

# Reengineering of processes of classification in neural intellectual information systems of city development

Vadim Belousov<sup>1,\*</sup>, Pavel Kurochka<sup>1</sup>, Tatyana Nasonova<sup>1</sup>, Tamara Narezhnaya<sup>2</sup>, Dmitriy Spitsov<sup>3</sup>

<sup>1</sup> Voronezh State Technical University, Moscow Avenue, 14, 394026, Voronezh, Russia

<sup>2</sup> Moscow State University of Civil Engineering, Yaroslavskoe sh. 26, Moscow, 129337, Russia

<sup>3</sup> Moscow Institute of Architecture, Rozhdestvenka street 11, Moscow, 107031, Russia

**Abstract.** In this article reengineering of processes of classification of objects in neural intellectual information systems of city development in the conditions of multi criteria approach is considered. The universal algorithm of calculation of cumulative total opinion of experts of the city district on the basis of a median of Kemeni who allows to order in the most reliable and objective way clusters rankings in the complex multilevel structures representing objects of municipal economy is offered. For receiving reliable results of an indistinct clustering it is necessary to reproduce the specified algorithm repeatedly at a certain number of clusters for various initial indistinct splitting and to compare values of criterion function of indistinct splitting which were received, for the purpose of adoption of the final decision on a necessary indistinct clustering.

## 1 Introduction

The urban environment of development represents set of difficult great number of agents system which represents a great number of the artificial in common working agents - officials, deputies of a city meeting, citizens who have activity, autonomy are capable to make decisions on the basis of the established set of rules, interacting with an environment and the next agents. The agent is forced to choose the most priority purposes - intentions.

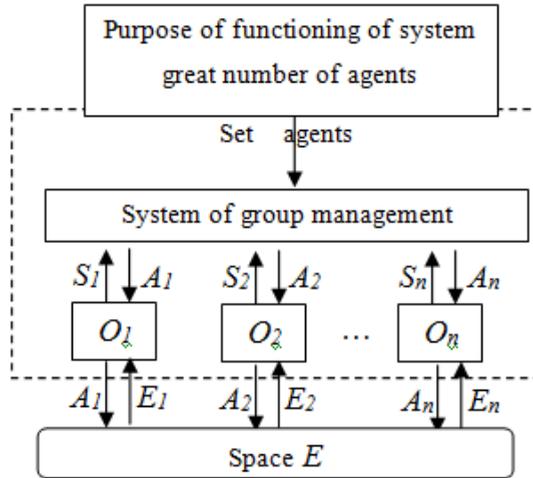
When developing the information management system city development most often applies technologies by means of which there is a formation of the mechanism of artificial intelligence of the agent most often: the clustering algorithms, neural networks, systems based on knowledge, natural language processing, trees of decisions, genetic algorithms, Bayes's theorem, fuzzy logic. The called intellectual systems based on knowledge can be joint within creation of great number of agents system. Let's develop a formal algorithm of functioning of great number of agents difficult system. Let's assume that the great number of agents system (fig. 1) consists of agents and system of group management [1].

Interaction of a great number of agents happens to other components of system as follows. Agents  $O_1 \div O_N$  accumulate information on their environment in the form of

---

\* Corresponding author: [belousov@vgasu.vrn.ru](mailto:belousov@vgasu.vrn.ru)

vectors of  $E_i$  and according to the  $S_i$  flowing a state send data to system of group management. Complexity of formation of a control system of a great number of agents consists in variety of tasks which they have to resolve. And here the fact that separate groups of agents are aimed at achievement of the local purposes is represented obvious. The structure, number, program and technical providing this or that group of agents depends on the local purpose and, directly, on opportunities of agents.



**Fig. 1.** Great number of agents.

Each agent of  $O_i$ , is capable to execute set of actions on the other hand. Each agent within achievement of goals of  $T_\mu$  in surrounding conditions:

$$f_\mu = \{f_{1\mu}, f_{2\mu}, \dots, f_{k_\mu\mu}\} \tag{1}$$

has to make set of the actions:

$$T_\mu = \{T_{1\mu}, T_{2\mu}, \dots, T_{q_\mu\mu}\} \tag{2}$$

which are characterized by some signs, such as type and intensity. Besides, on achievement of goals of  $T_\mu$  the set of other requirements:

$$Q_\mu = \{J(\mu), t_{dost}(\mu), n(\mu), \dots\}, \tag{3}$$

where by  $J(\mu)$  - criterion of efficiency,  $t_{dost}(\mu)$  - time of achievement of goals,  $n(\mu)$  - number of the agents involved for achievement of goals of  $T_\mu$  and other can be imposed.

If each of the  $vT_\mu$ ,  $v = \overline{1, q_\mu}$  elements, is present on type and intensity at the  $A_{ji}$  elements of the agent of  $O_i$  in the considered group, in other words  $T_{v\mu} \in A_i, v = \overline{1, q_\mu}$ , that for achievement of goals of  $T_\mu$  in the conditions of the project  $f_\mu^0$  will be required one agent.

Thus, the problem of identification of classes in the intelligent great number of agents information management system city development is urgent.

## 2 Problem definition

Let  $X = \{x_1, \dots, x_k\}$  - a set of objects which are exposed to the analysis on the basis of a set of quantitative and qualitative criteria  $G = \{G_1, \dots, G_n\}$ ;  $m$  - the number of experts who make assessment and testing of objects. It should be noted that an object is often investigated by several experts whose estimates can be contradictory or similar. It occurs owing to possible ambiguity of expert understanding of a solvable task, specificity of their knowledge, mistakes and inaccuracies.

Any  $r$  the qualifier faces the task of the multi criteria analysis consisting in classification of objects of a set  $X$  by criteria sets of  $G$ .

Let's say  $\mu_{G_i}(x_j)$  the number in the range  $[0,1]$  which characterizes belonging to a class  $x_j \in X$  by criterion  $G_i \in G$ : the more number  $\mu_{G_i}(x_j)$  the is higher assessment characterizing belonging to a class (group)  $x_j (j = \overline{1, k})$  on criterion  $G_i (i = \overline{1, n})$ .

Let's designate criterion  $G_i$  in the form of an indistinct set on a set  $\tilde{G}_i$  of classes  $X$  [1]:

$$\tilde{G}_i = \left\{ \frac{\mu_{G_i}(x_1)}{x_1}, \dots, \frac{\mu_{G_i}(x_k)}{x_k} \right\} \tag{4}$$

where  $\mu_{G_i}(x_j)$  designates as far as the element belongs to an indistinct set  $\tilde{G}_i$ .

Let's find degrees of accessory of an indistinct set (1) method of creation of functions of accessory by means of pair comparisons on Saati's scale. Let  $r$  be determined by the qualifier concerning each object  $x_j$  estimates by results of testing at equilibrium or no equilibrium criteria  $G_i$  by criteria respectively:

$$g_{j,i,r} = \mu_{G_i}, r(x_j) \text{ or } g_{j,i,r} = \mu_{G_i}, r^{\alpha_j}(x_j), \tag{5}$$

where  $\alpha_j$  - coefficient of relative importance of criterion:

$$G_i, \alpha_1 + \dots + \alpha_n = 1.$$

Let's say what on  $i$  criterion was given by qualifiers different values of estimates  $g_{j,i,r}$  which it is represent able through  $q_j^{y_i} (y_i = \overline{1, u_i})$  and the number of the qualifiers which appropriated to an object  $x_j$  assessment  $q_j^{y_i}$  corresponds:

$$k_{C_j}(q_i^{y_i}) \left( \sum_{y_i=1}^{u_i} k_{C_j}(q_i^{y_i}) = m \right) \tag{6}$$

Let's assume that estimates  $q_j^{y_i} (i = \overline{1, n})$  are ordered from bigger value to smaller. Then we will designate  $x_j$  any object  $G$  on a set of criteria through a multi set [2]:

$$C_i = \{k_{C_j}(q_1^1)q_1^1, \dots, k_{C_j}(q_1^{u_i})q_1^{u_i}, \dots, k_{C_j}(q_1^{u_n})q_1^{u_n}\} \tag{7}$$

The problem of classification of objects of a set of criteria comes down to classification of multi sets. Let's spread out set of all objects  $X$  to some classes by method of an indistinct clustering [3].

### 3 Indistinct clustering on the basis of multisets

Let's say that the set model has hierarchical structure. Change of external conditions attracts revision of number of levels of hierarchy and elements of structure. Are also calculated the weight of the set criteria. We will designate the main groups:

$$\xi^{ij} = \left\{ \xi_1^{ij}, \xi_2^{ij}, \dots, \xi_{M_{ij}}^{ij} \right\} \quad (8)$$

to which an object by the designated criteria can be carried in a look where  $i=1, \dots, p; j=1, \dots, n; p$  - number of parameters of assessment;  $n_i$  - amount of sub criteria of  $i$  of criterion;  $M_{ij}$  - the number of possible groups in  $j$  sub criteria of  $i$  of criterion.

The directions  $\xi_m^{ij}, m=1, \dots, M_{ij}$  are key subgroups within the corresponding sub criterion. They are in advance set and described. To everyone  $\xi_m^{ij}, m=1, \dots, M_{ij}$  we will designate a range of possible expressions:

$$\gamma_{\xi_m^{ij}} = \left\{ \gamma_{\xi_m^{ij}}^{(1)}, \gamma_{\xi_m^{ij}}^{(2)}, \dots, \gamma_{\xi_m^{ij}}^{(K_m^{ij})} \right\} \quad (9)$$

a set of linguistic formulations which correspond to different options.

Sizes  $\gamma_{\xi_m^{ij}}^{(k)}$  are calculated by experts, and each  $\gamma_{\xi_j}^k$  statement is estimated by a quantity of points  $V_{\xi_j}^k$ .

We will apply calculation of a median of sets to definition of the general assessment. And, at even quantity of estimates it makes sense to count an arithmetic average from the right and left median of sets [2].

Also it should be noted that ponder ability of the parameters allocated in sub criterion is often not identical. Experts will always note the big importance of some of them. To take into account opinions of experts about relative ponder ability of parameters, we will put weight  $w_{\xi_j}, j=1, \dots, M$  in compliance to each of them  $\sum_{j=1}^M w_{\xi_j} = 1$ , at the same time and

originally the  $w_{\xi_j} = \frac{1}{M}$ . Weight of parameters of sub criteria can be changed by experts further. Application of the procedure of pair comparisons with two outcomes is for this purpose justified.

The expert fills a matrix  $A = \|a_{ij}\|, i=1, \dots, M, j=1, \dots, M$ , at the same time:

$$a_{ij} = \begin{cases} 1, & \text{if } i \text{ more important } j \\ 0, & \text{if } i \text{ not more important } j \\ "-", & \text{if } i = j \end{cases} \quad (10)$$

Respectively, each binary relation built by rules of expression (3) is the matrix which is completely determined by the over diagonal part.

The number of elements pays off  $M \times \frac{(M-1)}{2}$ . Results of the procedure of pair comparisons influence revision of a set of scales defined earlier  $w_{\xi_j}$  :

$$w_{\xi_j} = \sum_j a_{ij} / \sum_{ij} a_{ij}, j = 1, \dots, M. \tag{11}$$

Calculation of size of concrete weight is almost unsalable task. Therefore, comparison of couples by experts has big practical applicability. Expression (4) is reasonable for calculation of scales  $w_{\xi_j}$  within one matrix of pair comparisons. But we need to consider opinion of a large number of experts, otherwise the result will be excessively subjective [4]. We will apply Kemeni's median to identification of the integrated indicator reflecting expert opinions.

Let's say that we have estimates of  $N$  experts. These estimates are presented in the form of the binary relations. The distance between expert estimates is size  $D$ :

$$D(A_1, A_2) = \sum |a_1(i, j) - a_2(i, j)|. \tag{12}$$

In other words, Kemeni's distance between the binary relations  $A_1, A_2$  described by matrixes:

$$\|a_1(i, j)\| \text{ and } \|a_2(i, j)\|$$

respectively is called the number  $D(A_1, A_2)$  where summation is made on all  $i, j$  from 1 to  $k$ , i.e. Kemeni's distance between the binary relations is equal to the sum of modules of differences, the elements standing on the same places in the matrixes corresponding to them [4].

To calculate total opinion of the commission of experts, we will resolve the following optimizing task:

$$\sum_{i=1}^N D(A_i, A) \rightarrow \min \tag{13}$$

The cumulative opinion of experts is a binary relation  $A$ , the sum of distances from which to expert answers will be minimum. It can be one of elements of a set  $A_i, i=1, \dots, N$ , i.e. one of opinions of experts of  $N$  and is the solution of the set optimizing task.

Let's develop an algorithm of identification of a median of Kemeni at which the new binary relation of  $A_N+1$  is a total opinion of expert group.

1. Let a over diagonal part in each binary relation of  $A_i, i=1, \dots, N$ , is set by  $N$  in the form of a binary vector of  $P_i, i=1, \dots, N$  with dimension  $R = M \times (M-1)/2$ .

2. Let's find a minimum of expression of  $F_j$  which is similar (6) on each element of the resulting  $P^*$  vector:

$$F_j = k_j \times |p_j^* - 1| + (N - k_j) \times p_j^*, j = 1, \dots, R, \tag{14}$$

where  $k_j$  - number of experts who delivered "1" in  $j$  cell of a vector of  $P_i, i=1, \dots, N$ .

3. We will transform a binary vector of  $P^*$  to the binary relation of  $A_N+1$ .

As  $\sum_{i=1}^N D(A_i, A) = 2 \sum_{j=1}^R F_j$ , by means of the described algorithm we can resolve an

optimizing task (6).

It is necessary to pay attention that such situation at even  $N$  at which  $k_j = N - k_j$  for some  $j$ , i.e. the number of the experts who gave a mark "0" equals to number of the experts who gave a mark "1" is possible. It can be considered as lack of coherence in expert group, and poll at least of one qualified specialist is necessary.

The developed algorithm of calculation of a median of Kemeni has simplicity, transparency and universality. The cumulative total opinion of commission of experts calculated on the set algorithm is the most reliable and objective that allows to order correctly rankings in complex multilevel structures and to create the adequate cumulative knowledge base of experts [3].

## 4 Conclusions

This approach to definition of classes of objects guarantees the analysis of the qualifiers given without bias, possible loss of information or/and its distortion. Quantitative and qualitative data including contradictory can act as criteria of testing.

Directly classification of data will be made in process of proximity to the best or worst object of city infrastructure on the basis of an indistinct clustering in metric space of multisets.

## References

1. V.E. Belousov, Messenger of VSU. Series: System analysis and information technologies **1**, 77-81 (2017)
2. V.E. Belousov, the Bulletin of VSU, Series: System analysis and information technologies **3**, 124-129 (2017)
3. V. Belousov, T. Averina, P. Kurochka, The 11th IEEE International Conference Application of Information and Communication Technologies. Aict2017 conference proceedings **1**, 164-169 (2017)
4. S. Khaykin, *Neural networks: full course, 2nd edition* (Williams publishing house, Moscow, 2006)