21. Fraze-Frazenko, A. Algorithm of study neural network for image recognition [Text] / A. Fraze-Frazenko // Technology audit and production reserves. – 2012. – Vol. 4, Issue 1 (6). – P. 33–34. Available at: http://journals.uran.ua/tarp/article/view/4781/4432
22. Fertsev, A. A. Uskorenie obucheniya neyronnoy seti dlya raspoznavaniya izobrazheniy s pomoschyu tehnologii NVidia Cuda [Text] / A. A. Fertsev // Vestnik Samarskogo gosudarstvennogo tehnicheskogo universiteta. – 2012. – Vol. 1, Issue 26. – P. 183–191.
23. Unglert, K. Principal component analysis vs. self-organizing maps combined with hierarchical clustering for pattern recognition in volcano seismic spectra [Text] / K. Unglert, V. Radić, A. M. Jellinek // Journal of Volcanology and Geothermal Research. – 2016. – Vol. 320. – P. 58–74. doi: 10.1016/j.jvolgeores.2016.04.014
24. Fakhar, K. Fuzzy pattern recognition-based approach to biometric score fusion problem [Text] / K. Fakhar, M. El Aroussi, M. N. Saidi, D. Aboutajdine // Fuzzy Sets and Systems. – 2016. doi: 10.1016/j.fss.2016.05.005
25. Perova, I. G. (Adaptive treatment of these medico-biological researches by methods of computational intelligence [Text] / I. G. Perova // Eastern-European Journal of Enterprise Technologies. – 2014. – Vol. 1, Issue 4 (67). – P. 24–28. doi: 10.15587/1729-4061.2014.21202
26. Demin, D. A., Pelikh, V. F., Ponomarenko, O. I. (1998). Complex alloying of grey cast iron [Text] / D. A. Demin, V. F. Pelikh, O. I. Ponomarenko // Litejnoe Proizvodstvo, 10, 18–19.
27. Demin, D. A. Change in cast iron's chemical composition in inoculation with a Si-V-Mn master alloy [Text] / D. A. Demin // Litejnoe Proizvodstvo. – 1998. – Vol. 6. – P. 35.

UDC 004.891.2:550.8.052
DOI: 10.15587/1729-4061.2016.75679

INTERVAL FUZZY MODELING OF COMPLEX SYSTEMS UNDER CONDITIONS OF INPUT DATA UNCERTAINTY

N. Kondratenko
PhD, Associate Professor*
E-mail: kondrn@gmail.com
O. Snihur
Postgraduate student*
E-mail: olha.snihur@gmail.com
*Department of information security
Vinnytsia National Technical University
Khmelnytske highway, 95, Vinnytsia, Ukraine, 21021

Запропоновано систему моделей на основі інтервальної нечіткої логічної системи класифікації, що дозволяє отримати вихід за умов відсутності частини вхідних значень. Система будується виходячи з експериментальних даних, допускає залучення одного або кількох експертів, а також інтеграцію сторонніх підмоделей на основі інших методів та технологій прийняття рішень

Ключові слова: нечітка логічна система, інтервальні нечіткі множини, недовизначеність, кластерний аналіз

Предложена система моделей на основе интервальной нечеткой логической системы классификации, позволяющая получить выход при условии частичного отсутствия входных значений. Система строится исходя из экспериментальных данных, допускает привлечение одного или более экспертов, а также интеграцию сторонних подмоделей на основе других методов и технологий принятия решений

Ключевые слова: нечеткая логическая система, интервальные нечеткие множества, недоопределенность, кластерный анализ

1. Introduction

Input data uncertainty is one of the key factors in complex natural systems modeling. These include ecological, social, economic, technical systems of various nature. Constructing a single analytic expression that would mathematically describe such a system is a highly complicated task, and it is only possible to make assumptions about the way the system operates based on an experimental data set.

In [1] a number of UN-factors are described, that have a defining impact on the experimental data set quality, including measurements imprecision, lack of conditions for direct observations of the object, incompleteness and ambiguity of the knowledge related to the subject area and the task at hand, unaccounted for (hidden) parameters impact, lack of expert knowledge about the subject area or inability to formalize them, as well as uncertainty caused by input feature space dimensionality (redundancy and noise) [1, 2].

All these factors are inherent in natural systems and processes in one way or another. As an example of modeling a system of this class, later in this paper we show how the condition of an artesian well can be evaluated at any given time from the beginning of hydrogeological exploration up to its full completion. This task is characterized by difficulties in accessing experimental data, since obtaining input data necessary for operation of any given model requires significant effort. It is therefore to be expected, that geological exploration which precedes putting an artesian well into operation lasts ranging from 6 months and up to several years [3].

Three stages can be defined within the flow of hydrogeological exploration (HGE): preliminary exploration, detailed exploration, and operational exploration [3, 4]. One of the characteristic features of the artesian well as an object of modeling is uncertainty, more specifically gaps in the input data vector characterizing it at any given moment in time until the full completion of all three research stages. This determines the necessity of developing mathematical models and technologies that would be able to provide results on early stages of work with the system, when the researcher does not possess the full data vector. Such models would allow to evaluate the prospects of further work with the system on early stages, and detect cases, when further work presents certain challenges. Based on this information additional research may be conducted, or a decision taken to terminate the operation altogether, that would allow to save physical and human resources.

2. Literature review and problem statement

Statistical models, which are successfully used for analysis of complex systems of various nature under conditions of input data certainty, appear inefficient on uncertain data. Applying statistical models on relatively small amounts of experimental data is especially dangerous, as the distribution laws obtained from them may be unstable [5]. Furthermore, statistical models and methods do not (fully) account for the expert knowledge of the area.

Artificial intelligence technologies such as classification and cluster analysis show good results on multi-dimensional data [6–9]; they are successfully applied for extracting hidden patterns and internal correlations within a data set [6, 10–12].

Techniques based on interval fuzzy sets allow to build mathematical models of complex systems and processes, which are capable of handling input data that contains gaps [13]. It is important that they do not impose any limitations in terms of model complexity, as it is only the response of the system to the given input data vector that is being modeled, without having to build a physical model of any internal processes occurring within the system, or causal relationships existing therein [6]. As of today, such models have wide application in medical diagnostics [13], pattern recognition [14, 15], for multimedia traffic modeling and classification [16], portfolio optimization [17], forecasting of time series [18, 19] etc. Artesian well prospects evaluation is another application for which the existing decision-making models can be adapted, and new ones proposed.

Fuzzy and neuro-fuzzy methods have long been applied for water quality evaluation: fuzzy inference based decision support methods and systems are presented in [20–22]; in [23] a method of fuzzy water quality evaluation for multiple monitoring points is proposed. Fuzzy clustering methods also yield positive results in this application [24, 25]. Working with groundwater is complicated due to their inaccessibility for direct examination. On the primary stages of hydrogeological exploration it is technically possible to collect only indirect knowledge. Information about system status is available only in individual points of the deposit; information about other areas is acquired by extrapolating the actual point data onto areas, for which no actual information is available [3]. That is why the modern groundwater quality evaluation methods and technologies [26–28] are in general no different from the methods used for surface water analysis. They too require direct access to aquifers, capabilities for performing pumping tests and research pumpings, unhindered test water sample obtainment, which means full completion of the HGE. Of all factors impacting groundwater quality and drinking water mining feasibility, the most attention is given to anthropogenic pollution [29, 30] and investigating aquifers' vulnerability to harmful substance present in the air, soil and surface waters [31, 32].

From this point of view models based on type-1 fuzzy sets have a significant limitation: they cannot directly process incomplete/uncertain input data [13]. The existing fuzzy inference technique does not allow to determine the output value in case the input vector is incomplete. Interval type-2 fuzzy models allow to account for and to model different types of uncertainties, including, in some cases, uncertainty originating from missing values [33]. Therefore in such conditions it is advisable to apply mathematical tools of type-2 fuzzy sets.

As a rule, in order to study behaviors of complex systems a single model is synthesized, and also a single criterion to measure discrepancies between the model output and the observations data. This approach only works when there is a functional relation between inputs and outputs of the system, and when observations are conducted with perfect precision. In case at least one of these conditions is not satisfied, it is recommended to build a system of models and conformity criteria [34, 35]. Natural systems and processes are an example of these due to low accessibility of observations data and no way to guarantee the precision of quantitative parameters measurements. Herewith, according to [34], the more complex the system and the less the certainty and accessibility of observations data, the more diverse the models selected by different criteria will be. This factor fully applies to natural systems as well. Based on all of the above, we propose to not restrict ourselves to using the capabilities of fuzzy set mathematics only, but to develop an approach based on a system of models, that allows for utilizing other decision making methods and technologies.

The given problem can be formulated as follows. Consider an experimental data set (X, Y):

$$X = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_m^1 \\ x_1^2 & x_2^2 & & x_m^2 \\ \dots & & & \\ x_1^n & x_2^n & \dots & x_m^n \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix},$$

where $X^i = \{x_1^i, \dots, x_m^i\}$ are results of examining a system W^i against parameters $p_1 \dots p_m$. For each vector X^i the value of the linguistic variable Y is known, and is the final conclusion for the system W^i assigned by an expert (a diagnosis, a quality class, operability, – depending on the application).

For any given system W^z defined by the input vector

$$X^z = \{x_1^z, \dots, x_m^z\}, \; X^z \notin X; \; x_{k<m}^z, \dots, x_m^z \in \varnothing$$

find the value of the linguistic variable Y. Considering the challenges described above, the problem requires synthesizing a system of models capable of fuzzy inference, taking expert knowledge into account, and including additional models and methods into the decision making process.

3. Research goal and objectives

The goal of the research is expanding the capabilities of existing decision making models and methods operating under conditions of input data uncertainty. We propose to build a system of models that would combine the advantages of an interval fuzzy inference based decision support system, as well as Data Mining technologies.

In order to reach the goal set for the present research, following objectives were to be achieved:

– develop a data classification model fit for operating on an uncertain input vector;

– propose ways of accounting for expert knowledge during decision making;

– propose an aggregated criterion that would give a generalized interval estimation of the output variable value based on multiple models;

– propose an alternative decision making model based on Data Mining technologies.

4. Methods and tools for modeling a natural system state under the conditions of input data uncertainty

The general look of the aggregated model of the decision making process under the conditions of uncertainty is shown in Fig. 1.

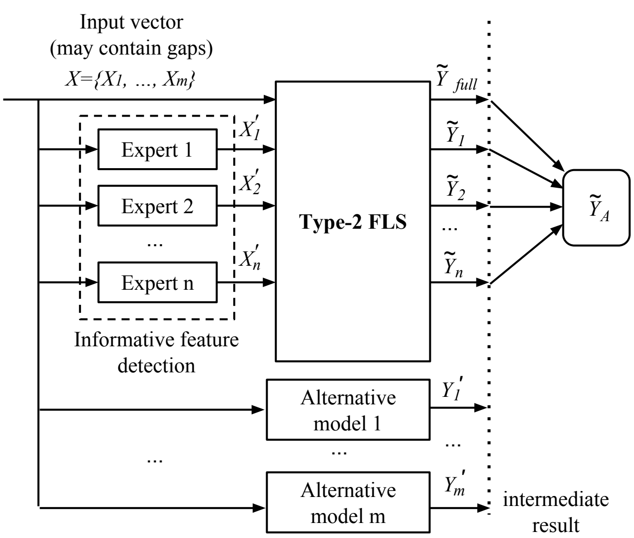

Fig. 1. Aggregated Decision Making Model

The source experimental data vector that may contain gaps is applied to the input of an interval type-2 fuzzy system unmodified, and also after going through an informative feature detection procedure. This procedure may be performed by one or more experts in the subject area. In case when more than one expert is present, every one of them generates one's own feature set, and as a result, a separate model. The interval output of the fuzzy logic system on the full vector is subsequently united with the outputs of the models resulting from uninformative features elimination. The union is performed according to the aggregating criterion Y_A.

The proposed system of models supports integration with one or more alternative models based on other decision making technologies or formal procedures that already exist in the area for solving the given problem. A decision making procedure based on a clustering method will be shown further as an example of such a model.

In case the interval output of an alternative model has the same dimensionality and qualitative nature as the fuzzy logic system output intervals, it is also taken into account when calculating an aggregated interval with the criterion Y_A. Otherwise the alternative models' outputs are presented to the user as separate intervals regardless of the main output.

4. 1. Interval type-2 fuzzy logic system for decision making

In order to solve the formulated problem, a classifier fuzzy logic system was built. An input vector is a set of system parameters' values $X^i = \{x_1^i, ..., x_m^i\}$. The knowledge base is formed by the known (X_i, Y_i) pairs, where an input vector is mapped to a linguistic estimation of the output variable Y value given by an expert. This way every input vector generates one rule. Rules antecedents are created by replacing a value x_j^i with the respective fuzzy term $A_{x_j}^i$; consequents are terms of the linguistic variable y, assigned by an expert for the vector X^i:

$$R^i : \text{IF } x_1 \in A_{x_1}^i \wedge$$
$$\wedge x_2 \in A_{x_2}^i \wedge ... \wedge x_m \in A_{x_m}^i \text{ THEN } y \in L_y^k \in \{L_1, ..., L_p\},$$

where x_i are the input variables, y is the output variable,

$$L_y \in \{L_1, ..., L_p\}$$

are term sets of the output variable. The term sets of the input and output variables are defined with Gaussian membership functions

$$\mu(x_i^j) = e^{-\left(\frac{x - b_i^j}{c_i^j}\right)^2}.$$

Given the (X, Y) data set, as well as the knowledge base synthesized from the experimental data, a type-1 fuzzy logic system with a crisp output $Y \in [0;10]$ is built. If necessary, membership functions parameters optimization is performed in order to improve the adequacy of reflecting the learning data by the model. After that the resulting type-1 membership functions are transformed into type-2 membership functions with uncertain means (Fig. 2).

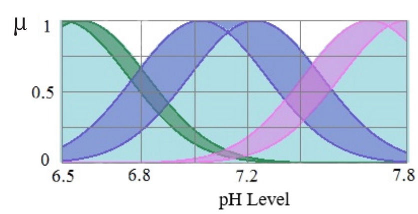

Fig. 2. An Example of an Interval Membership Function with an Uncertain Mean

Membership functions intervals bounds expansion is performed until the following condition is satisfied:

$$\forall x_i \in X \; F(x_i, P^{(k)}) = F(x_i, P^{(1)}),$$

where $P^{(1)}$ are initial membership functions parameters,

$$P^{(k)} = \{\mu_1 \cdot k, ..., \mu_p \cdot k\}, \quad k = k \pm 0{,}001,$$

$F(x_i, P^{(k)})$ is the systems output without defuzzification – the ID of the term with the highest coverage by the resulting membership function.

Fuzzy inference is performed according to the algorithm by Karnik and Mendel [36]. Interval membership grades of every rule are calculated as minimums of all antecedents:

$$\mu_{R_i} = \left[\min\left(\underline{\mu}_j^{(2)A_j^{(i)}}(x_j^*)\right); \min\left(\overline{\mu}_j^{(2)A_j^{(i)}}(x_j^*)\right) \right].$$

In order to find the left and right bounds of the output variable interval $[y_l; y_r]$, an output type-2 fuzzy set is built based on the calculated rules membership grades and rules' consequents interval values. The output value interval is obtained from the fuzzy set type reduction procedure. For the right interval bound:

1. Calculate

$$f_r^i = \left(\underline{\mu}_i + \overline{\mu}_i\right)/2; \quad y_r = \frac{\sum_{i=1}^{M} f_r^i w_r^i}{\sum_{i=1}^{M} f_r^i}; \; y_r' = y_r.$$

2. Find R (R=1...M–1), such that $w_r^R \le y_r' \le w_r^{R+1}$.

3. $$y_r = \frac{\sum_{i=1}^{R} \underline{f}^i w_r^i + \sum_{i=R+1}^{M} \overline{f}^i w_r^i}{\sum_{i=1}^{R} \underline{f}^i + \sum_{i=R+1}^{M} \overline{f}^i}; \; y_r'' = y_r.$$

4. If $y_r'' \ne y_r$, go to step 5, otherwise $y_r = y_r''$ and go to step 6.

5. $y_r' = y_r''$, back to step 2.

For the left bound of the interval:

1. Calculate

$$f_l^i = \left(\underline{\mu}_i + \overline{\mu}_i\right)/2; \quad y_l = \frac{\sum_{i=1}^{M} f_l^i w_l^i}{\sum_{i=1}^{M} f_l^i}; \; y_l' = y_l.$$

2. Find L (L=1...M – 1), such that $w_l^L \le y_l' \le w_l^{L+1}$.

3. $$y_l = \frac{\sum_{i=1}^{L} \overline{f}^i w_l^i + \sum_{i=L+1}^{M} \underline{f}^i w_l^i}{\sum_{i=1}^{L} \overline{f}^i + \sum_{i=L+1}^{M} \underline{f}^i}; \; y_l'' = y_l.$$

4. If $y_l'' \ne y_l$, go to step 5, otherwise $y_l = y_l''$ and go to step 6.

5. $y_l' = y_l''$, back to step 2.

The width of the resulting interval $[y_l; y_r]$ characterizes the degree of uncertainty associated with the decision taken.

4. 2. Models with input feature set dimensionality reduction

Experimental research shows, that the interval fuzzy classifier shown above does not always yield the expected result. A subset of modeling problems that deal with systems of unformalizable nature is comprised by problems with a significant number of input parameters. Such problems contain uncertainty related to the input feature space dimensionality. Some of its features might be redundant, others are not informative enough and act as sources of noise and anomalies in the experimental data set. In cases when it is almost impossible for a system to operate on the entire input feature set, we propose to reduce the dimensionality of the problem by eliminating the part of features that do not cause any apparent impact on the system's outcome. In the generic system of models presented in Fig.1 this function is performed by experts 1, ..., n, every one of whom offers one's own combination of informative features. Input variables eliminated by the experts are excluded from the input vector, and also from those rules of the knowledge base, in the antecedents of which they are present.

Involving several or even one expert does not always appear possible. A method described in [37] may be used to perform the role of an expert. It allows to account for both theoretical knowledge of an expert, and the quantitative data accumulated from real objects observations.

4. 3. A fuzzy clustering method with interval membership grades for decision making

Apart from the interval fuzzy set based decision making technique, which is the main part of the system of models discussed in this paper, an alternative model based on the modified PCM clustering method with interval outputs [11] is proposed. In the cluster analysis terminology the problem defined earlier may be reformulated as follows.

Consider an experimental data set X:

$$X = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_m^1 \\ x_1^2 & x_2^2 & & x_m^2 \\ \dots & & & \\ x_1^n & x_2^n & \dots & x_m^n \end{bmatrix},$$

where $X^i = \{x_1^i, ..., x_m^i\}$ are results of examining a system W^i against parameters $p_1...p_m$. In a general case the conclusion as to whether the system W^i belongs to one or more classes according to the evaluated parameter (quality, diagnosis etc.) for each of the vectors X^i is unknown, but it is known that the objects x^1, ..., x^n are distributed to form compact clusters in the input feature space $p_1...p_m$.

The requirement is to break the set X down into c clusters and calculate the membership grades for every of the c clusters for a given system W^z described by an input vector $X^z = \{x_1^z, ..., x_m^z\}$, $X^z \notin X$. The set X must satisfy the condition of being representative of the feature vectors universal set, i. e. the set X must contain representatives of all c classes.

The decision making process involves dividing the set comprised from the experimental data set X and the evaluated system's parameter vector X^z into cluster according to the method discussed in [11]. The final decision is taken based on the membership grades of the point defined by the vector X^z in every one of the c resulting clusters.

4. 4. An aggregated criterion

The end-to-end operation flow of the described system of models can be presented as the following sequence of actions.

1. Select all parameters, the values of which are known at the given moment.

2. Eliminate all other (unknown) parameters from rules antecedents.

3. Apply the vector comprised by known values to the input of the type-2 fuzzy system: $X \to Y_{full}$.

4. Dismiss (in any available way) the features that are uninformative and not informative enough and apply the resulting vector to the input of the type-2 fuzzy logic system, which results in the model $X_i' \to \tilde{Y}_i$.

5. Repeat step 4 for all available informative feature detection methods.

6. (optional) Obtain intermediate results according to alternative models 1..m.

7. Obtain the aggregated result of the main model according to the rule

$$\tilde{Y}_A = \tilde{Y}_{full} \cap \left(\bigcup_{i=1}^{n} \tilde{Y}_i \right).$$

5. Modeling results. Artesian well condition evaluation

5. 1. Interval fuzzy clustering for artesian well condition evaluation

The interval clustering method was applied to the technological problem of expert evaluation of an artesian well condition. An individual clustering object is a set of parameters values $X^i = \{x_1^i, ..., x_m^i\}$ of an artesian well; the parameters include those describing the distinctive features of geological composition, tectonic, climatic and hydrogeological conditions, data acquired by examining other wells operating in the area adjacent to the deposit, as well as results of the research conducted directly inside the well: geophysical research data, pumping test and research pumping results, regular hydrogeological observations, and parameters that characterize the groundwater quality. Cluster analysis is performed in the well feature set x_1, ... x_{84}. Some examples of the clustering features are shown in Table 1.

Table 1

Hydrogeological research parameters

Variable denotation	Parameter name	Values domain	Research stage No	Term Sets
x_1	Distance to human habitation, km	0–50	1	{L – low, M – medium, H – high}
x_2	Distance to interstate highways, km	0–50	1	{L – low, M – medium, H – high}
...				
x_{84}	Hydrogeological conditions by degree of knowledge availability	0–10	3	{A – category A, B – category B, C1 – category C1, C2 – category C2, P – category P}

A learning data set was compiled based on the archive research data of groundwater deposit wells located on the Right-Bank Geological Expedition territory. An input vector containing all well parameters x_1–x_{84} is mapped to a conclusion of an expert in hydrogeology as to its fitness for drinking water production for the period of the following 5 years. The learning data set consists of 20 samples, some of the samples are shown in Table 2.

Table 2

Well examination data (learning set)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x_1	1,5	3	12	22	18	15	15	26	37	35	27	4	48	50	50	30	25	22	39	31
x_2	43	12	12,5	25	4	11	32	2	5	2	10	3	3	0,8	4,5	1	40	14	1,5	3
...																				
x_{84}	5	4	3	2	5	4	4	9	9	7	9	9	8	8	10	10	10	10	10	10

The learning data set was divided into clusters according to method [11]. The number of clusters is considered predefined, c=3.

Membership grades for all well research samples from the learning set to every one of the clusters were calculated. The test data set consists of 10 samples, which do not form a subset of the learning data set. The membership grades calculation results for test data set samples to three clusters are shown in Table 3. Analyzing locations of cluster centers and the contents of clusters in terms of the "well perspective" concept allows to assign perspective values to clusters: cluster 1 – high, cluster 2 – sufficient, cluster 3 – insufficient.

In Table 3 and further on erroneous outputs of the system, i.e. those which do not match the expert's conclusion on the well, are highlighted with color.

Table 3

Test data set samples membership grades

№	Cluster 1		Cluster 2		Cluster 3		Result/ Interval width	Expert's Evaluation
	Left Bound	Right Bound	Left Bound	Right Bound	Left Bound	Right Bound		
1	0,165	0,178	0,618	0,918	0,249	0,491	Insufficient/0,3	Insufficient
...								
4	0,722	0,99	0,029	0,326	0,053	0,352	High/0,27	Insufficient
...								

5. 2. Type-1 fuzzy set based model

A type-1 fuzzy set based model is constructed as an intermediate stage for creating a type-2 FLS. An input vector consists of parameter values x_1, ... x_{84}. The knowledge base is created based on the experimental data from previous artesian wells explorations (Table 2). The data were processed by mapping every value from Table 2 a fuzzy variable term set from Table 1. A fragment of the resulting formalized knowledge base is given in Table 4.

On the full input data vector the system operation outcome matches the conclusion of an expert in hydrogeology in 20 cases out of 20 for the learning data set, and in 8 cases out of 10 on the test data set.

Let us model the type-1 system operation in case of uncertain input data. For this purpose Table 1 contains stages of the hydrogeological research on which values of respective parameters become known. The division into stages is relative and exists only for the sake of demonstration; in a generic case the requirement is for the system to produce results at any given moment of time between the beginning and the full completion of hydrogeological exploration. Values of the parameters that become available on a later stage are not taken into account on the current one. Columns corresponding to the unknown variables are also excluded from knowledge base rule antecedents. The third stage is equivalent to the complete certainty of all parameters.

Table 5 shows a partial result of the system's operation on the three stages. In general erroneous output was received in 6 cases out of 10 on the first stage, and in two cases on the second and third stages.

Table 4

A fragment of the fuzzy knowledge base

	R^1	R^2	R^3	R^4	R^5	R^6	R^7	R^8	R^9	R^{10}	R^{11}	R^{12}	R^{13}	R^{14}	R^{15}	R^{16}	R^{17}	R^{18}	R^{19}	R^{20}
x_1	H	H	H	M	M	L	M	M	L	H	M	H	H	M	M	M	L	L	L	H
x_2	H	H	M	M	L	L	H	H	M	M	M	L	M	M	M	M	M	L	L	L
...																				
x_{84}	B	C1	C1	C2	B	C1	C1	A	A	B	A	A	B	B	A	A	A	A	A	A
y	H	H	H	H	H	H	S	S	S	S	S	S	S	S	I	I	I	I	I	I

Table 5

Results of type-1 FLS operation on uncertain input data

№	System output			Expert's evaluation
	1st stage (18 features)	2nd stage (35 features)	3rd stage (84 features)	
1	4,04	3,52	4,12	insufficient
...				
4	5	5	3,36	insufficient
...				

5. 3. Type-2 fuzzy set based model and aggregated result

The experiment was repeated for the type-2 fuzzy set based model, partial results are shown in Table 6.

Table 6

Results of type-2 FLS operation on uncertain input data

№	1st stage (18 features)		2nd stage (35 features)		3rd stage (84 features)		Expert's evaluation
	System output	Interval width	System output	Interval width	System output	Interval width	
1	[1,39; 4,51]	3,12	[0, 24; 6,27]	6,03	[3,49; 8,97]	5,48	insufficient
...							
4	[1,18; 5,22]	4,04	[1,69; 3,76]	2,07	[1,86; 10]	8,14	insufficient
...							

The model operation result for one input vector of the test data set after feature set reduction is shown in Table 7.

Human expert 1 defined 39 informative features, expert 2 defined 41 features. The automated feature extraction method [11] allowed to define 32 informative features. Apart from the FLS interval output, the results also include outputs of two alternative models. Alternative model 1 is an interval cluster analysis decision making model. Alternative model 2 is a decision taken based on threshold values of the parameters regulated by the laws of Ukraine.

6. Artesian well condition evaluation model results discussion

Experimental tests of the interval clustering model particularly show a mismatch between the decision taken by the system and the expert conclusion in example 4 (Table 3). Sample 4 is close to cluster 1 by all parameters except one (radon concentration, 219 Bq/dm^3). Since the tested method does not have any capabilities for including other factors into consideration, except for Euclidean distance between points in a feature space, sample 4 was assigned to cluster 1 (High), although in truth water with such parameters is unsuitable for use. That is why a decision support system in the current application requires corrections introduced with expert knowledge, which is impossible to achieve with the capabilities of cluster analysis as an unsupervised learning technology. In other cases the result produced by the system matches the expert's decision for the respective sample; interval width may be regarded as a measure of uncertainty caused by lack of expert knowledge. It is fairly significant, as is to be expected for such a complex research object as a hydrogeological system.

As for the results received by the type-1 FLS, the high rate of erroneous outputs on early stages of the research, when input vectors contain a significant amount of gaps, confirms once more that the type-1 fuzzy inference engine is unfit for applications allowing for gaps in the input data. As to the errors on the third stage, when the input data are fully defined, the type-1 fuzzy set mechanisms do not allow to determine their source.

The type-2 fuzzy set based model enables quantitative assessment of the uncertainty associated with the obtained results. On the final stage of the research, when all parameter values are known, the uncertainty zone in many cases fills the entire domain of the output parameter, or most of it. However, on earlier stages, when the number of available input features is lower, the uncertainty interval is usually lower, which does not make sense from the information theory point of view [38]. This fact allows to conclude that high input feature space dimensionality complicates the system's work.

Table 7

Results of the aggregated model operation on uncertain input data

HGE stage No	Expert 1		Expert 2		Automated method		Full vector				Aggregated output
	Def. features/total	Model output	Def. features/total	Model output	Def. features/total	Model output	Def. features/total	Model output	Altern. model 1 output	Altern. model 2 output	
1	9/39	[1,39;1,51]	8/41	[0,43; 0,45]	9/32	[1,37; 1,42]	18/84	[1,39; 4,51]	x	x	[1,39; 1,51]
2	16/39	[0,24;0,27]	19/41	[0,01; 0,01]	16/32	[0,09; 0,03]	35/84	[0,24; 6,27]	x	x	[0,24; 0,27]
3	39/39	[0,09;3,97]	41/41	[0,43; 0,69]	32/32	[0,23;0,33]	84/84	[3,49; 8,7]	I:[0,62; 0,92]	insufficient	[3,49; 3,97]

Uncertainty associated with results from Table 6 was also present in the experiments with the type-1 fuzzy set model, but outputs represented with a single number did not make it possible to explain the mismatches between actual and expected results. Considering the results of the interval fuzzy set based mode, we conclude that the errors received during type-1 model testing are also the result of the input feature set redundancy.

Implementing a single aggregated model for decision making on a natural system's condition has following advantages compared to other existing approaches:

– generalization and reuse of computation methods utilized for receiving outputs of individual models;

– reducing the amount of time required for decision making in such areas as natural resources, rational exploration and environment protection;

– providing timely decisions in the dynamics of a natural process, and sufficient credibility of the decision taken.

The results of this research can potentially be applied in long-term natural processes study programs in order to shorten time required for decision making and to save resources necessary for it. This work in particular shows an application of the research results on intermediate stages of hydrogeological exploration for approximate evaluation of groundwater extraction perspective.

The present work was conducted in continuation of research [4, 11, 15, 37] in areas of fuzzy cluster analysis and fuzzy inference, and consists in bringing the results of all the previous work together to form a complete system ready for end-to-end application.

7. Conclusions

1. An interval type-2 fuzzy set based decision support system is proposed. Unlike systems based on type-1 fuzzy sets, the result of which is a membership grade represented as a single number, type-2 fuzzy sets allow to get an interval of the output linguistic variable's possible values as the system outcome. This interval appears as a result of uncertainties related to the way the expert knowledge is represented. These features allow the interval fuzzy system to function under conditions of incomplete input data, when type-1 fuzzy logic systems functioning is impossible.

2. Multiple models were built, which differ at the level of the input feature set defined by experts in the subject area as mandatory. While examining an input feature space every expert eliminates some of them as redundant, irrelevant, or such as introduce noise. As a result, a subset of the universal feature set is created. Individual features in this set make inputs of an interval fuzzy logic system and antecedents of fuzzy inference rules. This way every feature set defined by an expert generates a separate model with an interval output, which enables incorporating experts' experience in the decision making process, along with the information accumulated in the experimental data set. The multitude of models also allows to expand the concept of an expert and use automatic or automated informative feature extraction procedures alongside human experts.

3. A rule for constructing an aggregated output of the system is introduced, which allows to consolidate the results multiple models in a single interval. The aggregated criterion considers the result of system's operation over the full input vector, as well as the results of all models generated by reducing the input feature set. The latter procedure is normally performed by experts. As a result, the aggregated criterion is a generalized interval estimation of the system's status based on data available at the moment, and allows to get an idea of the uncertainty associated with the decision taken. Intermediate results, i.e. outputs of individual submodels, present value in terms of input feature set analysis. They may help determine the relation between the final decision and whether a particular parameter is considered or dismissed.

4. A way to integrate third-party models based on other decision taking methods and technologies is proposed. An alternative model is introduced, which is based on the modified PCM clustering method with interval membership grades. Presenting membership grades in interval form allows to consider and model uncertainties related to the lack of expert knowledge. The latter is especially important in the context of cluster analysis as an unsupervised learning technology.

References

1. Bardachev, Yu. N. Metodologicheskaya predpochtitelnost intervalnyih ekspertnyih otsenok pri prinyatii resheniy v usloviyah neopredelennosti [Text] / Yu. N. Bardachev, V. V. Krjuchkovskyj, T. V. Malomuzh // Visnyk Kharkivsjkogho nacionaljnogho universytetu. – 2010. – Vol. 890. – P. 18–28.
2. Pankratova, N. D. Sistemnyiy analiz i otsenivanie dinamiki ekologicheskih protsessov [Text] / N. D. Pankratova, V. V. Zavodnyk // Systemni doslidzhennja ta informacijni tekhnologhii. – 2004. – Vol. 2. – P. 47–59.
3. Borevskiy, B. V. Otsenka zapasov podzemnyih vod [Text] / B. V. Borevskiy, N. I. Drobnohod, L. S. Yazvin. – Kyiv: Vyscha shk. Golovnoe izd-vo, 1989. – 407 p.
4. Kondratenko, N. R. Ocinka perspektyvnosti arteziansjkoi sverdlovyny z vykorystannjam ekspertnykh znanj [Text]: Materialy statej P'jatoi Mizhnarodnoi naukovo-praktychnoi konferencii / N. R. Kondratenko, O. O. Snihur // Informacijni tekhnologhii ta komp'juterna inzhenerija", 2015. – P. 219–221.
5. Minaev, Yu. N. Metodyi i algoritmyi resheniya zadach identifikatsii i prognozirovaniya v usloviyah neopredelennosti v neyrosetevom logicheskom bazise [Text] / Yu. N. Minaev. – Moscow: Goryachaya Liniya-Telekom, 2003. – 205 p.
6. Cai, W. Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation [Text] / W. Cai, S. Chen, D. Zhang // Pattern Recognition. – 2007. – Vol. 40, Issue 3. – P. 825–838. doi: 10.1016/j.patcog.2006.07.011
7. De Andres, J. Bankruptcy forecasting: a hybrid approach using fuzzy c- means clustering and multivariate adaptive regression splines (MARS) [Text] / J. De Andres, P. Lorca, F. J. D. C. Juez, F. Sanchez-Lasheras // Expert Systems with Applications. – 2011. – Vol. 38, Issue 3. – P. 1866–1875. doi: 10.1016/j.eswa.2010.07.117
8. Wang, J. A modified FCM algorithm for MRI brain image segmentation using both local and non-local spatial constraints [Text] / J. Wang, J. Kong, Y. Lu, M. Qi, B. Zhang // Computerized Medical Imaging and Graphics. – 2008. – Vol. 32, Issue 8. – P. 685–698. doi: 10.1016/j.compmedimag.2008.08.004

9. Martyniuk, T. B. Applications of discriminant analysis methods in medical diagnostics [Text] / T. B. Martyniuk, L. M. Kupershtein, A. V. Medvid, A. V. Kozhemiako, W. Wojcik, O. Yuchshenko // Optical Fibers and Their Applications 2012. – 2013. – Vol. 8698. – P. 4. doi: 10.1117/12.2019733
10. Oliveira, J. V. Advances in Fuzzy Clustering and Its Applications [Text] / J. V. Oliveira, W. Pedrycz. – John Wiley & Sons Ltd., 2007. – 435 p.
11. Kondratenko, N. R. Intervaljna nechitka klasteryzacija na osnovi aljternatyvnykh kryteriiv jakosti [Text] / N. R. Kondratenko, O. O. Snihur // Naukovi visti NTUU «KPI». – 2012. – Vol. 2. – P. 59–66.
12. Martyniuk, T. B. Formalization of the Object Classification Algorithm [Text] / T. B. Martyniuk, A. V. Kozhemiako, L. M. Kupershtein // Cybernetics and Systems Analysis. – 2015. – Vol. 51, Issue 5. – P. 751–756. doi: 10.1007/s10559-015-9767-0
13. Kondratenko, N. R. Diaghnostyka ghipotyreozu na osnovi nechitkoi loghiky z vykorystannjam intervaljnykh funkcij nalezhnosti [Text] / N. R. Kondratenko, N. B. Zelinsjka, S. M. Kuzemko // Naukovi visti NTUU «KPI». – 2003. – Vol. 4. – P. 39–46.
14. Zeng, J. Type-2 Fuzzy sets for pattern classification: A review [Text] / J. Zeng, Z. Q. Liu // 2007 IEEE Symposium on Foundations of Computational Intelligence, 2007. – P. 193–200. doi: 10.1109/foci.2007.372168
15. Kondratenko, N. R. Nechitke modeljuvannja v zadachi ocinky jakosti zobrazhenj lazernykh pljam [Text] / N. R. Kondratenko, O. O. Snihur // Systemni doslidzhennja ta informacijni tekhnologhii. – 2014. – Vol. 3. – P. 7–16.
16. Liang, Q. MPEG MBR Video traffic modeling and classification using fuzzy technique [Text] / Q. Liang, J. M. Mendel // IEEE Transactions on fuzzy systems. – 2001. – Vol. 9, Issue 1. – P. 183–193. doi: 10.1109/91.917124
17. Zaychenko, Yu. P. Analiz investitsionnogo portfelya na osnove prognozirovaniya kursov aktsiy [Text] / Yu. P. Zaychenko, M. Esfandiyarfard, A. I. Zaika // Visnyk Nacionalnogo tekhnichnogo universytetutu Ukrainy «KPI». – 2007. – Vol. 47. – P. 168–179.
18. Liang, Q. Interval type-2 fuzzy logic systems: Theory and design [Text] / Q. Liang, J. M. Mendel // IEEE Transactions on Fuzzy Systems. – 2000. – Vol. 8, Issue 5. – P. 535–550. doi: 10.1109/91.873577
19. Mendel, J. M. Uncertainty, fuzzy logic, and signal processing [Text] / J. M. Mendel // Signal Processing. – 2000. – Vol. 80. – P. 913–933. doi: 10.1016/s0165-1684(00)00011-6
20. Ocampo-Duque, W. Assessing water quality in rivers with fuzzy inference systems: A case study [Text] / W. Ocampo-Duque, N. Ferré-Huguet, J. L. Domingo, M. Schuhmacher // Environment International. – 2006 – Vol. 32, Issue 6. – P. 733–742. doi: 10.1016/j.envint.2006.03.009
21. Gharibi, H. A novel approach in water quality assessment based on fuzzy logic [Text] / H. Gharibi, A. H. Mahvi, R. Nabizadeh, H. Arabalibeik, M. Yunesian, M. H. Sowlat // Journal of Environmental Management. – 2012. – Vol. 112. – P. 87–95. doi: 10.1016/j.jenvman.2012.07.007
22. Yan, H. Adaptive neuro fuzzy inference system for classification of water quality status [Text] / H. Yan, Z. Zou, H. Wang // Journal of Environmental Sciences. – 2010. – Vol. 22, Issue 12. – P. 1891–1896. doi: 10.1016/s1001-0742(09)60335-1
23. Zou, Z. Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment [Text] / Z. Zou, Y. Yun, J. Sun // Journal of Environmental Sciences. – 2006. – Vol. 18, Issue 5. – P. 1020–1023. doi: 10.1016/s1001-0742(06)60032-6
24. Iliev, B. A fuzzy technique for food- and water quality assessment with an electronic tongue [Text] / B. Iliev, M. Lindquist, L. Robertsson, P. Wide // Fuzzy Sets and Systems. – 2006. – Vol. 157, Issue 9. – P. 1155–1168. doi: 10.1016/j.fss.2005.12.014
25. Simeonov, V. Assessment of the surface water quality in Northern Greece [Text] / V. Simeonov, J. A. Stratis, C. Samara, G. Zachariadis, D. Voutsa, A. Anthemidis et. al. // Water Research. – 2003. – Vol. 37, Issue 17. – P. 4119–4124. doi: 10.1016/s0043-1354(03)00398-1
26. Dahiya, S. Analysis of groundwater quality using fuzzy synthetic evaluation [Text] / S. Dahiya, B. Singh, S. Gaur, V. K. Garg, H. S. Kushwaha // Journal of Hazardous Materials. – 2007. – Vol. 147, Issue 3. – P. 938–946. doi: 10.1016/j.jhazmat.2007.01.119
27. Singh, B. Use of fuzzy synthetic evaluation for assessment of groundwater quality for drinking usage: a case study of Southern Haryana, India [Text] / B. Singh, S. Dahiya, S. Jain, V. K. Garg, H. S. Kushwaha // Environmental Geology. – 2008. – Vol. 54, Issue 2. – P. 249–255. doi: 10.1007/s00254-007-0812-9
28. Dou, C. Transient Ground-Water Flow Simulation Using a Fuzzy Set Approach [Text] / C. Dou, W. Woldt, M. Dahab, I. Bogardi // Groundwater. – 2005. – Vol. 35, Issue 2. – P. 205–215. doi: 10.1111/j.1745-6584.1997.tb00076.x
29. Li, J. An integrated fuzzy-stochastic modeling approach for risk assessment of groundwater contamination [Text] / J. Li, G. H. Huang, G. Zeng, I. Maqsood, Y. Huang // Journal of Environmental Management. – 2007. – Vol. 82, Issue 2. – P. 173–188. doi: 10.1016/j.jenvman.2005.12.018
30. Nobre, R. C. M. Groundwater vulnerability and risk mapping using GIS, modeling and a fuzzy logic tool [Text] / R. C. M. Nobre, O. C. Rotunno Filho, W. J. Mansur, M. M. M. Nobre, C. A. N. Cosenza // Journal of Contaminant Hydrology. – 2007. – Vol. 97, Issue 3-4. – P. 277–292. doi: 10.1016/j.jconhyd.2007.07.008
31. Dixon, B. Applicability of neuro-fuzzy techniques in predicting ground-water vulnerability: a GIS-based sensitivity analysis [Text] / B. Dixon // Journal of Hydrology. – 2005. – Vol. 309, Issue 1-4. – P. 17–38. doi: 10.1016/j.jhydrol.2004.11.010
32. Dixon, B. Groundwater vulnerability mapping: A GIS and fuzzy rule based integrated tool [Text] / B. Dixon // Applied Geography. – 2005. – Vol. 25, Issue 4. – P. 327–347. doi: 10.1016/j.apgeog.2005.07.002
33. Kondratenko, N. R. Pidvyshhennya adekvatnosti nechitkykh modelej za rakhunok vykorystannya nechitkykh mnozhyn typu 2 [Text] / N. R. Kondratenko // Naukovi visti NTUU «KPI». – 2014. – Vol. 6. – P. 56–61.
34. Ivahnenko, A. G. Modelirovanie slozhnyih sistem po eksperimentalnyim dannyim [Text] / A. G. Ivahnenko. – Moscow: Rado i svyaz, 1987. – 120 p.

35. Kondratenko, N. R. Application of Type-2 Membership Functions in Fuzzy Logic Systems [Text] / N. R. Kondratenko // Research Bulletin of the National Technical University of Ukraine "Kyiv Politechnic Institute". – 2016. – Vol. 2. – P. 43–50. doi: 10.20535/1810-0546.2016.2.51636
36. Karnik, N. Type-2 Fuzzy Logic Systems [Text] / N. Karnik, J. Mendel, Q. Liang // IEEE Transactions on Fuzzy Systems. – 1999. – Vol. 7, Issue 6. – P. 643–658. doi: 10.1109/91.811231
37. Kondratenko, N. R. Evolyucijnyj poshuk informatyvnykh oznak iz zaluchennyam eksperta v zadachi ocinky yakosti arteziansjkoi vody [Text] / N. R. Kondratenko, O. O. Snihur // Visnyk Vinnycjkogho politekhnichnogho instytutu. – 2015. – Vol. 3. – P. 96–101.
38. Teoriya informatsii i ee prilozheniya (sbornik perevodov) [Text] / A. A. Harkevich (Ed.). – Moscow: Fizmatgiz, 1959. – 328 p.

UDC 681.325

DOI: 10.15587/1729-4061.2016.75595

SUMMATION OF BINARY CODES WITHOUT CARRY

M. Solomko
PhD, Associate Professor*
E-mail: doctrinas@ukr.net

L. Zubyk
Senior Lecturer**
E-mail: labrob@ukr.net

P. Olshansky
Senior Lecturer*
E-mail: p.v.olszanski@nuwm.edu.ua

V. Nazaruk
PhD, Senior lecturer*
E-mail: v.d.nazaruk@nuwm.edu.ua

*Department of Computer Engineering***
Department of Computer Science*
***National University of Water and Environmental Engineering
Soborna str., 11, Rivne, Ukraine, 33028

Розглянуто операцію додавання бінарних кодів без перенесення. Виявлено, що метод рекурсії забезпечує синтез системи бінарних кодів з кільцевою структурою при будь-якому початковому коді повної комбінаторної системи з повторенням, що й дозволяє використовувати обрану систему бінарних кодів для операції додавання без перенесення. Встановлена оцінка загальної складності обчислювального алгоритму суматора бінарних кодів

Ключові слова: суматор, комбінаторна система з повторенням, бінарні коди, додавання бінарних кодів, каскадна схема, клас комбінаторних систем, екземпляр класу, тезаурус, логарифмічна складність

Рассмотрена операция суммирования бинарных кодов без переноса. Выявлено, что метод рекурсии обеспечивает синтез системы бинарных кодов с кольцевой структурой при любом начальном коде полной комбинаторной системы с повторением, что и позволяет использовать выбранную систему бинарных кодов для операции суммирования без переноса. Установлена оценка общей сложности вычислительного алгоритма сумматора бинарных кодов

Ключевые слова: сумматор, комбинаторная система с повторением, бинарные коды, суммирование бинарных кодов, каскадная схема, класс комбинаторных систем, экземпляр класса, тезаурус, логарифмическая сложность

1. Introduction

Binary code is a general designation of the code, by which messages can be transmitted in sequences that have two characters (for example, "0" and "1"). In general, the number of combinations (codes) of n-digit binary code is equal to the number of locations with repetition of n elements by m

$$\hat{P}(n,m) = n^m. \tag{1}$$

For a binary code, the number of combinations equals:

$$\hat{P}(2,n) = 2^n, \tag{2}$$

where n is the digit capacity of a binary code.

The minimum possible number that can be written down by such a binary code equals 0. The maximum possible number that can be written down by such a binary code is determined by the formula

$$M = 2^n - 1. \tag{3}$$

Table 1

4-bit binary codes in lexicographical order

Numeric (literal) value	Binary code	Numeric (literal) value	Binary code
0	0000	8	1000
1	0001	9	1001
2	0010	A	1010
3	0011	B	1011
4	0100	C	1100
5	0101	D	1101
6	0110	E	1110
7	0111	F	1111

These two numbers fully determine the range of numbers that can be presented by a binary code (2). For example, for an 8-digit binary without a signed integer, the range of numbers is 0...255. For a 16-bit code, the range equals 0...65535.

The examples of binary codes are the code of Gray, Baudot code, Hamming code, ASCII, etc.