NEURAL NETWORKS AS A SUPPORTING TOOL IN CREDIT GRANTING PROCEDURE

Dorota Witkowska¹, Władysław Kaminski², Krzysztof Kompa³, Iwona Staniec⁴

Abstract
In this paper we discuss the application of artificial neural networks to the classification of the regional bank and leasing company clients. The clients of both institutions are classified taking into consideration creditworthiness, economic and financial situation of investigated firms along with decisions about allowing the credit. In the research, multilayer perceptron and radial basis function networks are used. The networks are trained applying back propagation and genetic algorithms along with Widrow – Hoff and linear vector quantization methods. The accuracy of the networks is evaluated in terms of classification errors.

JEL C45,
MSC 68T10

1. INTRODUCTION
During the transition it has been necessary to restructure the Polish economy. State enterprises have been privatized and many private firms have been founded. The lack of capital is the main problem that has been faced by the enterprises operating in Poland.
Looking for new (in Poland) ways of crediting the investments, the first leasing companies were founded in 1990. In 1997 there were more than one hundred leasing companies and 7% of all investments in Poland was financed through them.

Financial institutions have to check the probability that a transaction will be successfully completed. Although one can never eliminate the risk but attempts can be made to reduce it to the minimum. In order to achieve this goal, the economic analysis of an enterprise (as a potential borrower) is conducted. There are several tools such as the scoring point method, multivariate discriminant analysis and, recently, artificial neural networks (ANN) that are used to support the creditworthiness determination.

Applying the neural networks to firm evaluation is an example of the pattern classification problem. Yoon and Swales (1990) compare ANN to the discriminant analysis with respect to the prediction of stock price performance and find that the neural network is superior to discriminant analysis in its predictions. Rahimian et. al. (1993), Odom and Sharda (1993), Raghupathi et. al. (1993), Wilson and Shrada (1994), Baetge and Krause (1993), and Gately (1999) apply the neural network system to bankruptcy prediction. Rehkugler and Schmidt-von Rhein (1993), and Gately (1999) discuss the problems of application ANN to creditworthiness evaluation.

The aim of this paper is to present the results of our research on artificial neural networks application to evaluate the firms liability.

2. ARTIFICIAL NEURAL NETWORKS

Artificial neural networks are the information processing systems whose structure and function are motivated by the cognitive processes and organizational structure of neuro-biological systems (Hertz et al., 1991). The basic components of the network are highly interconnected processing elements called neurons, which work independently in parallel. Synaptic connections are used to carry messages from one neuron to another.

A single artificial neuron is the basic element of the neural network. It comprises several inputs \((x_1, x_2, \ldots, x_m)\) and one output \(y\) that can be written as

\[
y = f(e)
\]

for

\[
e = w_0 + \sum_{k=1}^{m} x_k w_k = w_0 + w^T x
\]

where \(w = [w_k]\) is the vector of weights assigned for each input \(x_k\), \(x = [x_k]\) is the vector of input variables, \((k=1, 2, \ldots, m)\). Weights are parameters of the activation function which maps any real input into a, usually, bounded range \([0, 1]\) or \([-1, 1]\). The activation (transfer) function may be linear or non-linear. If the activation function is linear then the output (1) may be written as

\[
f(e) = w_0 + w^T x
\]

The non-linear activation functions that are the most commonly used are

- threshold function:
  \[
f(e) = \begin{cases} 1 & \text{for } e \geq 0 \\ 0 & \text{for } e < 0 \end{cases}
\]

- logistic function:
  \[
f(e) = \frac{1}{1 + \exp(-\beta e)}
\]

- hyperbolic tangent:
  \[
f(e) = \tanh(\beta e) = \frac{\exp(\beta e) - \exp(-\beta e)}{\exp(\beta e) + \exp(-\beta e)}
\]

- Gaussian functions:
  \[
f(e) = \exp\left(-\frac{e^2}{2}\right)
\]

where \(\beta\) is a constant.
The computational structure of artificial neural networks has attractive characteristics such as graceful degradation, robust recall with noisy and fragmented data, parallel distributed processing, generalization to patterns outside of the training set, nonlinear modeling, and learning (Tours et al., 1993; Ripley, 1993).

ANNs are usually formed by a cascading group of single layers. There is an input layer, an output layer, and hidden layers. The neurons of different layers are densely interconnected through direct links. At the input layers, the neurons receive the values of input variables and multiply them through the network, layer by layer. The hidden layer neurons are often characterized as feature-detectors. The number of hidden layers and the number of neurons in each hidden layer can be selected arbitrarily. The initial weights of the connections can be chosen randomly. During training procedure the network adjusts its weights to produce a correct output for every input.

There are numerous ANN architecture designs, however they can be divided into three broad classes, on the basis of

- connections among layers of neurons,
- activation functions,
- the technique used to train free parameters (weights) in the network.

Taking into account the connection among layers of neurons we may distinguish feedforward and recurrent networks. Both of them are used in our research. We apply multilayer perceptron (MLP) that is an example of the former and the so called Hamming network that is an example of the latter. In our experiments, we use logistic activation function (5).

Another example of the feedforward networks is a radial basis function (RBF) network. Total transfer input-output function of the network is presented below:

$$y = \sum_{k=1}^{m} w_k \phi \left( \frac{\| x - c_k \|}{\delta_k} \right)$$

where, $w_k$ are the network adjustable weights connecting the network hidden nodes with the network output, $\phi \left( \frac{\| x - c_k \|}{\delta_k} \right)$ is the radially-symmetric transfer functions with centres $c_k \in \mathbb{R}^N$ ($k=1,...,m$), $\delta_k$ is the scaling factor, $\| \cdot \|$ denotes the Euclidean distance.

Note that in the RBF network, the only adaptable weights (i.e. parameters $w_k$ in (8)) are located between the hidden and the output network layers. These weights determine a linear combination of basis functions values and together with chosen basis functions centres $c_k$ decide about the shape of the generated mapping function $f(e)$ (Kaminski and Strumillo, 1997). Schematic diagrams of RBF and MLP are presented in Figures 1 and 2.

Usually three learning classes are distinguished: supervised, reinforcement and unsupervised learning. In our investigation we use the back propagation and genetic algorithms along with Widrow-Hoff rule which are supervised training methods. We also apply linear vector quantization network based on the rule "winner takes all" that is a method of unsupervised learning.

While applying the supervised learning, sample inputs and desired outputs must be given (the entire collection of cases learned is called a "training set"). During the training procedure the computed output is compared to the desired output. If the computed output is incorrect, then the weights are adjusted so as to make the computed output closer to the known output. Using the unsupervised learning the network does not know the correct answers and it is to find out the classification patterns (there is the so-called self-organizing network).

In many applications we deal with the objective (cost, criterion) function

$$Q = \sum_{i=1}^{P} (y_i - y_i^*)^2 = \sum_{i=1}^{P} (f(x_i) - y_i^*)^2$$

(9)
where $y_i$ is the estimated output; $y_i^*$ is the desired output, $P$ is the number of elements (cases) in the training set.

According to the fact that activation function is usually nonlinear, the cost function (9) has many local minima. The backpropagation algorithm, which is the most popular method of supervised learning, usually stops the training procedure in local minimum. In practice, we are interested in localization of the global minimum. The global optimisation methods, that have been developing rapidly in the last years, include genetic algorithms. These algorithms appear to be universal, flexible and efficient.

Genetic algorithms (GA) are based on evolutionary optimisation and belong to population-based methods (Fogel, 1994; Davis, 1991). Their concept is taken from biological evolution in competitive environments, and mathematical formulation of these algorithms relies on the operations which mimic the evolution process. Evolutionary based optimisation methods perform global search in the parameter space (i. e. neural network weights). Their successful applications for training neural networks have been already reported (Bulsari, 1995). In our research, we apply a modification of the genetic algorithm (Kaminski and Tomczak, 1998). It is worth mentioning that while applying RBF networks the global minimum is always obtained.

After the neural network learns up to the error threshold, the weight adaptation mechanism is turned off and the net is tested on known cases it had not seen before. The application of the neural network to unseen cases gives the true error rate (Baets and Venugopal, 1994).

3. STATISTICAL DATA

In our investigation, we used data regarding:
- 75 small firms, which applied for a credit in the regional bank in the period from September 1994 to June 1997. The value of the credit varied from $140 to $1600 of the monthly credit installments, which have been to be paid in the period not longer than 5 years.
- 71 small and medium size firms that applied for transport or production means lease in the leasing company in the period 1997 - 1998.

To consider an enterprise as a potential borrower, banks and leasing companies investigate financial statements along with the information about the firm and the investments that are to be financed. On the basis of these documents, the credit officers evaluate the firm’s economic and financial situation, and decide if the credit is to be allowed or the leasing agreement is to be signed.

The aim of our investigation is to apply artificial neural networks to classify the enterprises according to three questions that should be answered after the firm examination
(a) if the economic and financial situation is: bad, poor, satisfying, good or very good,
(b) if the investigated firm is: uncreditworthy, high risk, average or creditworthy client,
(c) if the credit should be allowed or the leasing object is to be leased.

The credit officers’ opinions are descriptive features. Therefore, to introduce them into ANN experiments, they are transformed into quantitative and integer variables, as following
(a) evaluation of the economic and financial situation of the firm, the output variable $y_1 \in [0; 4]$,
(b) evaluation of the creditworthiness of the firm, the output variable $y_2 \in [0; 3]$,
(c) decision about allowing the credit, the output variable $y_3 \in [0; 1]$.

On the basis of information provided by the firms, that applied for a credit in the bank, seven input variables are constructed.

---

5 Genetic algorithms are modern successors of Monte Carlo search methods, and they belong to the class of stochastic optimization algorithms. GA are usually a compromise between searching the whole set of feasible solutions and local optimization. But with high probability they lead to the global minimum.
$x_1^B$ - gross profit,
$x_2^B$ - net profit,
$x_3^B$ - the average net monthly profit,
$x_4^B$ - monthly repayment of the credit and the interest,
$x_5^B$ - value of maximal monthly interest,
$x_6^B$ - monthly liabilities to other banks and institutions,
$x_7^B = x_3^B - x_4^B - x_5^B - x_6^B$.

All data records are randomly arranged. The set of 75 vectors (describing each enterprise which applied for a credit) is divided into two separate subsets i.e. the training and testing samples. The training set includes 50 vectors $[x_1^B, x_2^B, ..., x_7^B, y_i]; i = 1,2,3$ and the testing set contains 25 vectors $[x_1^B, x_2^B, ..., x_7^B]$. In our investigation, regarding a classification of the leasing company clients, we apply two sets of input variables, which are based on

- original data from the documents obtained from the enterprises, so we use the following input variables
  
  $x_L^1$ - net value of the leasing object,
  $x_L^2, x_L^3$ - value of sales in the past year and in the past reporting period,
  $x_L^4$ - financial liquidity ratio,
  $x_L^5, x_L^6$ - value of profit on sales in the past year and the past reporting period,
  $x_L^7, x_L^8$ - value of current assets at the end of the past year and the past reporting period,
  $x_L^9, x_L^{10}$ - value of long-term loans at the end of the past year and the past reporting period,
  $x_L^{11}, x_L^{12}$ - value of liabilities at the end of the past year and the past reporting period,
  $x_L^{13}$ - amounts receivable at the end of the reporting period,

- ratios calculated by the leasing company experts
  
  $x_R^1$ - net value of the leasing object / sales in the past year,
  $x_R^2, x_R^3$ - monthly payment / average monthly sells in the past year,
  $x_R^4$ - monthly payment / average monthly sells in the past reporting period,
  $x_R^5$ - financial liquidity ratio,
  $x_R^6$ - profit on sales in the past year / sales in the past year,
  $x_R^7$ - incomes or profits in the past reporting period / sales in the past reporting period,
  $x_R^8$ - current assets in the past year / value of loans in the past year,
  $x_R^9$ - current assets in the past reporting period / value of loans in the past reporting period,
  $x_R^{10}$ - current assets in the past year / value of liabilities in the past year,
  $x_R^{11}$ - current assets in the past reporting period / value of liabilities in the past reporting period.

All data records are randomized and divided into the training and testing sets. The former contains 60 vectors including input and output variables $[x_1^L, x_2^L, ..., x_13^L; y_2]$ or $[x_1^R, x_2^R, ..., x_{10}^R; y_2]$ and the latter contains 11 vectors $[x_1^L, x_2^L, ..., x_{13}^L]$ or $[x_1^R, x_2^R, ..., x_{10}^R]$.

Applying the supervised learning, all desired values of variables $y_1, y_2$ and $y_3$ in the training set are integer, while the outputs generated by the neural networks are real ones therefore it is necessary to introduce the threshold function which maps real values into integer ones. We assume that

\[
y = \begin{cases} 
  a, & \text{if } \bar{y}+0.5 \geq a, \\
  (a, -1) & \text{if } \bar{y}+0.5 < a.
\end{cases} \tag{10}
\]

Unfortunately applying our system of data normalization we cannot introduce ratios with high variation and all such ratios are omitted.
where $y$ is the value of variables $y_1$, $y_2$ or $y_3$ after transformation, $\hat{y}$ is the estimated theoretical value of the network outputs, $a_i$ is the value of variable $y_1$, $y_2$ or $y_3$ ($a_i = 1, 2, 3, 4$ for $y_1$, $a_i = 1, 2, 3$ for $y_2$, and $a_i = 1$ for $y_3$).

4. EVALUATION OF THE CLASSIFICATION ACCURACY

Let us assume that there are $N$ objects (firms) $O_i (i = 1, 2, ..., N)$ that are to be classified to $P$ classes $A_p (p = 1, 2, ..., P)$ containing $n_p$ elements $N = \sum_{p=1}^{P} n_p$. Thus the pattern of the object recognition is known and it is possible to evaluate errors of classification by comparing elements which should belong to the groups $A_p$ with elements of classes $\tilde{A}_p$, where $\tilde{A}_p$ denotes the classes that are constructed on the basis on the results of ANN experiments. Let $\tilde{n}_p$ ( $N = \sum_{p=1}^{P} \tilde{n}_p$ ) denote the number of elements belonging to the class $\tilde{A}_p$. Then the general classification error $E$ is defined as follows

$$E = \frac{K}{N} * 100\%$$

(11)

where $K$ is the count of misclassified objects of the class $A_p$, i.e. the number of firms which actually belong to the class $A_p$ but they are recognized as the firms belonging to other classes $\tilde{A}_q$ ($q = 1, 2, ..., N; q \neq p$), $K = \sum_{p=1}^{P} K_p$ is the number of all misclassified firms. A general classification error informs about the percentage of misclassified objects. To analyze how this error is distributed among distinguished classes, the class $A_p$ classification error $E_p$ can be defined as

$$E_p = \frac{K_p}{n_p} * 100\%$$

(12)

$E_p$ informs about the share of objects that are classified to other classes instead of the class $A_p$. The error informing about the share of objects that are recognized as the elements of the $\tilde{A}_p$ class, while they actually belong to other classes, is called the class $A_p$ misclassification error $E^*_p$ and it is defined as follows

$$E^*_p = \frac{K^*_p}{n_p} * 100\%$$

(13)

where $K^*_p$ is the number of misclassified objects from the class $\tilde{A}_p$, i.e. the number of firms that actually belong to the different classes than $A_p$ but they are recognized as the $\tilde{A}_p$ objects (certainly, $K = \sum_{p=1}^{P} K_p = \sum_{p=1}^{P} K^*_p$), $N$ is the number of all firms.

Many problems can be treated as a dichotomous classification since the misclassification to the group of "creditworthy enterprises" (denoted as $\tilde{A}_1$) is more costly than the misclassification to the group of "uncreditworthy firms" (denoted as $\tilde{A}_2$), especially if the classification to the certain group of risk determines the credit granting decision. In such a situation it is useful to evaluate errors of the first and the second kind. These errors are in fact the class $A_p$ classification errors when it is assumed that there are only two classes (i.e. $A_1$ and $A_2$). The percentage error of the first kind can be defined as

$$E_1 = \frac{K_{11}}{n_2} * 100\% = \frac{K^*_{21}}{n_2} * 100\%$$

(14)

and the percentage error of the second kind
\[ E_2 = \frac{K_2}{n_1} \times 100\% = \frac{K_1}{n_1} \times 100\% \]  

(15)

It is worth mentioning that, from the bank’s point of view, the error of the first kind \( E_1 \) is crucial since it is more costly if the bank allows the credit to the enterprise that will not pay it off than if the financial institution rejects the firm which is capable to discharge the credit.

5. NEURAL NETWORKS TOPOLOGY AND TRAINING METHODS

The neural network construction depends on the problem that is to be solved. In our investigation we consider two financial institutions (i.e. the regional bank and the leasing company) that make their decisions on the basis of different data base.

5.1. Neural networks constructed for the regional bank clients

There are three classification problems ((a), (b), and (c)) that are to be solved by credit officers. We construct and train several neural networks. All of them consist of seven input variables: \( x_1^n, x_2^n, \ldots, x_7^n \), one hidden layer, and one (except the LVQ network) output neuron, represented by one of the variables: \( y_1, y_2 \) or \( y_3 \). The number of hidden elements varies regarding the output variable. To evaluate the economic and financial situation along with creditworthiness of the firm, the networks contain three hidden variables and to determine the decision about allowing the credit there are two or four hidden neurons. The activation function of all hidden neurons is logistic.

The neural networks, that are constructed to determine the credit granting decision (i.e. for the output variable \( y_3 \)), are trained applying

- supervised learning methods as: backpropagation, Widrow - Hoff (for the Hamming network) and genetic algorithms, along with
- unsupervised learning, i.e. linear vector quantization (LVQ) network (this network contains two output neurons).

The networks, that are applied to evaluate financial and economic situation of the enterprises (described by the output variable \( y_1 \) or \( y_2 \)), are trained using genetic algorithm (GA).

5.2. Neural networks constructed for the leasing company clients

The leasing company clients are classified applying RBF networks. The neural network constructed for 13 input neurons, representing variables \( x_1^L, x_2^L, \ldots, x_{13}^L \), generates the best results applying 23 radial functions \( \Phi() \) and \( \delta_k = 11 \). The network constructed for ten input variables, representing ratios \( x_1^R, x_2^R, \ldots, x_{10}^R \), contains 25 radial functions \( \Phi() \) and \( \delta_k = 14 \). All basis function centres \( c_k \in \langle 0; 1 \rangle \) are chosen randomly from uniform distribution. In both networks there is only one output neuron \( y_2 \) representing the class (- risk category) of clients distinguished by the leasing company. Clients are classified to the certain class in order to decide about the lease of means of production or transport since the leasing company is not willing to finance the investments if the enterprise is an uncreditworthy or a high risk client. Hence, taking into consideration decisions that are made by the leasing experts, the problem of firm evaluation can be also treated as a dichotomous classification, and the decision errors (11), (12) - (13) are also evaluated.

6. RESULTS OF EXPERIMENTS

The accuracy of the neural network classification is evaluated in terms of errors (11) - (13) under the assumption, that experts’ opinions are correct therefore their classification creates the patterns of
firms. The results of experiments are presented in tables where numbers in parentheses represent the position of the firm in the data set.

Results of the dichotomous classification of enterprises which applied for a credit in the regional bank are presented in Table 1. All networks "learned" how to recognize firms and the training errors equal zero, therefore they are not presented in Table 1. The decision errors always, but the Hamming network, regard to the firms denoted as (74) and (57). It is worth mentioning that both firms obtained the credit regardless the fact that the credit officers evaluate the economic and financial situation of the enterprise (74) as a poor one and the firm is recognized as the high risk client. The firm (57) is evaluated as an uncreditworthy client though the economic and financial situation is satisfactory. So it seems that the decisions generated by ANN are more accurate than the ones made by the bank.

The networks, containing two hidden neurons, that are trained applying backpropagation, Widrow - Hoff and LVQ methods did not recognize the firms denoted as (57), (74) and (54). The best results of classification for the net 7-2-1 are obtained applying the genetic algorithm, while the highest classification errors are obtained using the unsupervised training.

The ANN evaluation of the economic and financial situation (i.e. $y_1$ is the output variable), and the creditworthiness (i.e. $y_2$ is the output variable) of the enterprises is presented in Table 2. It is visible that the errors increase with the increase of the number of distinguished classes. As it can be noticed, for the output variable $y_2$ (with four classes distinguished), the general classification error for the training set equals 14%. Among seven misclassified firms, four are classified as "better" enterprises than they are (according to the credit officers’ opinions) and three others are recognized as "worse" firms. In the testing set ANN is more "optimistic" than the bank experts, and the general classification error is 20%. Note, that all misclassified enterprises are classified to the neighboring classes.

Classification into five groups (i.e. $y_1$ is the output variable) seems to be a difficult task for the neural network since the errors are very high, especially the class $\tilde{A}_p$, misclassification and the $A_p$ classification errors. Among 15 enterprises that are misclassified during the training procedure, eight obtain a "higher position" than they actually (i.e. according to the credit officers’ opinion) have and seven others are placed on the "lower position". In the testing set, ten enterprises are classified as "better" firms (than they are recognized by the credit officers) and three firms are recognized as "worse" firms. But also in this experiment, all misclassified enterprises are classified to the neighboring classes.

The results of the leasing company clients’ classification are presented in Tables 3 and 4. During the training procedure, regarding the classification based on variables $x_1^f, x_2^f, ..., x_5^f$, the neural network misclassified 29 firms (Table 3). Among the misclassified objects, nine enterprises, that actually belong to the class $A_3$, are recognized as the class $\tilde{A}_2$ firms and two enterprises are classified to the class $A_1$ firms. Among six misclassified firms that actually belong to the class $A_1$, five are recognized as the class $\tilde{A}_0$ firms and one enterprise as the class $\tilde{A}_2$ firm. All misclassified enterprises, that should be classified to the class $A_0$, are recognized as the class $\tilde{A}_1$ firms as well as all misclassified firms that actually belong to the class $A_2$. It means that only two enterprises (3.9%) are not classified to the neighboring classes.

Taking into account that the leasing agreement is signed only with the enterprises belonging to the classes $A_2$ or $A_3$, one should notice that, based on our classification, only 5 firms (belonging to the class $A_1$) would be allowed to lease the means of production. In the same time five enterprises, actually belonging to the classes $A_2$ or $A_3$, are classified as the class $\tilde{A}_1$ firms, therefore they would not sign the leasing agreement. Hence, the general decision error equals $E=16.67\%$ and the first kind error $E_1=15.15\%$, while the second kind error $E_2=18.52\%$. 
Analyzing the testing results, it is clear that only two firms are misclassified since both enterprises should be classified (according to the credit officers’ opinion) to the class $A_3$ and one of them is recognized as the class $\bar{A}_1$ firm, and another as the class $\bar{A}_2$ firm. That result gives the general decision error $E=9.09\%$ and the decision error of the second kind $E_2=33.33\%$ since, in the testing set, there is only one company which would not be financed by a leasing company though it belongs to the class $A_3$.

Applying ratios calculated by the credit officers (Table 4), the neural network had similar problems with a proper recognition of enterprises during the training procedure as the previous one. But it is not able to "generalize its knowledge" that results in essential testing errors. The neural network, with input variables defined as ratios $x^R_1, x^R_2, \ldots, x^R_{10}$, misclassified 27 firms. Among them nine firms, that should belong (according to the experts’ opinion) to the class $A_3$, are recognized as the class $\bar{A}_2$ firms. All misclassified firms, actually belonging to the class $A_2$, are recognized as the class $\bar{A}_1$ enterprises. Among seven misclassified firms, that should belong to the class $A_1$, three are recognized as the class $\bar{A}_0$ firms and four enterprises are classified to the class $\bar{A}_2$. Among the enterprises, that should be classified to the class $A_0$, six are recognized as the class $\bar{A}_1$ firms while one as the class $\bar{A}_2$ firm. So only three enterprises (5%) are not classified to the neighboring classes.

Examining the network accuracy in terms of decisions made by the leasing company, one can say that among misclassified firms, five would be allowed to lease the leasing object though they actually belong to the class $A_1$ and five of them will not sign the leasing agreement though they belong to the class $A_2$. So the general decision error equals $E=18.33\%$, the first kind error $E_1=15.15\%$, and the second kind error $E_2=18.52\%$.

Testing results show that six firms are classified as the class $\bar{A}_1$ firms, while three are originally the class $A_0$ firms and one should belong to the class $A_2$. One of the investigated enterprises is recognized as the class $\bar{A}_1$ firm though it is the class $A_3$ firm and two enterprises, that actually belong to the class $A_1$, are recognized as the class $A_3$ firms. Hence one of the misclassified firms would not be allowed to lease the means of production though it belongs to the class $A_3$ and three of them will sign the leasing agreement though they originally are the class $A_1$ firms. That result gives the general decision error $E=36.36\%$, the first kind error $E_1=37.5\%$, and the second kind error $E_2=33.33\%$. Among eight misclassified companies, six (77.8%) are recognized as belonging to the neighboring classes.

7. FINAL REMARKS

Applying the genetic algorithm to classification, a less complicated network is required than using the backpropagation algorithm or Widrof - Hoff rule. Since the population in the genetic algorithm (which denotes a possible penetration of the set of admissible solution) is bigger than repeatability of training vectors in the backpropagation algorithm. One can also notice that the supervised training gives better results than the unsupervised learning. Although, in many cases, we do not know the patterns of classification or they are not reliable. In such a situation the application of the unsupervised training is a unique possible solution.

To train the network many data records are required since the degrees of freedom influence the ability of the "knowledge generalization". In our experiments the training sets are pretty small and they influence the results of classification. Radial basis function networks require a small number of examples in the training set than MPL. The RBF system works as a fuzzy logic network that results from properties of the Gaussian function which possesses the so-called receptive field.

Our research shows that the general classification errors arise according to the increase of the number of the desired classes. Since the higher number of classes is distinguished, the difference among firms belonging to the different groups is less visible. The classification errors are always
higher for the testing set than for the training sample though the majority of misclassified enterprises are classified to the neighboring classes.

One may also add that the desired values, which are introduced to evaluate the accuracy of the ANN classification, represent the opinions expressed by the credit officers and we assume that these opinions are correct which does not have to be true. It is also obvious that, if we compare the results of any method of the firm recognition to the credit officers’ opinion instead of the actual situation (which can be described by a future cooperation between the client and the financial institutions only), then the classification results will be always worse than the classification made by the experts. Especially that the credit officers take into account not only information provided by the enterprise but also the position of the economic branch, represented by the analyzed firm, along with the history of cooperation between the investigated firm and the financial institution. That kind of information is not available, therefore we cannot introduce it into our experiments.

The results of experiments show that, regardless the classification errors, the artificial neural networks can be a useful tool for the financial institutions to support the decision making procedure. The classification based on original (i.e. provided by the firms) information is more adequate than using ratios evaluated by the credit officers. It results from the fact that every index "flattens" information that is used by data processing system though the ratio alone may be significant for the quality evaluation.

REFERENCES

---

7 Applying ANN to a dichotomous classification gives much better results than using the linear discriminant analysis (Witkowska, Staniec 1999), see also (Witkowska 1998).
<table>
<thead>
<tr>
<th>Network structure</th>
<th>Backpropagation algorithm</th>
<th>Genetic algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>class $A_0$, $y_3=0$</td>
<td>class $A_1$, $y_3=1$</td>
<td>$E_1$</td>
</tr>
<tr>
<td>7-2-1</td>
<td>7-4-1</td>
<td>7-2-1</td>
</tr>
<tr>
<td>$n_p$</td>
<td>4</td>
<td>21</td>
</tr>
<tr>
<td>$E_p$</td>
<td>25.00 (54)</td>
<td>0.00</td>
</tr>
<tr>
<td>$E_p^*$</td>
<td>50.00 (57, 74)</td>
<td>50.00 (57, 74)</td>
</tr>
</tbody>
</table>

Hamming network

<table>
<thead>
<tr>
<th>Network structure</th>
<th>class $A_0$, $y_3=0$</th>
<th>class $A_1$, $y_3=1$</th>
<th>$E_1$</th>
<th>$E_2$</th>
<th>$E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-2-1</td>
<td>7-4-1</td>
<td>7-2-1</td>
<td>7-4-1</td>
<td>7-2-1</td>
<td>7-4-1</td>
</tr>
<tr>
<td>$E_p$</td>
<td>25.00 (54)</td>
<td>0.00</td>
<td>9.52 (57, 75)</td>
<td>9.52 (57, 59)</td>
<td>25.00</td>
</tr>
<tr>
<td>$E_p^*$</td>
<td>50.00 (57, 74)</td>
<td>50.00 (57, 59)</td>
<td>4.76 (54)</td>
<td>0.00</td>
<td>50.00 (57, 59)</td>
</tr>
</tbody>
</table>

LVQ

<table>
<thead>
<tr>
<th>Network structure</th>
<th>class $A_0$, $y_3=0$</th>
<th>class $A_1$, $y_3=1$</th>
<th>$E_1$</th>
<th>$E_2$</th>
<th>$E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-2-2</td>
<td>7-4-2</td>
<td>7-2-2</td>
<td>7-4-2</td>
<td>7-2-2</td>
<td>7-4-2</td>
</tr>
<tr>
<td>$E_p$</td>
<td>25.00 (54)</td>
<td>25.00 (54)</td>
<td>14.29 (52, 72, 74)</td>
<td>9.52 (57, 74)</td>
<td>25.00</td>
</tr>
<tr>
<td>$E_p^*$</td>
<td>75.00 (52, 72, 74)</td>
<td>50.00 (57, 74)</td>
<td>4.76 (54)</td>
<td>4.76 (54)</td>
<td></td>
</tr>
</tbody>
</table>

Note: In Table percentage errors (11) - (15) are presented. Numbers in brackets denote the position of misclassified firms in the data set. Source: Authors’ calculations.
<table>
<thead>
<tr>
<th>Creditworthiness evaluation</th>
<th>Economic and financial situation evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>class $A_0$</td>
</tr>
<tr>
<td>Training set</td>
<td>$y_1=0$</td>
</tr>
<tr>
<td>$n_p$</td>
<td>13</td>
</tr>
<tr>
<td>$E_p$</td>
<td>7.69</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>$E_p^*$</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(1, 34)</td>
</tr>
<tr>
<td>Testing set</td>
<td>$y_1=0$</td>
</tr>
<tr>
<td>$n_p$</td>
<td>4</td>
</tr>
<tr>
<td>$E_p$</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(53, 72)</td>
</tr>
<tr>
<td>$E_p^*$</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(53, 72)</td>
</tr>
</tbody>
</table>

Note: In Table percentage errors (11) - (15) are presented. Numbers in brackets denote the position of misclassified firms in the data set.

Source: Authors’ calculations.
<table>
<thead>
<tr>
<th>Category of client</th>
<th>class $A_0$</th>
<th>class $A_1$</th>
<th>class $A_2$</th>
<th>class $A_3$</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_p$</td>
<td>15</td>
<td>18</td>
<td>8</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>$K_p$</td>
<td>9</td>
<td>6</td>
<td>3</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>$E_p$</td>
<td>60.00</td>
<td>33.34</td>
<td>37.50</td>
<td>57.90</td>
<td>48.33</td>
</tr>
<tr>
<td>Decision errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dichotomous</td>
<td>18.52</td>
<td>15.15</td>
<td></td>
<td></td>
<td>18.33</td>
</tr>
<tr>
<td>classifications</td>
<td>(2,4,12,23, 24)</td>
<td>(15,17,21,35,42)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testing set</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_p$</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>$K_p$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>$E_p$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>18.18</td>
</tr>
<tr>
<td>Decision errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dichotomous</td>
<td>33.33</td>
<td>0.00</td>
<td></td>
<td></td>
<td>9.09</td>
</tr>
<tr>
<td>classifications</td>
<td>(68)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: In Table percentage errors (11) - (15) are presented. Numbers in brackets denote the position of misclassified firms in the data set. Source: Authors` calculations.
TABLE 4: Clients evaluation applying ratios $x_1^R, x_2^R, ..., x_{10}^R$

<table>
<thead>
<tr>
<th>Category of client</th>
<th>class $A_0$</th>
<th>class $A_1$</th>
<th>class $A_2$</th>
<th>class $A_3$</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>y_2=0</td>
<td>y_2=1</td>
<td>y_2=2</td>
<td>y_2=3</td>
<td></td>
</tr>
<tr>
<td>$n_p$</td>
<td>15</td>
<td>18</td>
<td>8</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>$K_p$</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>9</td>
<td>45.00</td>
</tr>
<tr>
<td>$E_p$</td>
<td>46.67</td>
<td>38.89</td>
<td>50.00</td>
<td>47.37</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7,13,20,40,43,54,59)</td>
<td>(6,29,32,35,38,42,56)</td>
<td>(4,12,49,50)</td>
<td>(1,2,9,10,11,27,33,47,52)</td>
<td></td>
</tr>
<tr>
<td>Decision errors</td>
<td>dichotomous classifications</td>
<td>18.52</td>
<td>15.15</td>
<td>16.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4,12,14,49,50)</td>
<td>(32,35,42,56,59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testing set</td>
<td>y_2=0</td>
<td>y_2=1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_p$</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>$K_p$</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>72.73</td>
</tr>
<tr>
<td>$E_p$</td>
<td>0.00</td>
<td>75.00</td>
<td>100</td>
<td>50.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(63,66,67,69,70,71)</td>
<td>(65)</td>
<td>(62)</td>
<td></td>
</tr>
<tr>
<td>Decision errors</td>
<td>dichotomous classifications</td>
<td>33.33</td>
<td>37.50</td>
<td>36.36</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(62,65)</td>
<td>(66,67,69)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: In Table percentage errors (11) - (15) are presented. Numbers in brackets denote the position of misclassified firms in the data set. Source: Authors' calculations.
Fig. 1. Multilayer perceptron neural network

\[ \Phi_1(x) \Phi_2(x) \Phi_M(x) \]

\[ x_1 \ldots x_2 x_3 x_N \]

Fig. 2. Radial basis function