Municipal Creditworthiness Modelling by Clustering Methods

Petr Hájek, Vladimír Olej
Institute of System Engineering and Informatics
Faculty of Economics and Administration
University of Pardubice
Studentská 84, 532 10 Pardubice
Czech Republic
Petr.Hajek@upce.cz, Vladimír.Olej@upce.cz

Abstract

The paper presents the design of municipal creditworthiness parameters. Municipal creditworthiness modelling is realized by means of unsupervised methods, namely neural networks, cluster and fuzzy cluster analysis methods. Analysis of the gained results is based on clustering quality evaluation and further, on the comparison to results gained by the designed hierarchical structure of fuzzy inference system. Suitable interpretation of the created clusters is achieved this way.

1. Introduction

Municipal creditworthiness [1] is the ability of a municipality to meet its short-term and long-term financial obligations. It is determined by factors (parameters) relevant to the assessed object. High municipal creditworthiness shows a low credit risk, while the low one shows a high credit risk. Municipal creditworthiness evaluation is currently being realized by methods combining mathematical-statistical methods and expert opinion [1]. Their output is represented either by a score evaluating the municipal creditworthiness (scoring systems) or by an assignment of the municipalities to the j-th class $\omega_j \in \Omega$, $\Omega = \{\omega_1,\omega_2, \dots ,\omega_j, \dots ,\omega_c\}$ according to their creditworthiness (rating, unsupervised methods). Rating is an independent expert evaluation based on complex analysis of all known risk parameters of municipal creditworthiness, however, it is considered to be rather subjective. Municipalities are classified into classes $\omega_j \in \Omega$ by rating-based models. The classes are assigned to the municipalities by rating agencies.

Neural networks [2] are appropriate for municipal creditworthiness modelling due to their ability to learn, generalize and model non-linear relations. Nevertheless, the computational speed and robustness are retained. Municipal creditworthiness evaluation is considered to be a problem of classification, that is, it can be realized by various models of neural networks [2]. Classification can be realized by supervised methods (if classes $\omega_j \in \Omega$ are known) or unsupervised methods (if classes $\omega_j \in \Omega$ are not known). Statistical methods (discriminate analysis [3], logarithmic regression [3]), neural networks [4] and support vector machines [4] were used for the supervised methods. Statistical methods (e.g. multidimensional scaling [5]) were used for unsupervised methods. Only several municipalities of the Czech Republic have assigned the class $\omega_j \in \Omega$. Therefore, the article presents a design of municipal creditworthiness parameters of Czech municipalities and its modelling by neural networks [2], cluster analysis [6,7] and fuzzy cluster analysis [8]. The results of the methods are compared to the classification realized by the design of hierarchical structure of fuzzy inference system (HSFIS) [9]. The HSFIS represents a modelling of the decision-making process on the basis of expert knowledge in the field of municipal creditworthiness evaluation.

2. Municipal Creditworthiness Parameters Design

In [10] common categories of parameters there are mentioned namely economic, debt, financial and administrative categories. The economic, debt and financial parameters are pivotal. Economic parameters affect long-term credit risk. The municipalities with more diversified economy and more favourable social and economic conditions are better prepared for the economic recession. Debt parameters include the size and structure of the debt. Financial parameters inform about the budget implementation. Their values are extracted from the municipal budget. The design of parameters, based on previous correlation analysis of original parameters set [10] and recommendations of unable experts, can be realized as presented in Table 1.

Table 1. Municipal creditworthiness parameters design

Parameters	
Economic	$x_1 = PO_r$, PO_r is population in the r-th year.
	$x_2 = PO_r/PO_{r-s}$, PO_{r-s} is population in the year r-s, and s is the selected time period.
	$x_3 = U$, U is the unemployment rate in a municipality.
	$x_4 = \sum_{i=1}^{k}(PZO_i/PZ)^2$, PZO_i is the employed population of the municipality in the i-th economic sector, i = 1,2, … ,k, PZ is the total number of employed inhabitants, k is the number of the economic sector.
Debt	$x_5 = DS/OP$, $x_5 \in <0,1>$, DS is debt service, OP are periodical revenues.
	$x_6 = CD/PO$, CD is total debt.
	$x_7 = KD/CD$, $x_7 \in <0,1>$, KD is short-term debt.
Financial	$x_8 = OP/BV$, $x_8 \in R^+$, BV are current expenditures.
	$x_9 = VP/CP$, $x_9 \in <0,1>$, VP are own revenues, CP are total revenues.
	$x_{10} = KV/CV$, $x_{10} \in <0,1>$, KV are capital expenditures, CV are total expenditures.
	$x_{11} = IP/CP$, $x_{11} \in <0,1>$, IP are capital revenues.
	$x_{12} = LM/PO$, [Czech Crowns], LM is the size of the municipal liquid assets.

Based on the presented facts, the following data matrix **P** can be designed

$$\mathbf{P} = \begin{array}{c|ccccc:c} & x_1 & \dots & x_k & \dots & x_{12} & \omega \\ \hline o_1 & x_{1,1} & \dots & x_{1,k} & \dots & x_{1,12} & \omega_{1,j} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ o_i & x_{i,1} & \dots & x_{i,k} & \dots & x_{i,12} & \omega_{i,j} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ o_n & x_{n,1} & \dots & x_{n,k} & \dots & x_{n,12} & \omega_{n,j} \end{array},$$

where $o_1,o_2, \dots ,o_i, \dots ,o_n$ are objects (municipalities), $o_i \in O$, x_k is the k-th parameter, $x_{i,k}$ is the value of the parameter x_k for the i-th object o_i, $\omega_{i,j}$ is the j-th class assigned to the i-th object o_i, $\mathbf{p}_i = (x_{i,1},x_{i,2}, \dots ,x_{i,k}, \dots ,x_{i,12})$ is the i-th pattern, $\mathbf{x} = (x_1,x_2, \dots ,x_{12})$ is the parameters vector.

3. Design of model for municipal creditworthiness classification

Municipal creditworthiness modelling represents a classification problem. It is generally possible to define it this way: Let $F(\mathbf{x})$ be a function defined on a set A, which assigns picture $\hat{x}$ (the value of the function from a set B) to each element $\mathbf{x} \in A$, $\hat{x} = F(\mathbf{x}) \in B$, $F : A \rightarrow B$.

The problem defined this way it is possible to model by supervised methods (if classes $\omega_j \in \Omega$ of the objects are known) or by unsupervised methods (if classes $\omega_j \in \Omega$ are not known). Several Czech municipalities have assigned the class $\omega_j \in \Omega$. Therefore it is appropriate to model the municipal creditworthiness by e.g. neural networks (Kohonen's self-organizing feature maps (SOFM) [11] and adaptive resonance theory (ART) [12]), cluster analysis (CA) and fuzzy cluster analysis (FCA). It is possible to create clusters on the basis of the objects' similarity by using these methods. The results' quality of the presented methods can be evaluated by clustering quality indexes [13] or by classification error in case the assignment of objects $o_i \in O$ to classes $\omega_j \in \Omega$ is known. The assignment is realized by the design of hierarchical structure of fuzzy inference system (HSFIS). The number of clusters m = 7 is set by empirical experience and indexes of clustering quality [13]. The assignment of creditworthiness classes to clusters created by unsupervised methods makes the suitable interpretation by means of HSFIS base rules possible. Based on the mentioned facts, the design of municipal creditworthiness evaluation model is realized as presented in Figure 1.

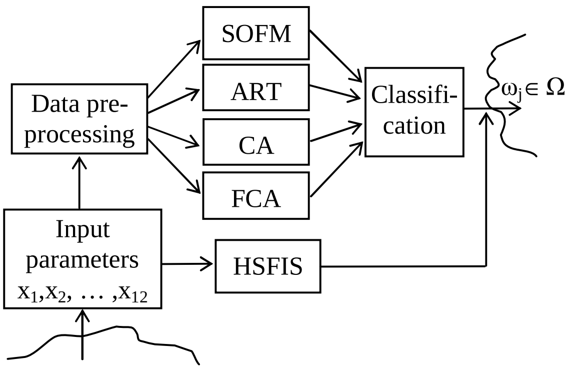

Figure 1. Design of municipal creditworthiness evaluation model

3.1. Clustering by Kohonen's self-organizing feature maps

Kohonen's self-organizing feature maps [2,11] are based on competitive learning strategy. The input layer serves the distribution of the input patterns $\mathbf{p}_i$, i = 1,2, … ,n. The neurons in the competitive layer serve as the representatives (Codebook Vectors), and they are organized into topological structure (most often a two-dimensional grid), which designates the neighbouring network neurons.

First, the distances d_j are computed between pattern $\mathbf{p}_i$ and weights of all neurons $w_{i,j}$ in the competitive layer according to the relation

$$d_j = \sum_{k=1}^{N} (\mathbf{p}_i - w_{i,j})^2 , \quad (1)$$

where j goes over s neurons of competitive layer, j = 1,2, … ,s, $\mathbf{p}_i$ is the i-th pattern, i = 1,2, … ,n, $w_{i,j}$ are synaptic weights. The winning neuron j* (Best Matching Unit (BMU)) is chosen, for which the distance d_j from the given pattern $\mathbf{p}_i$ is minimum. Synaptic weights of this neuron are adapted in order to approximate the i-th pattern $\mathbf{p}_i$. The aim of the SOFM learning is to approximate the probability density of the real input vectors $\mathbf{p}_i \in R^n$ by the finite number of representatives $\mathbf{r}_j \in R^n$, where j = 1,2, … ,s. When the representatives $\mathbf{r}_j$ are identified, the representative $\mathbf{r}_{j*}$ of the BMU is assigned to each vector $\mathbf{p}_i$. In the learning process of the SOFM, it is necessary to define the concept of neighbourhood function, which determines the range of cooperation among the neurons, i.e. how many representatives $\mathbf{r}_j$ in the neighbourhood of the BMU will be adapted, and to what degree. Gaussian neighbourhood function is in common use, which is defined as

$$h(j^*, j) = e^{\left(-\frac{d_E^2(j^*, j)}{\lambda^2(t)}\right)} , \quad (2)$$

where h(j*,j) is neighbourhood function, $d^2_E(j^*,j)$ is Euclidean distance of neurons j* and j in the grid, $\lambda(t)$ is the size of the neighbourhood in time t. After the BMUs are found, the adaptation of synaptic weights $w_{i,j}$ follows. The principle of the sequential learning algorithm [11] is the fact, that the representatives $\mathbf{r}_{j*}$ of the BMU and its topological neighbours move towards the actual input vector $\mathbf{p}_i$ according to the relation

$$w_{i,j}(t+1) = w_{i,j}(t) + \eta(t)h(j^*, j)[\mathbf{p}_i(t) - w_{i,j}(t)] , \quad (3)$$

where $\eta(t) \in (0,1)$ is the learning rate. The batch learning algorithm of the SOFM [11] is a variant of the sequential algorithm. The difference consists in the fact that the whole training set O_{train} passes through the SOFM, and only then the synaptic weights $w_{i,j}$ are adapted. The adaptation is realized by replacing the representative $\mathbf{r}_j$ with the weighted average of the input vectors $\mathbf{p}_i$ [11]. The input parameters of the designed SOFM model are specified in Table 2. Using the SOFM as such can detect the structure in the data. The K-means algorithm can be applied to the adapted SOFM in order to find clusters as presented in Figure 2.

Table 2. Input parameters of the SOFM model

Parameter	Learning algorithm	s	Initial $\lambda(t)$	Final $\lambda(t)$	Epochs
Value	Batch	108	3	1	24

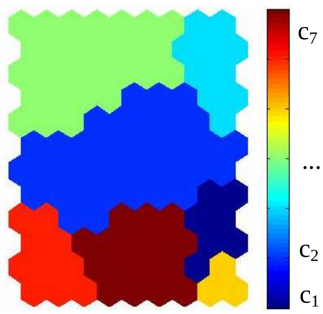

Figure 2. Clustering of the SOFM by K-means algorithm

The K-means algorithm belongs to non-hierarchical algorithms of cluster analysis, where patterns $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_i, \ldots, \mathbf{p}_n$ are assigned to clusters $c_1, c_2, \ldots, c_r, \ldots, c_m$. Interpretation of clusters is realized by the values of parameters $x_1, x_2, \dots, x_{12}$ for individual representatives $\mathbf{r}_j$ (Figure 3).

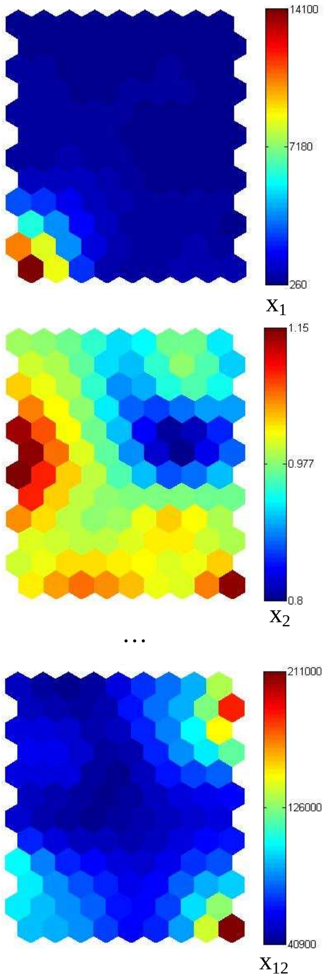

Figure 3. Values of parameters $x_1, x_2, \dots, x_{12}$ for individual representatives $\mathbf{r}_j$

3.2. Clustering by adaptive resonance theory

Adaptive resonance theory [12] represents the model of neural network based on unsupervised learning, by which new information can be learnt without damaging information stored previously. Neural network ART1 [12] works with binary values, while ART2 [14] works with the real values of patterns. Basic feature of the ART-type neural networks consists in the fact that weights have to be adapted by every exchange of pattern $\mathbf{p}_i$ between comparative and recognition layers till a stable state is achieved. This process consists of following stages: neural network's initialization, recognition, comparison, search and adaptation. Within the initialization, the initial state of neural network is set, i.e. the weights among neurons of comparative and recognition layers $w^P_{i,j}$ and $w^R_{j,i}$ and the vigilance ρ, where $0 \geq \rho \leq 1$. Input patterns are transformed through feed-forward weights $w^P_{i,j}$ to neurons in recognition layer during the recognition process as follows

$$y_j = \sum_{i=1}^{n} w^P_{i,j} x_i \ , \tag{4}$$

where y_j is output of the j-th neuron, $j = 1,2, \dots, m$. An external part of the ART neural network, so-called vigilance test [12] is responsible for data flow control. Following the transformation, the input pattern $\mathbf{p}_i$ is compared to every stored pattern in the recognition layer. The neuron is selected for which $y_{j*} = \max(y_j)$, where $j = 1,2, \dots, m$. The degree of similarity between an input pattern $\mathbf{p}_i$ and recognition patterns is evaluated by the vigilance test. The gained result S represents the vigilance according to the relation

$$S = \frac{\sum w^P_{i,j*} x_i}{\sum x_i} > \rho \ . \tag{5}$$

If $S>\rho$, then the corresponding class $\omega_j \in \Omega$ is found for the input pattern $\mathbf{p}_i$, else the process continues in order to find the pattern in the recognition layer for which the relation holds. If the pattern is found that passed the vigilance test, then the weights $w^P_{i,j}$ and $w^R_{j,i}$ are adapted as follows

$$w'^P_{i,j*} = \frac{w^R_{j*,i} x_i}{0.5 + \sum_{w=0}^{N-1} w^R_{j*,i} x_i} \ , \ w'^R_{j*,i} = w^R_{j*,i} x_i \ . \tag{6}$$

The neural network model of ART2-type for municipal creditworthiness modelling is specified by following parameters: the vigilance $\rho = 0.5$, learning rate $\eta(t) = 0.1$ and number of epochs $n_e = 100$. It is possible to realize the interpretation of clusters $c_1, c_2, \dots, c_7$ by the values of clusters' representatives (Table 3). The structure of the ART2-type neural network is presented in Figure 4.

Table 3. Clusters' representatives

	x_1	x_2	x_3	x_4	…	x_{12}
c_1	5991.3	1.02	8.44	0.19	…	64188.8
c_2	413.0	1.10	3.67	0.16	…	48148.4
…	…	…	…	…	…	…
c_7	324.5	1.09	11.31	0.19	…	63311.9

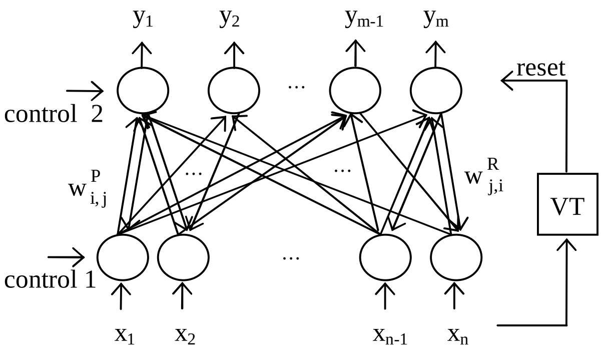

Figure 4. Structure of the ART2-type neural network, where VT is vigilance test

3.3. Clustering by cluster analysis methods

The cluster analysis [6,7] belongs to the methods which deal with the search for the similarity among multidimensional data objects and with their classification to clusters. The classes are not assigned to data objects. The number of clusters is mostly unknown, too. The found clusters represent the data structure only with reference to the selected parameters. This method does not contain a technique capable of distinguishing the significant and insignificant parameters, it only distinguishes the clusters. The goal of the municipal creditworthiness evaluation is the classification of the objects (municipalities) to the creditworthiness classes. In terms of definition of the cluster analysis scope, the data are standardized by normalization of each of the parameters to its Z-score [6]. The standardization facilitates mutual comparison of parameters' values (their average is 0 and the standard deviation is 1). The positive values are then above-average and the negative are below the average. All the parameters are of quantitative type. Therefore the distance measures can be used. Further it is necessary to choose the clustering algorithm and to resolve upon the expected number of the clusters. Both the mentioned decisions have influence on the results interpretation. There are two basic algorithms of clustering, namely hierarchical and partitioning algorithms [7]. The hierarchical algorithms construct a tree structure of the clusters, so-called dendrogram [7]. These algorithms are not suitable for the analysis of extensive samples, the results are affected by outlying objects and the undesirable preceding combinations persist in the analysis. In the partitioning algorithms [7] (K-modes, K-means algorithms, etc.) the objects are assigned to the number of clusters given in advance. First step is the setting of initial cluster centres and all the objects situated inside the given distance to a cluster centre are assigned to this cluster. The choice of the initial cluster centres is crucial. The K-means algorithm is used for the municipal creditworthiness modelling.

The goal of the K-means algorithm is the minimization of objective function J as follows

$$J = \sum_{r=1}^{m}\sum_{i=1}^{n}\left\|\mathbf{p}_i^r - \mathbf{c}_r\right\|^2 , \tag{7}$$

where $\mathbf{p}_i^r$ is the i-th pattern belonging to the r-th cluster and $\mathbf{c}_r$ is the cluster centre of the r-th cluster.

The results of the cluster analysis are negatively affected by the existence of outlying objects and by multicolinearity of the parameters [6]. The outlying objects are identified by the Mahalanobis distance and removed in consequence. The multicolinearity has not been noticed. The K-means algorithm for municipal creditworthiness modelling is specified by following parameters: number of clusters m = 7, initial cluster centres and maximum number of iterations. The initial cluster centres are set up by the hierarchical algorithm (Ward's method [6]), while the maximum number of iterations $n_i = 100$.

The results can be interpreted by the rules designed in terms of the known classification of the objects to the clusters. For each cluster c_r, where $r \in \{1,2, \dots ,7\}$, rules $V_{r,k}$ were created, where k is the sequence of a rule for the r-th cluster c_r. The algorithm PART (partial decision trees) was used for the rules creation [15]. The rules are presented in Table 4.

Table 4. Rules representing clusters $c_1, c_2, \dots ,c_7$

$V_{1,1}$	IF $x_{10} \le 0.324$ AND $x_7 > 0.861$ AND $x_3 \le 13.58$ AND $x_5 \le 0.206$ AND $x_2 > 0.896$	THEN r=1
$V_{1,2}$	IF $x_6 \le 1021.1$ AND $x_{10} \le 0.332$ AND $x_7 > 0.25$ AND $x_1 \le 1513$ AND $x_5 \le 0.311$ AND $x_3 \le 18.627$ AND $x_6 > 87.43$	THEN r=1
$V_{1,3}$	IF $x_5 \le 0.071$ AND $x_3 \le 9.15$ AND $x_3 \le 8.387$	THEN r=1
$V_{1,4}$	IF $x_4 \le 0.249$ AND $x_{10} \le 0.266$ AND $x_5 \le 0.006$	THEN r=1
$V_{2,1}$	IF $x_7 \le 0.438$ AND $x_8 \le 0.511$ AND $x_2 > 0.917$	THEN r=2
…		
$V_{7,4}$	IF $x_5 \le 0.071$ AND $x_3 > 9.15$ AND $x_6 \le 5807.03$ AND $x_1 > 110$ AND $x_3 > 12.632$	THEN r=7

First, rules $V_{r,k}$ are induced initially. Then, they are refined by the combination of decision tree generation and separate and conquer paradigm. The set of rules is optimized this way while high computational costs and over pruning are eliminated.

3.4. Clustering by fuzzy cluster analysis methods

Clusters are disjoint subsets of data objects' set in cluster analysis. The clusters can overlap in fuzzy cluster analysis. Let $\mu_r \subseteq O$, where μ_r are fuzzy sets, O is set of data objects, $O = \{o_1, o_2, \dots, o_n\}$ and r is the cluster index, then

$$0 \le \mu_r(o_i) \le 1, \sum_{r=1}^{m} \mu_r(o_i) = 1, 0 < \sum_{i=1}^{n} \mu_r(o_i) < 1, \qquad (8)$$

where $\mu_r(o_i)$ is the membership degree of the i-th object into the r-th cluster, m is the number of the clusters, o_i is the i-th object, $i = 1,2, \dots ,n$. Each object must belong at least to one cluster (8). The sum of all membership degrees of the i-th object into all m clusters equals to 1 (8). All the objects of the set O might not belong to one cluster with maximum membership degree and each cluster might not be empty (8).

Municipal creditworthiness modelling is realized by fuzzy C-means algorithm (FCM) and Gustafson Kessel algorithm (GKA) [8]. Contrary to the FCM, by which it is possible to detect the clusters of spherical shape only, the clusters of different shapes and orientations can be detected by the GKA algorithm, because each cluster is described by its centre and by special matrix H_r, $r = 1,2, \dots ,m$ in the GKA algorithm. Covariance matrix S_r is applied in calculation of the matrix H_r according to the following

$$S_r = \frac{\sum_{i=1}^{n} (\mu_r^{(a)}(\mathbf{u}_i))^b (\mathbf{u}_i - \mathbf{s}_r^{(a)})^T (\mathbf{u}_i - \mathbf{s}_r^{(a)})}{\sum_{i=1}^{n} (\mu_r^{(a)}(\mathbf{u}_i))^b}, \qquad (9)$$

where a is the order of fuzzy pseudo-partition, $\mathbf{u}_i$ is the vector of membership degrees of the i-th object, b is a parameter specified in advance, $\mathbf{v}_r$ is vector of coordinates of the r-th centre. The matrix H_r is calculated this way

$$H_r = (\det(S_r))^{\frac{1}{n}} S_r^{-1}. \qquad (10)$$

Input parameters of the FCM and GK algorithms are presented in Table 5.

Table 5. Input parameters of the FCM and GK algorithms, where a represents the weighting exponent which determines the fuzziness of the clusters, φ represents the termination tolerance of the clustering method, β represents the maximal ratio between the maximum and minimum eigenvalue of the covariance matrix.

Input parameters of the FCA	m	a	φ	β
Value for the FCM algorithm	7	2	0.001	-
Value for the GF algorithm	7	2	0.001	10^{15}

The clusters $c_1, c_2, \dots ,c_7$ can be interpreted by rules. In general, the rules can be presented as follows

IF x_1 is $A_1^{(k)}$ AND x_2 is $A_2^{(k)}$ AND … x_k is $A_j^{(k)}$ AND … AND x_N is $A_m^{(k)}$ THEN c_r, (11)

where:
- $k = 1,2, \dots ,N$, N represents the number of input variables,
- $j = 1,2, \dots ,m$, represents the assignment of linguistic variables relating to sets X_k,
- $A_1^{(k)}, A_2^{(k)}, \dots ,A_j^{(k)}, \dots ,A_m^{(k)}$ represent linguistic variables corresponding to fuzzy sets $\mu_1^{(k)}(x), \mu_2^{(k)}(x), \dots ,\mu_j^{(k)}(x), \dots ,\mu_m^{(k)}(x)$.

The rules created by the GK algorithm are presented this way

IF x_1 is Low AND x_2 is Low AND x_3 is Middle AND x_4 is Low AND x_5 is Low AND x_6 is Middle AND x_7 is Low AND x_8 is Low AND x_9 is Low AND x_{10} is Low AND x_{11} is Low AND x_{12} is High THEN c_1

IF x_1 is Middle AND x_2 is Middle AND x_3 is Middle AND x_4 is Middle AND x_5 is High AND x_6 is Low AND x_7 is Middle AND x_8 is Low AND x_9 is High AND x_{10} is High AND x_{11} is High AND x_{12} is Low THEN c_2

… (12)

IF x_1 is Middle AND x_2 is High AND x_3 is Middle AND x_4 is Middle AND x_5 is Middle AND x_6 is High AND x_7 is Middle AND x_8 is Low AND x_9 is High AND x_{10} is High AND x_{11} is High AND x_{12} is Low THEN c_7

Non-spherical clusters with various orientations can be created by the GK algorithm (Figure 5) contrary to the FCM algorithm (Figure 6).

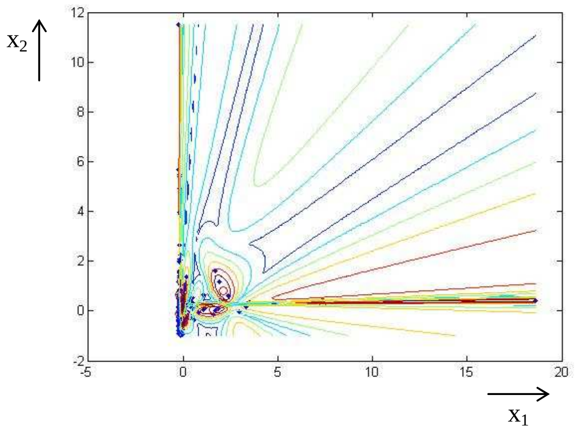

Figure 5. Clusters created by the GK algorithm Cluster shapes are denoted by ellipses. Clusters created by the FCM algorithm are spherical. Clusters created by the GK algorithm are non-spherical and variously oriented.

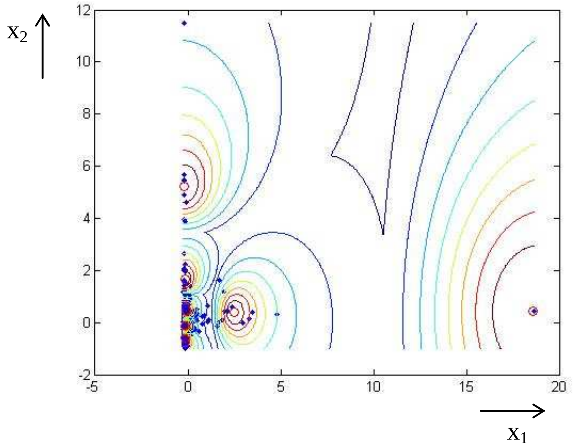

Figure 6. Clusters created by the FCM algorithm

4. Hierarchical structure of fuzzy inference systems design for municipal creditworthiness evaluation

General structure of fuzzy inference system (FIS) [16] contains the fuzzification process by means of input membership functions, construction of base rules (BRs) or automatic extraction of rules from the input data, application of operators (AND, OR, NOT) in rules, implication and aggregation within rules and the defuzzification process of obtained outputs to the crisp values. Based on the general structure of FIS, three fundamental types of FIS can be designed [16], i.e. Mamdani-type, Takagi-Sugeno-type and Tsukamoto-type.

Let $x_1, x_2, \dots, x_k, \dots, x_N$ be the input variables defined in the reference sets $X_1, X_2, \dots, X_k, \dots, X_N$ and let y be the output variable defined in the reference set Y. Then FIS has N input variables and one output variable. Each set X_k, $k = 1,2, \dots, N$, can be divided into p_j, $j = 1,2, \dots, m$ fuzzy sets $\mu_1^{(k)}(x), \mu_2^{(k)}(x), \dots, \mu_{pj}^{(k)}(x), \dots, \mu_m^{(k)}(x)$. Individual fuzzy sets $\mu_1^{(k)}(x), \mu_2^{(k)}(x), \dots, \mu_{pj}^{(k)}(x), \dots, \mu_m^{(k)}(x)$, $k = 1,2, \dots, n$; $j = 1,2, \dots, m$ represent the assignment of linguistic variables relating to sets X_k. The set Y is also divided into p_z, $z = 1,2, \dots, o$ fuzzy sets $\mu_1(y), \mu_2(y), \dots, \mu_z(y), \dots, \mu_o(y)$. The fuzzy sets $\mu_1(y), \mu_2(y), \dots, \mu_z(y), \dots, \mu_o(y)$ represent the assignment of linguistic variables for the set Y. Then the Mamdani-type FIS rule can be put as follows [14]

IF x_1 is $A_1^{(k)}$ AND x_2 is $A_2^{(k)}$ AND … AND x_N is $A_{pj}^{(k)}$ THEN y is B, (13)

where:
- $k = 1,2, \dots, N$, $j = 1,2, \dots, m$,
- $A_1^{(k)}, A_2^{(k)}, \dots, A_{pj}^{(k)}$ represent linguistic variables corresponding to fuzzy sets $\mu_1^{(k)}(x), \mu_2^{(k)}(x), \dots, \mu_{pj}^{(k)}(x), \dots, \mu_m^{(k)}(x)$,
- B represents linguistic variable corresponding to fuzzy sets $\mu_1(y), \mu_2(y), \dots, \mu_z(y), \dots, \mu_o(y)$, $z = 1,2, \dots, o$.

Let's have a given Mamdani-type FIS. Then the number of rules in this FIS is defined according to the relation

$$N_R = z^N, \tag{14}$$

where:
- N_R is number of rules,
- z is number of membership functions in FIS,
- N is number of input variables.

Due to a great number of N, FIS can be ineffective with regard to the increase of N_R. The design of the HSFIS is one of the ways leading to the decrease of rules number N_R [9]. The BRs reduction lowers the computational cost and makes FIS interpretation possible. The combination of tree and cascade structure of the HSFIS is used for municipal creditworthiness evaluation (Figure 7). The decrease of rules number N_R is obtained by the design of this model. In addition to rules reduction, the model design should reproduce the expert's decision-making process in municipal creditworthiness evaluation with the intent to consider the similarity and mutual relations of parameters $x_1, x_2, \dots, x_{12}$.

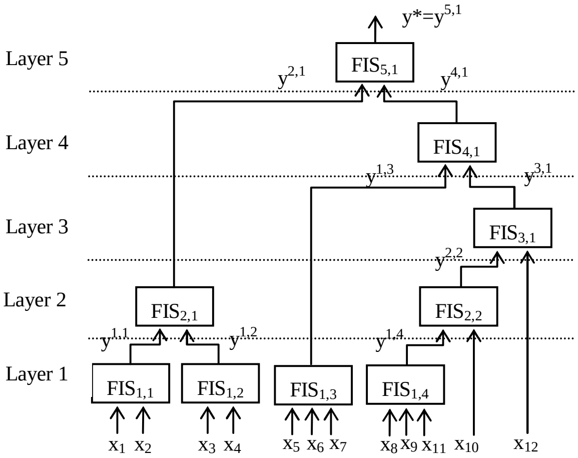

Figure 7. Design of HSFIS for municipal credit-worthiness evaluation, where $x_1, x_2, \dots, x_{12}$ are input variables, $y_{1,1}, y_{1,2}, \dots, y_{5,1}$ are outputs of subsystems $FIS_{1,1}, FIS_{1,2}, \dots, FIS_{5,1}$, $L = 5$ is the number of HSFIS layers.

The HSFIS model can be formalized by BRs $R^{h_{1,1}}, R^{h_{1,2}}, \dots, R^{h_{5,1}}$ and outputs $y^{1,1}, y^{1,2}, \dots, y^*$ of HSFIS subsystems this way:

Layer 1: $FIS_{1,1}$: $R^{h_{1,1}}$: IF x_1 is $A_1^{h_{1,1}}$ AND x_2 is $A_2^{h_{1,1}}$ THEN $y^{1,1}$ is $B^{h_{1,1}}$,

$FIS_{1,2}$: $R^{h_{1,2}}$: IF x_3 is $A_3^{h_{1,2}}$ AND x_4 is $A_4^{h_{1,2}}$ THEN $y^{1,2}$ is $B^{h_{1,2}}$,

… (15)

Layer 5: $FIS_{5,1}$: $R^{h_{5,1}}$: IF $y^{2,1}$ is $B^{h_{2,1}}$ AND $y^{4,1}$ is $B^{h_{4,1}}$ THEN y^* is $B^{h_{5,1}}$,

$$y^{1,1}(B^{h_{1,1}}) = \frac{\sum_{j=1}^{p} y_j^{1,1} \times \mu_{B^{h_{1,1}}}(y_j^{1,1})}{\sum_{j=1}^{p} \mu_{B^{h_{1,1}}}(y_j^{1,1})}, \quad (16)$$

$$y^{1,2}(B^{h_{1,2}}) = \frac{\sum_{j=1}^{p} y_j^{1,2} \times \mu_{B^{h_{1,2}}}(y_j^{1,2})}{\sum_{j=1}^{p} \mu_{B^{h_{1,2}}}(y_j^{1,2})}, \quad (17)$$

…

$$y^*(B^{h_{5,1}}) = \frac{\sum_{j=1}^{p} y_j^{5,1} \times \mu_{B^{h_{5,1}}}(y_j^{5,1})}{\sum_{j=1}^{p} \mu_{B^{h_{5,1}}}(y_j^{5,1})}, \quad (18)$$

where:
- $x_1, x_2, \dots, x_{12}$ are input parameters,
- $A_1^{h_{1,1}}, A_2^{h_{1,1}}, \dots, A_{12}^{h_{5,1}}$ represent linguistic variables corresponding to fuzzy sets $\mu_1^{h_{1,1}}(x_j), \mu_2^{h_{1,1}}(x_j), \dots, \mu_{12}^{h_{5,1}}(x_j)$,
- $B^{h_{1,1}}, B^{h_{1,2}}, \dots, B^{h_{5,1}}$ represent linguistic variables corresponding to fuzzy sets $\mu^{h_{1,1}}(y^{1,1}), \mu^{h_{1,2}}(y^{1,2}), \dots, \mu^{h_{5,1}}(y^*)$,
- $\mu_{B^{h_{1,1}}}(y_j^{1,1}), \mu_{B^{h_{1,2}}}(y_j^{1,2}), \dots, \mu_{B^{h_{5,1}}}(y_j^{5,1})$ are membership functions values of the aggregated fuzzy set for values $y_j^{1,1}, y_j^{1,2}, \dots, y_j^{5,1}$ from the reference sets.

The numbers and shapes of input and output membership functions and BRs are defined for the designed model. Classification of municipalities into classes $\omega_j \in \Omega$ according to creditworthiness by the HSFIS is presented in Figure 8.

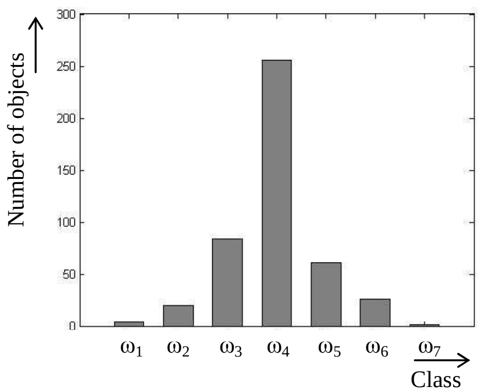

Figure 8. Classification of municipalities into classes by the HSFIS

5. Analysis of the results

Municipal creditworthiness modelling is realized by following unsupervised methods: SOFM (completed by K-means algorithm), ART, CA (K-means algorithm) and FCA (GKA, FCM). Every object o_i, $o_i \in O$ is assigned to one of the clusters $c_1, c_2, \dots, c_r, \dots, c_m$ by means of the SOFM, ART and CA. It is possible to assign an object o_i to clusters $c_1, c_2, \dots, c_r, \dots, c_m$ with certain membership degree by the FCA. In order to realize the comparison of mentioned methods, every object $o_i \in O$ is assigned to the cluster c_r for which the membership degree is maximum. Clusters can be interpreted by cluster centres (average values of all parameters within one cluster), group of distant points, decision trees or by rules [7]. The labelling of clusters $c_1, c_2, \dots, c_r, \dots, c_m$ with classes $\omega_1, \omega_2, \dots, \omega_j, \dots, \omega_c$ is possible due to suitable interpretation. Classification of municipalities $o_1, o_2, \dots, o_n$ into classes $\omega_1, \omega_2, \dots, \omega_j, \dots, \omega_c$ by unsupervised methods is presented in Figure 10.

Quality of clustering can be compared based on clustering quality indexes [13] or by classification error in case the assignment of objects $o_i \in O$ to classes $\omega_j \in \Omega$ is known in advance. Following clustering quality indexes are selected for the comparison: Separation index (S), Xie-Beni index (XB) and Dunn index (DI) [13]. The S index uses a minimum distance separation for partition validity. The XB index aims to quantify the ratio of the total variation within clusters and the separation of clusters. On the contrary to S index and XB index, the DI index uses a hard partition clustering results. They are defined according to the following

$$S(r) = \frac{\sum_{i=1}^{n}\sum_{r=1}^{m}(\mu_{i,r})^2 \left\|\mathbf{p}_i - \mathbf{v}_r\right\|^2}{n \times \min_{r,k}\left\|\mathbf{v}_k - \mathbf{v}_r\right\|^2}, \tag{19}$$

$$XB(r) = \frac{\sum_{i=1}^{n}\sum_{r=1}^{m}(\mu_{i,r})^2 \left\|\mathbf{p}_i - \mathbf{v}_r\right\|^2}{n \times \min_{i,r}\left\|\mathbf{p}_i - \mathbf{v}_r\right\|^2}, \tag{20}$$

$$DI(r) = \min_{r \in m}\left\{\min_{k \in m, k \neq r}\left\{\frac{d_{min}(\mathbf{v}_r, \mathbf{v}_k)}{\max_{l \in m}\{d_{max}(\mathbf{v}_r, \mathbf{v}_k)\}}\right\}\right\}, \tag{21}$$

where k,l,r are cluster indexes, $\mu_{i,r}$ is membership degree of the i-th object to the r-th cluster, $\mathbf{p}_i$ is the i-th pattern, $\mathbf{v}_r$ is centre of the r-th cluster, d_{min} is minimum and d_{max} is maximum Euclidean distance.

Many experiments were carried out. The mean values of the indexes for clustering methods are presented in Table 6.

Table 6. Clustering quality indexes

	SOFM	ART	CA	FCM	GK
S	0.0029	0.0048	0.0002	0.0004	0.0002
XB	7.280	7.816	9.217	15.189	6.033
DI	0.0503	0.0323	0.0464	0.0013	0.0003

The results show, the best separated clusters are created by the GK algorithm. It achieves the minimum values of all indexes. However, indexes S and XB prefer the FCA methods with the intent that the membership degree $\mu_{i,r}$ of the i-th object to the r-th cluster is taken into account. The other methods provide similar results. Further, the results are compared to the classification realized by the HSFIS model. The results of the classification evaluated by mean square error MSE and mean error ME are presented in Table 7.

Table 7. Classification errors

	SOFM	ART	CA	FCM	GK
MSE	1.909	2.622	2.002	1.522	1.949
ME	1.024	1.170	1.175	0.916	1.046

The comparison of classification results of unsupervised methods (Figure 9) to those ones of the HSFIS model shows that the classification made by the FCM algorithm and SOFM is the most similar to that one made by the HSFIS model. Classification made by the ART2-type neural network represents the classes gained by the HSFIS with regard to the ME only.

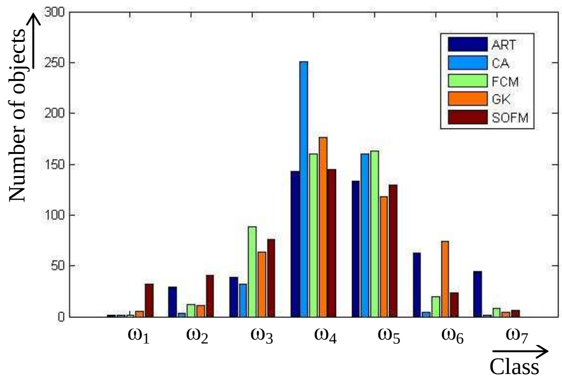

Figure 9. Classification of municipalities into classes by unsupervised methods

6. Conclusion

The paper presents the design of municipal creditworthiness parameters. Municipal creditworthiness modelling is realized by unsupervised methods. The following methods are selected for the comparison: SOFM, ART, CA and FCA. The SOFM makes the quick detection of both distinct and similar clusters possible. It is efficient in handling large data sets and is also robust when the data set is noisy. The ART is characterized by stabilized learning and the degree of match required can be controlled by a user. It is a stable and quick neural network with low number of input parameters. It makes the over-learning possible without stored knowledge degradation. The CA allows the creation of spherical clusters only. The advantage of the CA consists in low computational costs and easy realization. The drawback of the CA is the convergence of objective function J to local minimum. On the contrary to mentioned methods, the FCA makes the clusters' overlapping possible. As the result, the input patterns belong to each cluster with certain membership degree. Clusters with non-spherical shapes can be detected by the GK algorithm which results in higher computational costs.

The goal of the paper is the comparison of the methods with respect to clustering quality. It is evaluated by both clustering quality indexes (S, FB, DI) and classification error since the assignment of municipalities to classes is known in advance. The assignment is realized by means of the HSFIS model. The assignment of creditworthiness classes to clusters created by unsupervised methods makes the suitable interpretation of clusters possible. As the results show, the GK algorithm is suitable with respect to clustering quality indexes and SOFM and FCM algorithm are preferable due to low classification errors MSE and ME. The gained results can be used in the design of models for municipal creditworthiness evaluation by rating agencies, banks and municipalities. The presented methods are suitable for modelling of economic, social and environmental processes, as well as technical processes.

It would be beneficial to realize the municipal creditworthiness modelling by hybrid methods combining the advantages of unsupervised methods and fuzzy inference systems in further research.

Experiments were carried out in MATLAB 6.5 (unsupervised methods) and Matlab Simulink (HSFIS) in MS Windows XP operation system.

7. References

[1] V. Olej, P. Hájek, "Modelling of Municipal Rating by Unsupervised Methods", WSEAS Transactions on Systems, WSEAS Press, No.6, 2006, pp. 1679-1686.

[2] S. S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice-Hall, Upper Saddle River, 1999.

[3] J. Galindo, P. Tamayo, "Credit Risk Assessment using Statistical and Machine Learning: Basic Methodology and Risk Modeling Applications", Computational Economics, Springer-Verlag, Vol.15, No.1-2, 2000, pp. 107-143.

[4] Z. Huang, "Credit Rating Analysis with Support Vector Machines and Neural Networks: A Market Comparative Study", Decision Support Systems, Elsevier, No.37, 2004, pp. 543-558.

[5] C. Mar Molinero, P. Apellaniz, C. Serrabo-Cinca, "A Multivariate Analysis of Spanish Bond Ratings", International Journal of Management Science, Elsevier, Vol.24, No.4, pp. 451-462.

[6] W. Hardle, L. Simar, *Applied Multivariate Statistical Analysis*, Springer-Verlag, New York, 2003.

[7] A. K. Jain, M. N. Murty, P. J. Flynn, "Data Clustering: A Review", ACM Computer Surveys, ACM, Vol.31. No.3, 1999, pp. 264-323.

[8] F. Höppner, F. Klawonn, R. Kruse, T. Tunkler, *Fuzzy Cluster Analysis*, John Wiley and Sons, Chichester, 1999.

[9] V. Olej, P. Hájek, "Hierarchical Structure of Fuzzy Inference Systems Design for Municipal Creditworthiness Modelling", WSEAS Transactions on Systems and Control, WSEAS Press, Vol.2, 2007, pp. 162-169.

[10] P. Hájek, *Municipal Creditworthiness Modelling by Computational Intelligence Methods*, Ph.D. Thesis, University of Pardubice, Pardubice, 2006.

[11] T. Kohonen, *Self-Organizing Maps*, Springer-Verlag, New York, 2001.

[12] G. A. Carpenter, S. Grossberg, "Adaptive Resonance Theory: Stable Self-organization of Neural Recognition Codes in Response to Arbitrary Lists of Input Patterns", Proc. of the 8th Annual Conference of the Cognitive Science Society, 1986, pp. 45-62.

[13] B. Stein, S. Meyer zu Eissen, F. Wissbrock, "On Cluster Validity and the Information Need of Users", Proc. of the International Conference on Artificial Intelligence and Applications (AIA 03), Benalmádena, Spain, 2003, pp. 216-221.

[14] G. A. Carpenter, S. Grossberg, "ART2: Self-organization of Stable Category Recognition Codes for Analog Input Patterns, Applied Optics", OSA, Vol.26, No.23, 1987, pp. 4919-4930.

[15] G. Holmes, M. Halland, E. Frank, "Generating Rule Sets from Model Trees", Proc. of the 12th Australian Joint Conference on Artificial Intelligence, Sydney, Australia, 1999, pp. 1-12.

[16] V. Olej, *Modelling of Economic Processes on the basis of Computational Intelligence*, M&V, Hradec Králové, 2003, (in Slovak).