at improving information technology of decision support in complex ergatic systems of critical application.

Using the given information technology will allow to generate and evaluate relevant alternatives for decision – making made by a human operator in conditions of inconclusive, inaccurate and contradictory information in real time with the following upgrading of the distributed database. Due to this the paper is considered to be of great importance.

Analysis of publications on the problem

Supporting the relevance of the knowledge base (KB) operation is a difficult and time-consuming process. Proposed ways of the KB updating and its support [1-3] suggest professional technical knowledge, moreover, they presuppose the existence of large arrays of address databases received from various sources. In papers [4-6] methods of examination and diagnosis in order to maintain the KB relevance are considered. There are some disadvantages of the KB upgrading in using expert methods [3] connected with the customer's inability to define needs exactly to the system being developed. Also, physiological problems may appear. Developing the KB system, the expert can prevent the transfer of his knowledge, fearing of being replaced with the machine after a little while [7]. Considered ways of the KB upgrading, in our opinion, are the complex procedure and they take much time.

This paper presents information technology of the KB upgrading by means of preformatting, with using fuzzy classification, relevant alternatives for decision making, with the further their expert evaluation, which in turn will allow to boost operativeness of the KB upgrading process in critical application systems in managing of complex technological facilities in real time.

The purpose of the paper

Improving information technology allowing in real time to generate and evaluate relevant alternatives for decision support made by a human – operator with the further upgrading of the distributed database (DB).

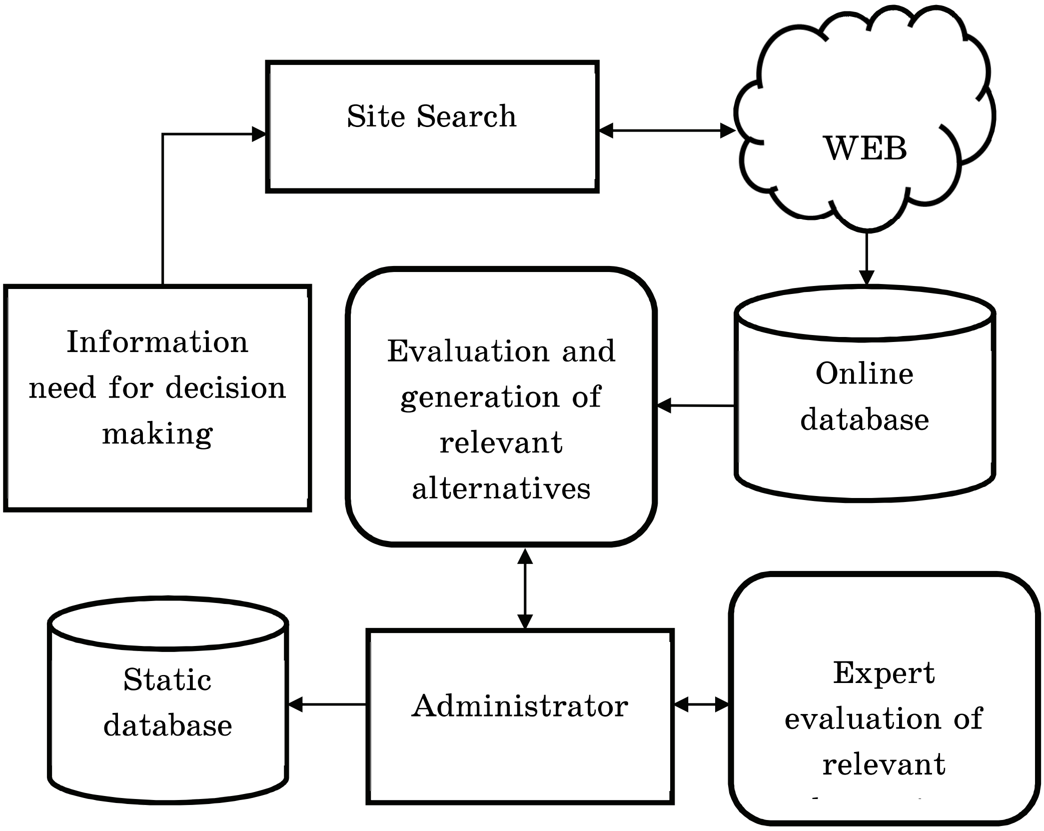

Figure 1 – Information technology of generating, evaluating and upgrading of decision making alternatives in the distributed DB

First of all, it is necessary to divide lots of documents D into two groups (classes): C_1 - relevant and C_2 - irrelevant documents. The following factors affect the document relevance [8]:

F_1 - number of repetitions of keyword query in full scale in the document;

F_2 - number of repetitions of search request in document subtitles;

F_3 - number of repetitions of synonyms of key words.

To solve this problem, we use the method of fuzzy classification [9]. Introduce fuzzy variables V_1, V_2, V_3, which we call as well as informative signs F_1, F_2, F_3. Let all the variables have two single therms: "low number", "high number". Then we will label therms V_1 by ″sign1″, ″nosign1″; therms V_2 by ″sign2″, ″nosign2″; therms V_3 by ″sign3″, ″nosign3″. Quantitative informative signs F_1, F_2, F_3 are labeled by x_1, x_2, x_3. Where $x_1, x_2, x_3 \in N$, that is a set of natural numbers.

Membership functions $\mu_{11}(x_1)$ and $\mu_{12}(x_1)$ the therms ″nosign1″ and ″sign1″ respectively from the input variable x_1 are shown in fig.2.

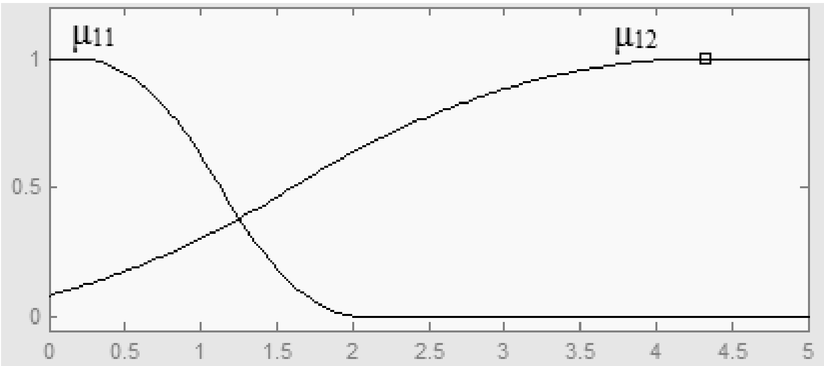

Figure 2. Membership functions

They represent standard polynomial functions of Fuzzy Logic Toolbox package, each of which is characterized by two parameters, that is $\mu_{11}(x_1) = zmf(x_1, [a_1\ b_1])$, $\mu_{12}(x_1) = smf(x_1, [e_1\ d_1])$. Similarly, functions $\mu_{21}(x_2) = zmf(x_2, [a_2\ b_2])$, $\mu_{22}(x_2) = smf(x_2, [e_2\ d_2])$, $\mu_{31}(x_3) = zmf(x_3, [a_3\ b_3])$, $\mu_{32}(x_3) = smf(x_3, [e_3\ d_3])$ $\mu_{32}(x_3) = smf(x_3, [e_3\ d_3])$ are membership functions of therms ″nosign2″, ″sign2″, ″nosign3″, ″sign3″ correspondently. Here $a_1, b_1, e_1, d_1, a_2, b_2, e_2, d_2, a_3, b_3, e_3, d_3$ - are real parameters. Because of informative signs $x_1, x_2, x_3 \geq 2$ the probability of having the document relevant increases significantly [8], let us take into account membership functions in the scope $0 \le x_i \le b \quad (i = \overline{1,3})$, $b \ge 5$. The value b one can define in the optimization process during the training sample.

Classification problem consists of assigning an object, specified by the vector of informative signs $V = (x_1, x_2, x_3)$, to one of advance certain classes $\{C_1, C_2\}$, that is, consists of performing of mapping the form:: $V = (x_1, x_2, x_3) \to u \in \{C_1, C_2\}$.

Classification based on fuzzy inference is made on the knowledge base in the form:

IF	$x_1 = \text{sign1}$	and	$x_2 = \text{sign2}$	and	$x_3 = \text{sign3}$
OR	$x_1 = \text{sign1}$	and	$x_2 = \text{sign2}$	and	$x_3 = no\text{sign3}$
OR	$x_1 = \text{sign1}$	and	$x_2 = no\text{sign2}$	and	$x_3 = \text{sign3}$
OR	$x_1 = no\text{sign1}$	and	$x_2 = \text{sign2}$	and	$x_3 = \text{sign3}$
OR	$x_1 = no\text{sign1}$	and	$x_2 = no\text{sign2}$	and	$x_3 = \text{sign3}$
OR	$x_1 = no\text{sign1}$	and	$x_2 = \text{sign2}$	and	$x_3 = no\text{sign3}$

THEN u=C_1 (1)

IF	$x_1 = \text{sign1}$	and	$x_2 = no\text{sign2}$	and	$x_3 = no\text{sign3}$
OR	$x_1 = no\text{sign1}$	and	$x_2 = no\text{sign2}$	and	$x_3 = no\text{sign3}$

THEN u=C_2.

Then membership degrees of an object , informative signs are given by a vector $V^* = (x_1^*, x_2^*, x_3^*)$, to classes C_1, C_2 from knowledge base (1), are calculated as:

$$\mu_{C_1}(V^*) = (\mu_{12}(x_1^*) \wedge \mu_{22}(x_2^*) \wedge \mu_{32}(x_3^*)) \vee (\mu_{12}(x_1^*) \wedge \mu_{22}(x_2^*) \wedge \mu_{31}(x_3^*)) \vee$$
$$\vee (\mu_{12}(x_1^*) \wedge \mu_{21}(x_2^*) \wedge \mu_{32}(x_3^*)) \vee (\mu_{11}(x_1^*) \wedge \mu_{22}(x_2^*) \wedge \mu_{32}(x_3^*)) \vee$$
$$\vee (\mu_{11}(x_1^*) \wedge \mu_{21}(x_2^*) \wedge \mu_{32}(x_3^*)) \vee (\mu_{11}(x_1^*) \wedge \mu_{22}(x_2^*) \wedge \mu_{31}(x_3^*)) \quad (2)$$

$$\mu_{C_2}(V^*) = (\mu_{12}(x_1^*) \wedge \mu_{21}(x_2^*) \wedge \mu_{31}(x_3^*)) \vee (\mu_{11}(x_1^*) \wedge \mu_{21}(x_2^*) \wedge \mu_{31}(x_3^*)).$$

Here $\wedge$ and $\vee$ mean operation of finding the minimum and maximum correspondently.

As a solution, there is selected class with a maximum degree of membership:

$$u^* = \underset{\{C_1, C_2\}}{\arg} \max(\mu_{C_1}(V^*), \mu_{C_2}(V^*)) \tag{3}$$

As a criterion of training fuzzy classifier [2] let us choose the simplest criterion [9] - a criterion of error rates minimization of classification during the training sample:

$$\frac{100\%}{N} \sum_{k=1}^{N} \Delta_k(P) \to \min, \tag{4}$$

where $\Delta_k(P) = \begin{cases} 1, & if \quad u_k \neq F(P, V_k) \\ 0, & if \quad u_k = F(P, V_k) \end{cases}$ - an error of classification of the *k-th* object, specified by the vector of informative signs V_k; N − the number of objects in the training sample (or - the number of pairs of "input-output) (V_k, u_k), $k = \overline{1, N}$ classifier (2)); $P(a_1, b_1, e_1, d_1, a_2, b_2, a_3, b_3, e_3, d_3)$ − is the vector of the parameters of the membership function of the fuzzy therms x_1, x_2, x_3 from knowledge base (1); $F(P, V_k)$ − is the result of the classification on the fuzzy basis (1) with parameters P if the input value is V_k.

Training fuzzy classifier, therefore, is finding the vector P that minimizes the distance between the results of the logical inference and experimental data from the sample (V_K, u_K).

As the minimized function (4) is an integer, the most appropriate tool is the genetic algorithm for finding extremum. For its realization we use gatool MatLAB function

Considered above the classification process (2), (3) on knowledge base (1) not only divides documents into two groups: relevant and irrelevant, but also it allows to array relevant documents in order of relevance. In fact, the more the variable $\mu_{C_1}(V^*)$, the more relevant the document, specified by the vector of informative signs $V^*(x_1^*, x_2^*, x_3^*)$.

After preliminary evaluation, formed relevant alternatives are sent by the administrator on the e – mail to the experts for evaluating. For

 ISSN 1562-9945

solving similar problems, Delphi approach is used the most commonly, which is the complex logical and mathematical procedures, directed to getting information from professionals, its analysis and generalization aimed at preparing and making rational decisions.

Let us suppose there are n documents, which each of *m* experts has in order of decreasing (or increasing) of privacy strength of some feature. Label by x_{ij} the rank (location number) of *j-th* object (j = 1,2, ..., n) in ranking *i-th* expert (i = 1,2, ..., m). Sum of ranks in ranking of *i-th* expert [2]:

$$\sum_{j=1}^{n} x_{ij} = 0{,}5n(n+1) \tag{5}$$

If it is difficult for experts to rank objects, they can give the same ranks to two or more objects. In this case the total number of different ranks in ranking will be less than *n*. It is necessary to put in order such ranking when the condition is realized (5). For this object with similar ranks the rank is given which equals to average value of location numbers which these objects share between them.

Information received from *m* experts is generalized by calculating of the sum of ranks x_i, given by all the experts to each of *n* object. The result of generalization is locating objects in order of increasing of the sum of ranks.

Expert's coordination in objects ranking is worth evaluating by the concordance coefficient.

$$K_0 = \frac{12s}{m^2(n^3 - n) - m\sum_{i=1}^{m} T_i}, \tag{6}$$

where

$$s = \sum_{j=1}^{n} dj^2;$$

$$T_i = \sum_{\mu=1}^{n} (t_{\mu i}^3 - t_{\mu i});$$

$$d_j = x_j - 0{,}5m(n+1)$$

Variables d_j define deviation of sums of ranks of *j-th* object from the average value of sums of ranks across all objects. Variables $t_{\mu i}$ – are numbers of μ*-th* ranks repetition in ranking of *j-th* expert (number of tied ranks in ranking of *i-th* expert). Coefficient K_0 equals to one when all the experts ranked objects identically and it equals to zero at uniform sums of ranks of all the objects.

For definition and estimation automation of main aspects of expert evaluation coordination, the administrator interface was developed (fig.3), which will allow to define the degree of memberships of relevant alternatives quickly and their significance for KB updating (fig.4).

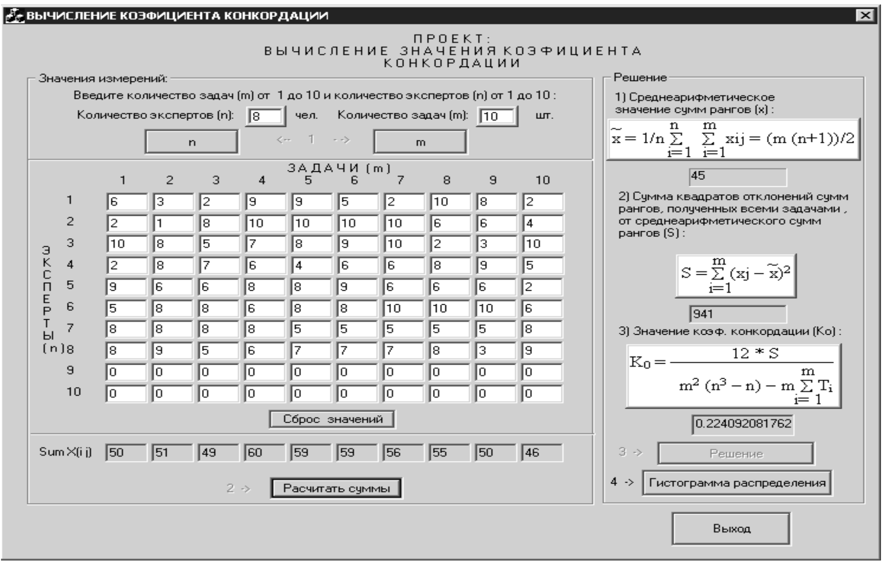

Figure 3- Expert's coordination calculation interface

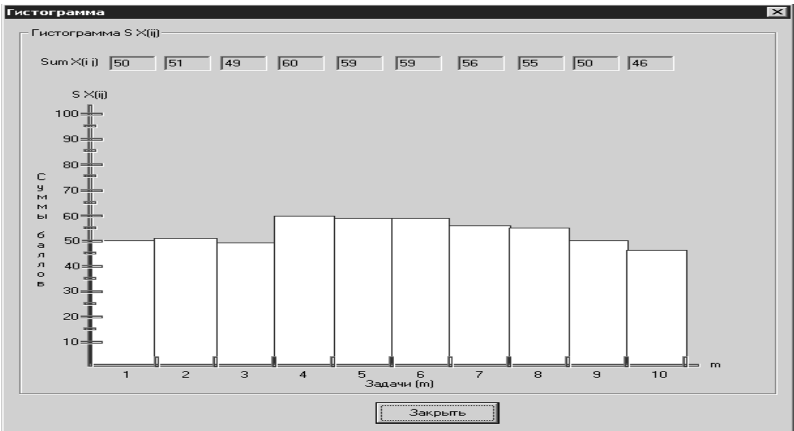

Figure 4- Bar chart of degree division of alternatives value for making decision

Conclusion

Using the given information technology will allow to generate and evaluate relevant alternatives for decision – making according to information request of a human – operator in conditions of inconclusive, inaccurate and contradictory information in real time with the following upgrading of the distributed database.

Supposed technology will allow to increase the efficiency of the KB upgrading process and can be used in critical application systems in managing of complex technological facilities in real time.

REFERENCES

1. Rotshtein A.P. Intellektualnye texnologii identifikacii: nechetkaya logika, geneticheskie algoritmy, nejronnye seti. [Intellectual identification technologies: fuzzy logic, genetic algorithms, neural networks] Publ., - Vinnitsa: Universum-Vinnitsa, 1999. – 320 p. (In Russian)
2. Kovalenko I.I. Koncepciya postroeniya sistemnyx informacionnyx texnologij podderzhki ekspertnogo analiza scenariev [The concept of building system information technologies of support expert analysis of scenarios] / I.I. Kovalenko, V.I. Perederij, A.V. Shved // Sistemni texnologiї.[System technologies.] Publ., Regional interuniversity collection of scientific works. - Issue 6 (71). - Dnepropetrovsk, 2010. – p. 74–88. (In Russian)
3. Perederij V. I. Zastosuvannya merezhi Bajesa z ocinki stupenya znachimosti vplivayuchix faktoriv na operatora v avtomatizovanix sistemax prijnyattya relevantnix rishen [Application of Bayesian network assess the significance of the influencing factors on the operator in automated systems relevant decision-making] / V.I. Perederij, V.I. Litvinenko // Komp'yuterni nauki ta informacijni texnologiї.[Computer Science and Information Technology] Publ., Bulletin of the National University "Lviv Polytechnic". - Issue № - 733 - Lviv, 2012. -- p. 120-128. (In Ukrainian)
4. Sapronov Yu.G. Ekspertiza i diagnostika obektov i sistem servisa [Examination and diagnostics of objects and systems of service] Moscow: Publ.,: Academia, 2008. 224 p. (In Russian)
5. Uzga-Rebrovs O. Nenoteiktibu parvaldisana 3 / O. Uzga-Rebrovs. – Rezekne: R. A. Izdevnieciba, 2010. – 560 lpp.
6. Asuncion A., Newman D.J. UCI Machine Learning Repository. Irvine (USA): University of California, School of Information and Computer Science, 2007.

7. Hayes – Roth F., Waterman D., Lenat D. Postroyeniye ekspertnykh system. [Building of expert systems] – M.: World, 1987. – p. 430
8. Bayesovskiye metody v zadache otsenki relevantnosti pri poiske raboty v internet [Bayes methods in task of relevance estimation in job hunting on the Internet]/ D.B.Shekhter, A.B.Chaduik, A.L.Chervinskiy-Ivashura//Problemy programuvannya. [Problems of programming.]- 2006. — Issue № 2-3. — p. 519-525. (In Ukrainian)
9. Shtovba S.D. Vnedreniye v teoriyu nechetkikh mnozhestv I nechetkuyu logiku.[Introduction into the theory of fuzzy set and fuzzy logic]// http://matlab.exponenta.ru/ fuzzylogic/book1/1_7_5_5.php.