Fuzzy Model for the Assessment of Operators' Work in a Cadastre Information System

Dariusz Król[1], Grzegorz Stanisław Kukla[1],
Tadeusz Lasota[2], and Bogdan Trawiński[1]

[1] Institute of Applied Informatics, Wrocław University of Technology,
Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
dariusz.krol@pwr.wroc.pl, grzegorz_kukla@o2.pl, trawinski@pwr.wroc.pl
[2] Faculty of Environmental Engineering and Geodesy, Agricultural University of Wrocław, Norwida 25/27, 50-375 Wroclaw, Poland
tadeusz.lasota@wp.pl

Abstract. One of critical tasks of cadastre system maintaining is the input of changes into its database. Managers of information centres often complain they have no adequate tools for the assessment of work of cadastre system operators. In the paper a fuzzy model is proposed which goal is to provide a useful tool for management of an information center. The architecture of the fuzzy system comprises five main modules of operators' work statistics, fuzzification, inference, defuzzification and visualization. For each input criterion i.e. productivity (P), complexity (C), time (T) and quality (Q) as well as for output assessment triangle and trapezoid membership functions have been defined. The statistics module provides initial parameters of the model and values of input criteria. The model based on change records saved in cadastre database produces the assessments of operators' work for defined periods of time automatically.

1 Introduction

Cadastre systems are designed for the registration of parcels, buildings and apartments as well as their owners and users. Those systems have complex data structures and sophisticated procedures of data processing. The maintenance of real estate cadastre registers is dispersed in Poland. There are above 400 information centres located by district local self-governments as well as by the municipalities of bigger towns which exploit different cadastre systems. One of critical tasks of cadastre system maintaining is the input of changes into its database. The changes are allowed only on the basis of legal documents such as sale purchase agreements, extracts from perpetual books, results of surveyors' works and others. Managers of information centres often complain they have no adequate tools for the assessment of work of cadastre system operators. Reviews with the managers revealed that using only productivity expressed by the number of changes input within a given period is highly insufficient. Criteria that they mentioned first of all were complexity of changes calculated as a mean number of objects

B. Gabrys, R.J. Howlett, and L.C. Jain (Eds.): KES 2006, Part III, LNAI 4253, pp. 774–781, 2006.

which were changed in database falling per one change, average time of inputting one change and quality of work determined by the percentage of changes without any corrections. The assessment of operators' work is subjective and imprecise to some extent. If, for example, during one month an operator input 87 changes of mean complexity of 9.3 objects per change, with mean time of 13.4 minutes and without any corrections then what was his work: good, average or bad? Such evaluation has fuzzy nature. It is much easier for a manager to estimate each criterion separately using such linguistic values as high, medium and low. So it was the main reason for developing our fuzzy model of operators' work assessment.

Many books have been written on fuzzy logic and control [2], [8], numerous articles deal with fuzzy expert systems [1], [3], [6]. The fuzzy model presented in the paper is composed of typical elements. However having the statistics of cadastre system operators' work it is possible to apply them as initial parameters of fuzzification process. Based on managers' suggestions four following input variables have been designed: productivity, complexity, time and quality. Moreover using the statistics as the input of the model allows obtaining final assessments automatically. It has been assumed that experts will tune the initial parameters and the centre managers, who are the primary group of users of the system, will be able to modify such parameters as the average values of input criteria, which play the role of points of reference, or weights of rules.

2 The Model of the Assessment of Operators' Work

The fuzzy model proposed in the paper will constitute the basis of the fuzzy system which is intended to rationalize the management of information centres, to improve the organization of work and to determine wages of outsource workers. The architecture of the fuzzy system is shown in Fig. 1. It comprises five main modules of operators' work statistics, fuzzification, inference, defuzzification and visualization.

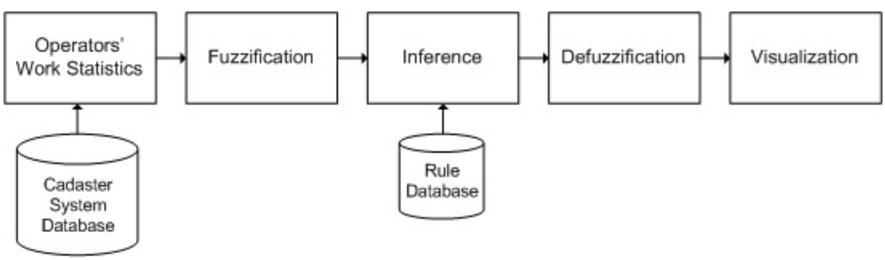

Fig. 1. Architecture of the fuzzy system for the assessment of operators' work

2.1 Input and Output of the Model

For each input criterion i.e. productivity (P), complexity (C), time (T) and quality (Q) as well as for output assessment triangle and trapezoid membership functions have been defined (see Fig. 2). The statistics module provides initial parameters of the model and values of input criteria. The idea of obtaining the final assessment consists in calculating the average value of P, C and T criteria taking into account the change records saved in cadastre database for all operators and for long period of time, e.g. a year or a half of year. These average values are used as the reference values of 100% for calculating what percentage of corresponding average value a given operator achieved within the assessment period. Values of time variable are reversed in the scale form 0% to 200% in order to achieve lower input values for longer times of introducing changes into the system. Data for quality variable are applied directly, because this criterion is expressed in percents. Standard deviations, calculated separately for each criterion for a long period of time, determine the width of the basement of triangle and trapezoid fuzzy sets by adding or subtracting them from 100%. The domain for input P, C and T are 0-200 percentage values, with 0 being the lowest and 200 the highest mark. Data for quality variable are applied directly, because this criterion is expressed in percents. The domain for output, represented by five fuzzy sets, is an arbitrary assessment scale from 1 to 200, with 0 being the lowest and 200 the highest mark.

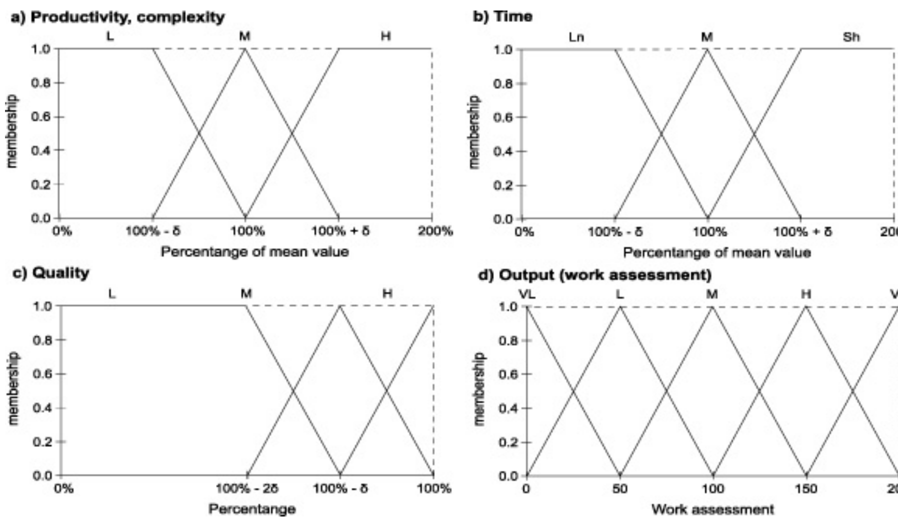

Fig. 2. Membership functions of input and output variables, where VL, L, M, H, VH denote very low, low, medium, high, very high respectively, Ln, Sh mean long and short

2.2 Formal Description of Linguistic Variables

Let x_P be a linguistic variable of productivity, and its term set be T_P= {low, medium, high}, where each term in T_P is characterized by a fuzzy set in a universe of discourse $U_P = [0, 200]$, then the membership functions of these fuzzy sets are as follows:

$$low(x_P) = \begin{cases} 1 & \text{if } x_P \leq 100 - \delta_P \\ 1 - (x_P + \delta_P - 100)/\delta_P & \text{if } 100 - \delta_P < x_P < 100 \\ 0 & \text{if } x_P \geq 100 \end{cases}$$

$$medium(x_P) = \begin{cases} 0 & \text{if } x_P \leq 100 - \delta_P \\ 1 - |(x_P - 100)|/\delta_P & \text{if } 100 - \delta_P < x_P < 100 + \delta_P \\ 0 & \text{if } x_P \geq 100 + \delta_P \end{cases}$$

$$high(x_P) = \begin{cases} 0 & \text{if } x_P \leq 100 - \delta_P \\ 1 - (100 + \delta_P - x_P)/\delta_P & \text{if } 100 - \delta_P < x_P < 100 + \delta_P \\ 1 & \text{if } x_P \geq 100 + \delta_P \end{cases}$$

where δ_P is the standard deviation of productivity calculated for a long period of time.

Let x_C be a linguistic variable of complexity, and its term set be T_C={low, medium, high}, where each term in T_C is characterized by a fuzzy set in a universe of discourse $U_C = [0, 200]$. Then the membership functions of these fuzzy sets are as follows:

$$low(x_C) = \begin{cases} 1 & \text{if } x_C \leq 100 - \delta_C \\ 1 - (x_C + \delta_C - 100)/\delta_C & \text{if } 100 - \delta_C < x_C < 100 \\ 0 & \text{if } x_C \geq 100 \end{cases}$$

$$medium(x_C) = \begin{cases} 0 & \text{if } x_C \leq 100 - \delta_C \\ 1 - |(x_C - 100)|/\delta_C & \text{if } 100 - \delta_C < x_C < 100 + \delta_C \\ 0 & \text{if } x_C \geq 100 + \delta_C \end{cases}$$

$$high(x_C) = \begin{cases} 0 & \text{if } x_C \leq 100 - \delta_C \\ 1 - (100 + \delta_C - x_C)/\delta_C & \text{if } 100 - \delta_C < x_C < 100 + \delta_C \\ 1 & \text{if } x_C \geq 100 + \delta_C \end{cases}$$

where δ_C is the standard deviation of complexity calculated for a long period of time.

Let x_T be a linguistic variable of time, and its term set be T_T={low, medium, high}, where each term in T_T is characterized by a fuzzy set in a universe of discourse $U_T = [0, 200]$. Then the membership functions of these fuzzy sets are as follows:

$$low(x_T) = \begin{cases} 1 & \text{if } x_T \leq 100 - \delta_T \\ 1 - (x_T + \delta_T - 100)/\delta_T & \text{if } 100 - \delta_T < x_T < 100 \\ 0 & \text{if } x_T \geq 100 \end{cases}$$

$$medium(x_T) = \begin{cases} 0 & \text{if } x_T \leq 100 - \delta_T \\ 1 - |(x_T - 100)|/\delta_T & \text{if } 100 - \delta_T < x_T < 100 + \delta_T \\ 0 & \text{if } x_T \geq 100 + \delta_T \end{cases}$$

$$high(x_T) = \begin{cases} 0 & \text{if } x_T \leq 100 - \delta_T \\ 1 - (100 + \delta_T - x_T)/\delta_T & \text{if } 100 - \delta_T < x_T < 100 + \delta_T \\ 1 & \text{if } x_T \geq 100 + \delta_T \end{cases}$$

where δ_T is the standard deviation of time calculated for a long period of time.

Let x_Q be a linguistic variable of quality, and its term set be T_Q={low, medium, high}, where each term in T_Q is characterized by a fuzzy set in a universe of discourse $U_Q = [0, 100]$. Then the membership functions of these fuzzy sets are as follows:

$$low(x_Q) = \begin{cases} 1 & \text{if } x_Q \leq 100 - 2\delta_Q \\ 1 - (x_Q + \delta_Q - 100)/\delta_Q & \text{if } 100 - 2\delta_Q < x_Q < 100 - \delta_Q \\ 0 & \text{if } x_Q \geq 100 - \delta_Q \end{cases}$$

$$medium(x_Q) = \begin{cases} 0 & \text{if } x_Q \leq 100 - 2\delta_Q \\ 1 - |(x_Q - 100)|/\delta_Q & \text{if } 100 - 2\delta_Q < x_Q \leq 100 \end{cases}$$

$$high(x_Q) = \begin{cases} 0 & \text{if } x_Q \leq 100 - \delta_Q \\ 1 - (100 + \delta_Q - x_Q)/\delta_Q & \text{if } 100 - \delta_Q < x_Q \leq 100 \end{cases}$$

where δ_Q is the standard deviation of quality calculated for a long period of time.

Let y be a linguistic variable of assessment, and its term set be T_A={very low, low, medium, high, very high}, where each term in T_A is characterized by a fuzzy set in a universe of discourse $U_A = [0, 200]$. Then the membership functions of these fuzzy sets are as follows:

$$very\ low(y) = \begin{cases} 1 - y/50 & \text{if } 0 \leq y < 50 \\ 0 & \text{if } y \geq 50 \end{cases}$$

$$low(y) = \begin{cases} 1 - |(y - 50)|/50 & \text{if } 0 \leq y < 100 \\ 0 & \text{if } y \geq 100 \end{cases}$$

$$medium(y) = \begin{cases} 0 & \text{if } y \leq 50 \\ 1 - |(y - 100)|/50 & \text{if } 50 < y < 150 \\ 0 & \text{if } y \geq 150 \end{cases}$$

$$high(y) = \begin{cases} 0 & \text{if } y \leq 100 \\ 1 - |(y - 150)|/50 & \text{if } 100 < y \leq 200 \end{cases}$$

$$very\ high(y) = \begin{cases} 0 & \text{if } y \leq 150 \\ 1 - |(y - 200)|/50 & \text{if } 150 < y \leq 200 \end{cases}$$

2.3 Rule Database and Inference Process

It has been assumed that the rule database should contain simple IF-THEN rules where the condition consists of only two input variables combined by AND operator and the conclusion is built by one variable. An example of a rule is as follows: **IF Productivity is medium AND Complexity is low THEN Assessment is low**. Thus the rules for one pair of input criteria can be given in the form of a matrix shown in Fig. 3. So 6 matrices for 9 rules has been designed for C-P, C-T, C-Q, P-T, P-Q, T-Q combinations. Having such form of rule database, the experts will not have difficulties to determine the output values for each pair

		Input variable C		
Input variable P		L	M	H
	L	VL	L	M
	M	L	M	H
	H	M	H	VH

– input values
– output values

VL – very low
L – low
M – medium
H – high
VH – very high

Fig. 3. Representation of rule database in matrix form for C and P input

of input values. In order to express the strength of rules belonging to particular combination, rule weights were designed with initial values of w_{C-P}=0.30, w_{C-T}=0.25, w_{C-Q}=0.20, w_{P-T}=0.10, w_{P-Q}=0.10 and w_{T-Q}=0.05 as the multipliers of rule conditions in aggregation step.

In order to assure that each rule will have influence on the final assessment following operators has been used: PROD for aggregation of rule conditions, PROD for activation of rule conclusions and ASUM for accumulation of output membership functions, where PROD means algebraic product and ASUM denotes algebraic sum [5]. In defuzzification step the center of gravity method is used.

3 Evaluation of the Fuzzy Model

Statistics module and fuzzy model have been programmed using C# and Java languages. Work statistics for the period of 6 months from January to June 2005 have been calculated for 13 operators in one of information centre (see Fig. 4).

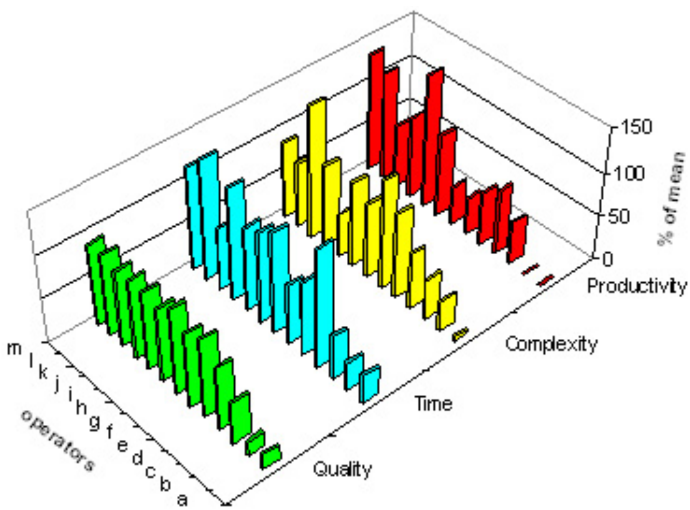

Fig. 4. Values of input criteria for 13 operators for the period of six months

The results of statistics module have been applied as model parameters and as input values. The data have been processed by fuzzy model using rule weights and final assessments for each operator and for each month have been obtained as the output.

The values of input criteria and the output produced by the model for one operator are presented in Fig. 5. The weights line denotes the result obtained using rule weights and the second one indicates assessments calculated without rule weights.

Data given in Fig. 5 provide the basis for a centre manager to take decisions.

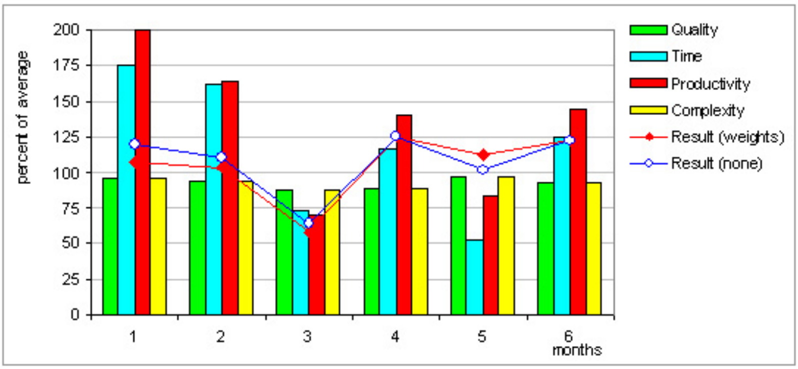

Fig. 5. Input criteria and the final assessment for one selected operator

4 Conclusions

The model for the multicriterial assessment of cadastre system has been designed and programmed with the aid of an expert and information centre managers. The evaluation of the model has proved that it is reasonable to use operators' work statistics as the initial parameters and the input values for the model. The model based on change records saved in cadastre database produces the assessments for defined periods of time automatically. The architecture of proposed model enables the centre managers to modify system parameters to customize them to local circumstances.

It is planned to carry out an evaluation experiment with the participation of the centre managers. The number of input criteria will be increased by incorporating additionally the statistics of system reports usage and the subjective managers' assessment of the usefulness of a given operator during selected period of time. Tuning tests will also be performed in order to determine how the output depends on the number of input linguistic values and rule weights and other model parameters. Finally, this idea will be used in the Web cadastre system.

References

1. Bae S.M., Park S.C., Ha S.H.: Fuzzy Web Ad Selector Based on Web Usage Mining. IEEE Intelligent Systems November/December 2003
2. Driankov D, Hellendoorn H., Reinfrank M.: An Introduction to Fuzzy Control Springer-Verlag Berlin 1993
3. Drigas A.., Kouremenos S., Vrettos S., Vrettaros J., Kouremenos D.: An expert system for job matching of the unemployed. Expert Systems with Applications **26** (2004) 217–224
4. Eirinaki M., Vazirgiannis M.: Web Mining for Web Personalization. ACM Transactions on Internet Technology, **3** (2003) 1-27
5. IEC 1131 - Programmable Controllers. Part 7 - Fuzzy Control Programming. Committee Draft CD 1.0 (Rel. 19 Jan 97)
6. Levary R.R., Lin C.Y.: Modelling the Software Development Process Using an Expert Simulation System Having Fuzzy Logic. Software-Practice And Experience **21** (1991)133–148
7. Saini H.S., Kamal R., Sharma A.N.: Web Based Fuzzy Expert System for Integrated Pest Management in Soybean. International Journal for Information Technology **8** (2002)
8. Yager R.R., Filev D.: Essentials of fuzzy modeling and control Wiley New York 1994