On Creation of Expert Knowledge Base in Information Systems of Decision Making Support

Lyudmila V. Borisova, Inna N. Nurutdinova, and Valery P. Dimitrov

Abstract—We have considered designation and structure of the knowledge acquisition block of the expert system for technical and technological service of complex machines. The main aspects of fuzzy expert knowledge representation designated for setting parameters of the machines technological adjustment have been studied. The block of expert knowledge acquisition and representation is one of the main parts in decision making problems under uncertainty. This paper suggests the technique of fuzzy knowledge base generation based on various criteria of consistency including those, that take into account different hierarchy of expert knowledge. This technique makes it possible to determine an optimal term-set for a linguistic variable, which is required to construct a general membership function and describe input and output parameters of the system. The method was applied to the subject domain of the combine harvesting of grain.

Keywords—Consistency of expert knowledge, expert system, Fishburne coefficients, fuzzy knowledge, membership function.

I. INTRODUCTION

Perspective direction in improving the methods of technological adjustment of machine is a development of decision support systems (expert systems) [1] – [4]. Specific peculiarities of the subject domain and the main requirements to information systems have defined the composition of its components. The expert system (ES) under consideration has standard blocks: harvesting quality data block, data input interactive block, knowledge base, knowledge acquisition block, answer block, solutions explanation block, teaching components block (Fig. 1).

For the purpose of adaptation of the ES knowledge base to real-life conditions, we have provided the possibility of loading into this system the knowledge indicated by an expert as well as the standard functions: loading, saving and editing of the knowledge base.

We have shown previously [4] that the tasks of initializing and adjusting the technical parameters of the machines can be considered as a problem of decision making in a fuzzy environment.

Fig. 1 block diagram of the hardware-software system «Electron expert»

Besides traditional structural constituents the method of technological adjustment includes a component providing informational support for an operator (decision-maker) and also automation of decision making during technological adjustment of the machine (Fig. 2).

Knowledge bases founded on fuzzy knowledge, i.e. fuzzy production systems, are widely used when developing ES in the sphere of operating complex machines. As a rule, expert data is hard to formalize in terms of traditional mathematical approaches and this stipulated to apply in this field a theory of fuzzy sets [5], [6]. At present the instrument of the theory of fuzzy sets covers very different fields of research and is used along with other approaches [6] – [9].

Implementation of the approach founded on application of fuzzy expert knowledge consists of three main stages: fuzzification, composition and defuzzification. At the stage of fuzzification it is necessary to present the statements of the problem solution in a linguistic form. This approach seems to be reasonable, since we don’t have any exact description of any state, both environmental factors and adjustable parameter
of the machine. The main position is that expert knowledge is, in essence, presented in a linguistic form. With the help of membership functions (MF’s) of all the terms of input linguistic variables (LV’s) on the basis of the pre-assigned definite values out of universes of input LV’s we determine a degree of confidence that an input LV possesses the value – the certain term.

The system of decision making deals with fuzzy knowledge and concepts and makes it possible to draw conclusions on the basis of fuzzy logic rules, and this actualizes the problem of most appropriate representation of fuzzy expert information (see Fig. 2). To create such information (at the stage of fuzzification) it is necessary to determine MF’s of all LV’s of a domain model, as well as the optimal number of LV’s terms. At that the set of LV values should be so that maximum conformity of expert information is provided, and at the same time it should be enough for revealing regularities and interconnections of factors. When representing linguistic values of qualitative characters as numerical elements of ordinal scales the information becomes rough, its valuable component that characterizes expert’s individual experience and knowledge disappears. Approximate representation of MF values of semantic space terms can result in inadequacy of fuzzy models to subjective opinions and initial data. To describe attributes of some subject domain experts can apply different sets of their linguistic values. In one case there appear difficulties connected with insufficiency of values, in other case – connected with their excessiveness. As a result of which, increase of fuzziness and mismatch of information obtained from experts is to be expected. The necessity of creating appropriate initial information actualizes the question of criteria according to which the choice of optimal sets of values of linguistic scale should be performed while evaluating this or that attribute. While describing real objects an optimality criterion of choice of LV terms must meet requirements of minimal uncertainty for experts and maximal consistency of expert information [6], [10]. From the practical standpoint this problem comes to establishing an optimal set of a linguistic scale used for estimating parameters of the domain model and an optimal number of LV terms. On top the number of terms is limited for reasons of measurement accuracy of the parameter under consideration. And the lower bound must be such one that it would be possible to recognize and describe interaction of input parameters with output ones. When solving the stated problem it is necessary to estimate conformity of fuzzy expert knowledge.

II. PROBLEM FORMULATION

Creation of expert knowledge base, that is appropriate to a certain subject domain, is a key stage in developing expert systems. Thereupon the problem of developing methods for creating expert information base is relevant. Representation of expert knowledge implies definition of final LV’s set, terms for each LV, construction a MF and estimation of expert information consistency. This article aims at systematizing the process of generating an extensible and editable expert knowledge base. An important aspect is exploitation of all available knowledge and prevent of any distortion of the information. For this purpose it is necessary to formulate the criteria of optimal model selection for different semantic groups of the considered properties of the subject domain.

III. DESCRIPTION OF THE METHODS

First, we establish the meaningful LV’s in the given subject domain. After that, we collect the expert knowledge that serves as a basis for the generation of the fuzzy model. The next stage is the construction of MF’s for the LV’s.

An important point when constructing a MF is definition of basic and extended term-sets. In general case the basic term-set of LV has the form [6]:

\[ T_i = \{ T_1^i, T_2^i, \ldots, T_m^i \}, \quad (i \in K = \{1, 2, \ldots, m\}). \]

Here: \( \langle T, X; \tilde{C} \rangle \) is a fuzzy variable corresponding the term \( T \in T; \tilde{C} \) is a carrier of the fuzzy set \( \tilde{C} \). In accordance with physical meaning LV terms are determined on the real axis R.

We will consider normal fuzzy sets for which upper bound of the MF is equal to 1. Fuzzy sets may be both unimodal and possessing tolerance field.

Solving the problems of mathematical simulation of complex systems applying the instrument of fuzzy sets requires performance of great deal of operations with different fuzzy variables. For ease of performing operations, and also for input-output and storage of data it is advisable to work with MF of standard form. One of the effective methods of approximating fuzzy sets is approximation with the help of functions of (L-R)-type [6].

To choose an optimal model as a criterion of consistency it is expedient to apply indices of general and pairwise consistency. When carrying out analysis of conformity of fuzzy expert information in the first stage additive and multiplicatic indices of general consistency are usually used, and according to their values a conclusion about the conformity of the models of expert assessment is stated. In the second stage the analysis of matrix of pairwise consistency of the models \( X_i \) and \( X_j \) of experts is carried on.
General consistency of a set of models of expert estimation of an attribute is determined by additive $k$ and multiplicative $\hat{k}$ indices [10], [11]:

$$k = \frac{1}{m} \sum_{i=1}^{n} \frac{1}{m} \oint \frac{\min \mu_{\alpha i}(x) \, dx}{\max \mu_{\alpha i}(x) \, dx} \quad \hat{k} = \prod_{i=1}^{m} \frac{1}{m} \oint \frac{\min \mu_{\beta i}(x) \, dx}{\max \mu_{\beta i}(x) \, dx}. \quad (1)$$

Where: $l = 1, 2, ..., m$ is the term number, $i = 1, 2, ..., n$ is the expert number, $\mu_{\alpha i}(x)$ is the MF which was preset by the $i$-th expert for the $l$-th term.

The matrix $K_{m}$ of pairwise consistency of the models $X_{i}$ and $X_{j}$ of experts is developed on the basis of indices of conformity $k_{ij}$ between the models of two experts, $i$-th and $j$-th in terms of $l$-th term for $m$-term model [6], [8], [11]:

$$k_{ij} = \frac{1}{m} \oint \frac{\min \mu_{\alpha i}(x), \mu_{\beta j}(x) \, dx}{\max \mu_{\alpha i}(x), \mu_{\beta j}(x) \, dx}. \quad (2)$$

On the basis of matrices of pairwise consistency of models for all the terms there is a matrix $K_{m}$ of consistency of the models $X_{i}$ and $X_{j}$ with all the terms. Its elements are determined by the formulae [9], [12]:

$$k_{ij} = \frac{1}{m} \sum_{l=1}^{m} k_{ij}^{l}. \quad (3)$$

Where: $m$ is a number of terms.

The analysis of additive and multiplicative indices and also matrices of pairwise consistency for the models with different number of terms may be a criterion for choosing an optimal number of MF terms [13].

As an alternative approach to estimation of expert data we can use a minimization method of weighted mean quadratic deviation $F_{m}$ of the parameters estimated by experts, from average values of these parameters:

$$F_{m} = \sum_{i=1}^{n} \sum_{j=1}^{m} \omega_{i} \sum_{j=1}^{4} \left( a_{i}^{j} - a_{j}^{i} \right)^{2} \to \min. \quad (4)$$

Where: $a_{i}^{u}$ and $a_{i}^{l}$ are tolerance limits of a fuzzy number $\mu_{\alpha i}(x)$; $a_{i}^{u}$ and $a_{i}^{l}$ are left and right coefficients of fuzziness accordingly, $a_{j}^{i}$ are their averaged values, $\omega_{i}$ are weight coefficients of experts.

From the necessary condition of the function extremum $F_{m}$ (4) we obtain:

$$a_{j}^{i} = \sum_{i=1}^{n} \omega_{i} a_{j}^{i}. \quad (5)$$

With the specified weight coefficients and constant number and composition of experts $F_{m}$ depends only on the number of the model terms. The optimal number of terms will be the one for which $F_{m}$ will adopt the least value.

Thus formed expert data serve for obtaining a generalized MF which is then applied in the mechanism of fuzzy logical derivation. And productivity and effectiveness of the derivation is provided, to a great extent, by maximal consistency of the expert data. There appears a question of the choice of weight coefficients that is not a trivial one. As an initial approximation for solving application problems it is commonly accepted to use equal weight coefficients for all the experts, and that is natural only with the same qualification of experts. It is this approach that is applied in the implemented software system for entering and updating expert knowledge [14]. However, experts’ assessments are based not only on their qualification, that is often different, but also on the use of indirect means of objective control of different accuracy. The necessity of implementing different weighting coefficients of experts is obvious. In the present paper we suggest to apply the numbers of Fishburne for calculating weighting coefficients [15]. Application of the rule of Fishburne will make it possible to take into account a significance level of experts’ assessments. Let us introduce experts $r_{i}$ and establish the relation: $r_{1} \geq r_{2} \geq ... \geq r_{n}$. The set of Fishburne weights for the system of strict preferences is determined by the formulae:

$$\omega_{i} = \frac{2(N - i + 1)}{N(N+1)}. \quad (6)$$

Where: $N$ is the number of experts, $i$ is the number of an expert by significance.

For the mixed system of preferences, when along with preferences the system incorporates indifference ratios, weighting coefficients of Fishburne have the form:

$$\omega_{i} = \frac{a_{i}}{b}, \quad a_{i} = \begin{cases} a_{i}, & \text{if } r_{i-1} = r_{i} \\ a_{i+1}, & \text{if } r_{i-1} > r_{i} \end{cases}, \quad i = N, ..., 2; \quad (7)$$

$$r_{N} = 1, \quad b = \sum_{i=1}^{N} a_{i}. \quad$$

Different considerations can be used for experts ranking, for example, the degree of consistency of their data with the assessments of other experts.

IV. APPLICATION OF THE METHODS

Without loss of generality, we will demonstrate the suggested method of estimating the consistency of the expert information on the subject domain “Combine harvesting of grain crops”. Due to the generality of the mathematical formulation, this choice doesn’t restrict the application domain.

The main blocks of ES (Fig. 1) intended for decision making by the console operator in harvesting conditions were
considered in [4], [16], [17]. The proposed methods aim at improving of knowledge acquisition block (Fig. 1).

Analysis of the subject domain has shown that for examining the question of choice of optimal set of the linguistic scale, used for estimating factors of external environment, adjustable parameters of the machine and quality index of operation, it is expedient to perform the analysis of expert data consistency.


The estimations of MF’s for LV’s were given by four experts. The MF’s were constructed based of the expert information.

As an example we present MF’s for the external factor “Stand of grain humidity”:

«Stand of grain humidity, %» - \( \mu_{SGH} : X \rightarrow [0; 1] \)

Tuple of LV<STAND OF GRAIN HUMIDITY, %> = {DSG, AHSG, HSG %}, [0 – 22], > SGH = {DSG, AHSG, HSG %}.

It is generally assumed that when stand of grain humidity is more than 22%, which corresponds the term of the given LV – “very humid”, then harvesting is not carried on.

Typical functions of trapezoidal and triangle types have been applied for describing the terms [18]. A trapezoidal form of fuzzy number is a quadruple

\[
\begin{align*}
\mu_c(x) &= \begin{cases} 
0, & \text{if } x \leq a \\
\frac{x-a}{c-a}, & \text{if } a < x < c \\
1, & \text{if } c \leq x \leq d \\
\frac{b-x}{b-d}, & \text{if } d < x < b \\
0, & \text{if } x \geq b
\end{cases}
\]

For representing leftmost \( \mu_c(x) \) and rightmost \( \mu_b(x) \) terms we used the functions:

\[
\begin{align*}
\mu_l(x) &= \begin{cases} 
1, & \text{if } x \leq a \\
\frac{b-x}{b-a}, & \text{if } a < x < b \\
0, & \text{if } x \geq b
\end{cases}
\]

\[
\mu_r(x) &= \begin{cases} 
0, & \text{if } x \leq a \\
\frac{x-a}{b-a}, & \text{if } a < x < b \\
1, & \text{if } x \geq b
\end{cases}
\]

The values of MF’s coefficients listed in [19] have been chosen for calculations. We will present, as an example, the graphs of MF’s for LV1, fixed by four experts for the 3-term model (Fig. 3). Thus range of variable \( x \) is from 0 to 1 (normalized value). Range of parameters \( a, b, c, d \) is from 0 to 1.
in interactive mode necessary configurations of expert systems (subsystems of knowledge input) with different sets of LV’s and different ways of constructing MF’s.

Results of calculations of additive $k$ and multiplicative $\tilde{k}$ indices for all the models are given in Table I.

The analysis of the obtained indices of general conformity of expert data shows that the most coordinated for LV1 and LV3 is the 2-term model, and for LV2 – the 3-term model.

The results of calculations of matrices of pairwise consistency for all the models make it possible to calculate weighted-average quadratic deviations $F_m$ of the parameters, estimated by experts, from the averaged values of these parameters. With this we use both equal weight coefficients and weights calculated according to the rule of Fishburne for strict (6) and mixed systems of preferences (7). Table II shows one of the matrices of pairwise consistency used for illustrating models of the LV1, as well as the experts’ ranks and the Fishburne weights.

The experts’ ranking was carried out on the basis of the criterion of the greatest pairwise consistency, for this we used sums of matrix lines elements.

The parameters of the generalized MF and the values $F_m$ from the condition (5) for all the models have been calculated. The results of calculations of the value $F_m$ are presented in Table III.

In Table IV we have compared the results of definition of the optimal number of LV terms which were obtained on basis of conformity index analysis and applying the method of minimization of weighted mean quadratic deviation $F_m$ of individual parameters, preset by experts, from the averaged values of these parameters.

### Table I. Results of calculations of indices $k$ and $\tilde{k}$

<table>
<thead>
<tr>
<th>LV</th>
<th>Model</th>
<th>$k$</th>
<th>$\tilde{k}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LV1</td>
<td>2-term</td>
<td>0.817</td>
<td>0.816</td>
</tr>
<tr>
<td></td>
<td>3-term</td>
<td>0.784</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>4-term</td>
<td>0.669</td>
<td>0.657</td>
</tr>
<tr>
<td>LV2</td>
<td>2-term</td>
<td>0.809</td>
<td>0.809</td>
</tr>
<tr>
<td></td>
<td>3-term</td>
<td>0.815</td>
<td>0.812</td>
</tr>
<tr>
<td></td>
<td>4-term</td>
<td>0.653</td>
<td>0.644</td>
</tr>
<tr>
<td></td>
<td>2-term</td>
<td>0.835</td>
<td>0.832</td>
</tr>
<tr>
<td></td>
<td>3-term</td>
<td>0.757</td>
<td>0.746</td>
</tr>
<tr>
<td></td>
<td>4-term</td>
<td>0.696</td>
<td>0.687</td>
</tr>
</tbody>
</table>

### Table II. Matrix of pairwise consistency for the 3-term model of the LV1

<table>
<thead>
<tr>
<th>Matrix $K_m$</th>
<th>m</th>
<th>Rank</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 0.905 0.877 0.859</td>
<td>2</td>
<td>3</td>
<td>0.2</td>
</tr>
<tr>
<td>0.905 1 0.951 0.859</td>
<td></td>
<td>2</td>
<td>0.3</td>
</tr>
<tr>
<td>0.877 0.951 1 0.903</td>
<td></td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>0.859 0.859 0.903 1</td>
<td></td>
<td>4</td>
<td>0.1</td>
</tr>
<tr>
<td>1 0.889 0.837 0.745</td>
<td>3</td>
<td>3</td>
<td>0.2</td>
</tr>
<tr>
<td>0.889 1 0.861 0.757</td>
<td></td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>0.837 0.861 1 0.787</td>
<td></td>
<td>2</td>
<td>0.3</td>
</tr>
<tr>
<td>0.745 0.757 0.787 1</td>
<td></td>
<td>4</td>
<td>0.1</td>
</tr>
<tr>
<td>1 0.855 0.815 0.813</td>
<td>4</td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>0.855 1 0.9 0.703</td>
<td></td>
<td>2</td>
<td>0.3</td>
</tr>
<tr>
<td>0.815 0.9 1 0.704</td>
<td></td>
<td>3</td>
<td>0.2</td>
</tr>
<tr>
<td>0.813 0.703 0.704 1</td>
<td></td>
<td>4</td>
<td>0.1</td>
</tr>
</tbody>
</table>

### Table III. Values $F_m$ for the LV’s under consideration

<table>
<thead>
<tr>
<th>LV</th>
<th>Model</th>
<th>$F_m$ (equal weight coeff.)</th>
<th>$F_m$ (Fishburne weight coeff.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LV1</td>
<td>2-term</td>
<td>0.0125</td>
<td>0.01065</td>
</tr>
<tr>
<td></td>
<td>3-term</td>
<td>0.008125</td>
<td>0.00705</td>
</tr>
<tr>
<td></td>
<td>4-term</td>
<td>0.01125</td>
<td>0.01115</td>
</tr>
<tr>
<td>LV2</td>
<td>2-term</td>
<td>0.014688</td>
<td>0.00985</td>
</tr>
<tr>
<td></td>
<td>3-term</td>
<td>0.005938</td>
<td>0.0055</td>
</tr>
<tr>
<td></td>
<td>4-term</td>
<td>0.004063</td>
<td>0.0035</td>
</tr>
<tr>
<td>LV3</td>
<td>3-term</td>
<td>0.004375</td>
<td>0.00358</td>
</tr>
<tr>
<td></td>
<td>4-term</td>
<td>0.007188</td>
<td>0.0047</td>
</tr>
<tr>
<td></td>
<td>5-term</td>
<td>0.010625</td>
<td>0.0084</td>
</tr>
</tbody>
</table>

### Table IV. Optimal models for LV’s

<table>
<thead>
<tr>
<th>LV</th>
<th>Optimal number of terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>LV1</td>
<td>By conformity indices</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>LV2</td>
<td></td>
</tr>
<tr>
<td>LV3</td>
<td></td>
</tr>
</tbody>
</table>

The data of Tables III and IV allow us to make a conclusion that for developing a generalized MF with the purpose of maximal consistency of expert data it is more preferable to apply weight coefficients of Fishburne with ranking of experts according to the degree of consistency of their data with those of others. Fig. 4 presents a chart of generalized MF for the 3-term model LV1 taking into account weight coefficients of Fishburne.

Fig. 4 generalized MF for the 3-term model of LV1
Consequently, under the conditions requiring the greatest consistency of expert assessments base the application of weight coefficients of Fishburne gives the best results. We should note that the suggested method is relevant for other variants of ranking expert data, for example, according to the level of experts’ qualification.

V. CONCLUSION

In this article, the method of forming a knowledge base of an expert system has been developed. It includes an analysis of fuzzy expert data based on the criteria of consistency. In order to achieve maximal correspondence of the formalized data to the real situation, we assigned weight coefficients to the estimates given by different experts, using Fishburne numbers for their ranking. The suggested technique was used to create the expert database in the subject domain “Combine-harvesting of grain crops”. Through application of our technique, we have determined basic term-sets and optimal MF for all important environmental factors and for adjustable parameters of a grain combine. Optimal model selection has been demonstrated on several examples of LV with 2-, 3- and 4-term MF’s estimated by four experts. The characteristics of general and pairwise consistency of experts’ models as well as the parameters of the generalized MF in case of equal weight coefficients and Fishburne weights have been calculated. The methods of an optimal model selection used in next stages (composition and defuzzification) of deriving the solutions have been illustrated. The developed technique of creating an expert knowledge base makes it possible to perform analysis of expert data suitability for application at the stages of composition and generation of logical derivation.

REFERENCES


Lyudmila V. Borisova graduated in 1990 with a degree in agricultural machinery at the Rostov-on-Don State Academy of Agricultural Engineering, Rostov-on-Don, Russia. She received her degree of doctor of Engineering from the Don State Technical University (DSTU), Rostov-on-Don, Russia in 2007. She is now Head of Department of Economics and Management in Engineering, DSTU, Rostov-on-Don, Russia. Her research interests include information systems of decision making support, quality management systems. She developed Expert Systems of Decision Making in the operation combine harvests. Dr. Borisova is Honorary Worker of Higher Professional Education of Russia, member of the Dissertation Council of DSTU, Rostov-on-Don, Russia. She has been involved in technical committees for several conferences.

Inna N. Nurutdinova graduated in 1982 with a degree in physics at the Rostov State University, specialty, Rostov-on-Don, Russia. She received her PhD in physical and mathematical sciences from the Institute of Semiconductors of Academy of Science of Ukraine, Kiev, Ukraine in 1992. She is now associate professor at the Department of Applied Mathematics, DSTU, Rostov-on-Don, Russia. Her research interests include mathematical modeling of systems and processes, application the theory of fuzzy sets, information systems of decision making.

Valery P. Dimitrov graduated in 1975 with a degree in design and production of agricultural machines at the Rostov institute of agricultural engineering, Rostov-on-Don, Russia. He received his degree of doctor of Engineering from DSTU, Rostov-on-Don, Russia in 2002. He is now head of Department of Quality Management, DSTU, Rostov-on-Don, Russia. His research interests include information systems of decision making support, operation of agriculture machines, quality management systems. He developed Expert Systems of Decision Making in the operation combine harvests. Dr. Dimitrov is Honorary Worker of Higher Professional Education of Russia, Chairman of the Dissertation Council of DSTU, Rostov-on-Don, Russia. He has been involved in technical committees for several conferences.