

NEURAL NETWORKS USAGE IN THE EVALUATION OF EUROPEAN UNION COFINANCED PROJECTS

Tadeusz A. GRZESZCZYK

Faculty of Management
Warsaw University of Technology, 02-524 Warszawa, Poland
email: t.grzeszczyk@wz.pw.edu.pl

Abstract: Research concerns the implementation of modern computing technologies in the evaluation of projects cofinanced by the European Union. Crucial element of this research is the enrichment of currently used evaluation methods with modern mechanisms basing on artificial intelligence. The article deals with the possibility analysis of neural networks usage in such applications.

Key words: evaluation of European projects, modern computing technologies, neural networks.

1 Introduction

Scientific research on modern, broad and crucial issues useful in expansion of the management theory and enrichment of management practice were firstly realized in the Institute for Organization of Production Systems. Currently the research is continued in the Faculty of Management at the Warsaw University of Technology. The author of the article, basing on past experience and research in the process of formulation of new paradigms of management theory, points out the issue (not very popular nowadays) of management and assessment of projects cofinanced by the European Union (EU project evaluation).

Project management (also known as enterprise management) is currently a subject to intensive research worldwide (especially in Anglo-Saxon countries). In Poland this discipline is relatively less researched. European Project is a new concept, which stands for a management of a project of particular type. It fulfills all systematic disciplines determined by the theory of management. Additionally it fulfills the requirements determined by the European Commission. These disciplines are connected with the possibility to gain cofinancing of the project by the European Union. The author has a considerable experience with methods and systems of evaluation of European problems.

First experience with projects cofinanced by the EU was gathered in Poland during 2004-2006. Apart from problems connected with the lack of experts who could evaluate applications, there were no proper IT tools present that would aid the management of the union funds. Difficulties with implementation of IT System for Monitoring and Finance Control of Structural Funds and Cohesion Fund were also one of major issues.

There are no effective tools which would support the decision making for the implementation and mediating experts. This is the reason for relatively low effectiveness of the cofinancing projects application evaluation process. In current financial perspective, considerably larger amount of financial resources was dedicated to Poland that in the previous programming period. Lack of proper tools supporting the project evaluation process can lead to the loss of union funds.

Therefore the issue of European project evaluation is extremely important and still up-to-date. Moreover it was not properly researched even in countries with considerable traditions and experience in EU funds management. Research difficulties are connected with the fact that methodology and terminology in the scope of European project are far from completion and coherency. In Polish theory of management there was no research of possibility of using computing technologies (including neural networks) for the evaluation of European projects.

The genesis of the research originates from the necessity and possibility to undertake interdisciplinary methodological research on European projects evaluation. Author provides proof for the usefulness of new computing technologies that can be supplementary for currently used project evaluation methods. These technologies can be based on the system with neural networks.

2 European project evaluation process

Evaluation (also known as assessment) has many definitions. Therefore the evaluation of project realization can be determined according to different points of view of different criteria and applications etc. One of the classic definitions claims that evaluation is "activity

acknowledged as efficient or inefficient in the universal of synthetic meaning” [11]. According to the author of the definitions “in general meaning efficiency is every factor of a job well done, which means that efficiency is effectiveness, economy etc.”. On the other hand the synthetic sense is expressed in the generalization and research of many virtues of a job well done.

With the participation of Poland in the European Union all research dealing with the European project evaluation gain meaning in successive years. However, research advancement level should not be compared with other countries, which have been members of the EU for many years.

During 1991-1999 the MEANS (Methods of Evaluating Structural Policies) program was realized under the direction of Committee of Independent Experts and supported by European Commission. Teams of theorists and practitioners elaborated evaluation systems that, with minor modifications, are still in use. Six volume paper edition and updated Internet version are the main source of information about methodological guidelines of European Commission in the scope of evaluation of European and other projects [6]. The author (to certain extent) includes these guidelines in his research, focusing on proposition of new methods for evaluation that are supplementary for currently used methods.

The following definition of evaluation is most common in the publications of European Commission. Evaluation of a policy, program or project is a judgment on its value including previously taken assumptions and criteria based on gathered and analyzed information [9].

Evaluation of EU cofinanced projects is a process consisting of few distinguishable stages [6]:

- structuring of the project evaluation process,
- data gathering,
- data analysis,
- summary of the research and formulation of project evaluation,
- editing of the evaluation process report,
- use of evaluation process results.

The results of performed process structuring are certain assumptions and aims of the evaluation process of a given project. The following stage of data gathering assures specification of the subjective and objective scope of the research as well as gathering data about evaluated project. Aim of next evaluation process stage is the assessment analysis of data stored during data gathering stage. Results of this stage should allow the realization of one of the final evaluation stages – evalu-

ation formulation. In other words: giving judgment based on gathered analysis results. Number of methods can be associated with every stage:

- structuring of the project evaluation process method,
- data gathering method,
- data analysis method,
- formulation of project evaluation supporting method.

These methods were described and characterized e.g. in monograph [5]. The article focuses on the presentation of issues concerning the use of neural networks in the data analysis stage of the European projects evaluation process.

Literature dealing with evaluation of European projects (e.g. [10]) lists, among others, the following methods:

- general evaluation method:
 - GIS (Geographical Information Systems),
 - Shift-Share,
 - Macro-economic model,
- detailed (deepened) evaluation method:
 - Comparison groups,
 - Regression analysis.

The following section of the article focuses on the possibility of using neural networks in regression analysis.

3 Regression analysis in the process of European projects evaluation

Primary and secondary data gained in the data gathering phase can be subjected to different statistical analysis, also used many other data analysis methods. There are many software tools supporting the analysis process (Statistica from Statsoft, www.statsoft.pl).

Also the commonly known Excel spreadsheet (among other functions) allows using different statistical methods e.g. linear and non-linear regression models. Possible uses of this popular program for statistical calculations and forecasting are described in many publications.

Descriptive statistics can be used to perform the analysis of gathered data (connected with the realization of the project). It is used to show relations between variables. Especially the regression equations are used to research and describe the relations between variables. Regression function is an analytical way of assigning of the explained variable value (dependent)

to particular values of explanatory variables (independent).

When solving the regression problem one should aim at determining a formal description method (regression model) of real relations present due to the realization of the project. Created regression model should represent those real relations. Regression equation is a basic element of the created model, which includes:

- input variables (independent, explanatory) – of qualitative or quantitative character,
- output variables (dependent, explained) – usually of quantitative character (in most cases there is one quantitative variable).

Root-cause analysis is another one of the significant concepts. It is used to study the relations between the cause and effect. In case of using such analysis to evaluate the EU projects, it is necessary to assess:

- cause in the form of implementation of a particular project,
- real effects, which could have appeared as a result of the implementation.

Root-cause analysis can be of two kinds:

- inductive – research of real causes of particular effects (these causes can result from the realization of the project or be totally independent),
- deductive – qualitative and quantitative validation of the hypothesis (based on gathered empirical data), which assumes direct influence of project implementation on the appearance of positive effects.

Regression analysis is one of the deductive methods to research the causes. It is used to answer the following question: did the realization of European project contribute to positive effects in the region? Two kinds of factors are explanatory variables in this case:

- connected with the realized European project,
- other factors, not connected directly with the implemented project.

Factors mentioned above can be described with variables: nominal (step) and continuous variable. Relations between the variables can be linear or non-linear. Regression analysis is basing on a logical model of an event and many empirical observations.

The concept of regression is connected with econometric model, which can e.g. be used to perform the summary evaluation of structural funds and cohesion fund influence on Polish economy [12]. Econometric models can be used to evaluate the effects of a complex project in minor scale e.g. in certain region.

Econometric models have three of the following forms:

- Time series models

Seldom used with relation to the project effects evaluation due to the considerable amount of data necessary to perform such analysis. In practice they are currently possible to be used e.g. in case of pre-accession funds (longer period of functioning in Poland than the structural funds) where variables stand for quarterly funds or monthly observations. There are many, relatively easy to operate, software programs suitable for analysis, smoothing, extrapolation and forecasting of time scales. In simple cases one can use a spreadsheet (e.g. Excel) or some other software available in the market e.g. [7].

- Multi-equation econometric models

Used to perform the regional, national and international economical systems functioning simulations. They can be used as an evaluation tool for program (many projects) influence on social and economical situation of a region, country etc. Such complex models will not be presented in this article due to its seldom use in the EU cofinanced project evaluation.

- Single-equation regression model (single-equation econometrics model).

They are used in quantitative analysis of the relations between the Y dependent variable and the X_1, X_2, \dots, X_n independent variables.

Macro-economic models are one of the most complex evaluation methods, because they need to take into account a great number of factors, which need to be identified and relations between these factors have to be established. Estimation of effects resulting from the cofinancing e.g. with the use of structural funds can be achieved through a comparison analysis of two models. One of the models describes region or economy of the whole country before the cofinancing, second one relates to the case when researched area was replenished with European funds. Usually as a result of financing particular investments is realized, which influence the situation in EU member countries and their regions.

Macro-economic models will not be described because of set object of research. Such models are usually used e.g. for the evaluation of structural funds effects or the operational programs on national or regional level. Seldom are they used in the analysis of a single project, because such econometrical models are a very complicated analysis tool. Elaboration of such models is much more costly than the other analysis methods. On the other hand use of ready econometrical models is con-

venient, quick, precise and usually guarantees solid data analysis results. However, their usefulness in most projects is relatively low.

In case of ex-ante analysis, during the construction of the model, experience, gained in the creation of single-dimensional econometrical models designated for ex-post evaluation for similar projects, can be very useful. They allow a simpler and faster finding of statistically significant and non-correlated independent variables. Preparation of a model for ex-post evaluation is, in general, easier. Estimation of a preliminary model can be relatively easily performed on the basis of gathered data, which relates to the realization period of a researched project. Regression analysis (especially the research of statistical significance of model elements and partial regression factor) allows determining the degree of influence of project's realization on the explained variable.

It is also possible to perform a relatively precise evaluation of the degree of the influence. In this case the regression analysis is much more precise method than other, previously mentioned, methods (displacement analysis and comparison groups).

The ex-post evaluation with the use of regression analysis is usually performed few years after the completion of the project, due to the significance of long-term effects of the project. Resources used in the realization of the project, expected and unexpected effects and the general efficiency of the project is analyzed. Previously created econometric regression model can be used in the analysis (see Fig. 1).

Typical problem that is solved with the regression equations is the analysis leading to determination of the percentage of positive results, reached with the evaluated project (see Fig. 2). Such analysis (different than in other data analysis methods) enables to reach relatively precise results. For example Shift-Share displacement analysis allows only gaining an approximate orientation in the level of implemented project's influence on the situation in the region. Let alone the use of regression equations allows a proper verification of the estimated project results.

Next chapter of the article describes the rules of using neural networks for the construction of regression models that are useful in the data analysis necessary in project evaluation.

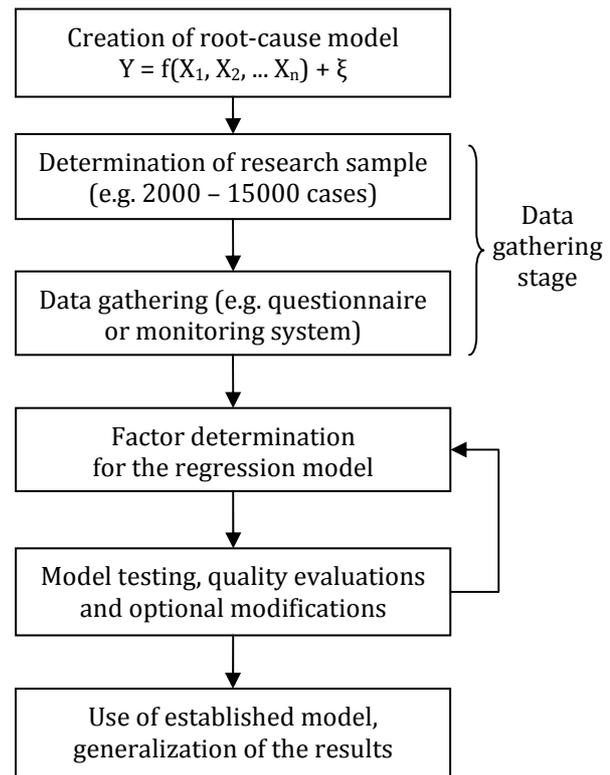


Figure 1. Creation and application of econometric regression model (source: self study on the basis of [10])

4 Use of neural networks for the construction of regression models

Artificial intelligence instruments can be of great value in the project data analysis process [2, 4]. Artificial intelligence is a new branch of IT science, which deals with both the search of new possibilities of computer usage and new methods of programming and new methods of computer problem solving [14].

In the classical set theory there are only two possible variants: given element can be either the part of the set or not. However, in many sciences and techniques such approach is not sufficient. It does not give a full description of a researched event or phenomenon. Many medical, economical, political and other concepts (connected with project realization) have the, so called, non-acute character. They are described with common language, which usually does not allow precise definition of corresponding numeric values.

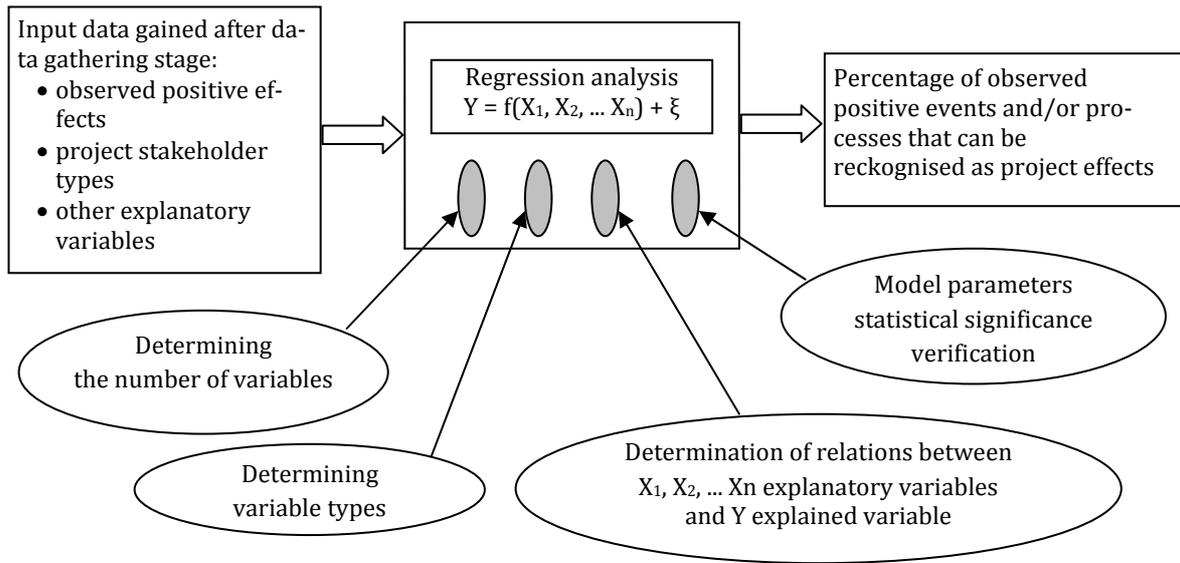


Figure 2. Example of regression model application
(source: self study on the basis of [10])

In such cases it is possible to determine the boundary area the scope that can, but does not have to, include elements significant for the realization of project evaluation process. Finding proper tools to analyze non-acute concepts allows the elaboration of decision making algorithms in uncertain conditions.

Concluding in uncertain conditions takes place when gathered data concerning a problem, which needs to be solved, is incomplete, e.g. processed data is not completely defined or is measured with set error margin. Artificial intelligence (or computational intelligence) methods are specially designed for such purposes.

One of the subclasses of artificial intelligence is the artificial neural networks. Their functioning is based on neural network learning process or the supervised learning process, which results in construction of event models, with the use of random algorithms.

Neural networks are the mathematical models, created in biological sciences, used for the projection of living organisms' nervous systems functioning. Research in this area of science has a long history. Firstly, it was used to solve problems from the area of technical sciences. Gradually neural networks started to be used in management sciences (e.g. to forecast social and economical events [3, 8]).

Neural networks, as a modeling instrument for complex social and economical processes, can support the decision making processes in the evaluation of projects. It enables the generalization of gathered data and approximation of any relationships. Neural networks are showing the knowledge in the process of learning through numerous trials. They allow to automatically creating the model, which projects the complex relations between the input variables (independent) and the output variable – dependent (see Fig. 3). Creation of such model is based (generally) on selection of the neural network, determination of its structure as well as selection of the network parameters in the learning process by the evaluator.

Neural networks have many different uses, one of which is the data analysis. In further deliberations the topic of basic rules for neural networks usage in research of economical events useful in evaluation of European projects, will be presented. Neural networks are especially useful in case of considerable diversity and low precision level of input data. It is not necessary to establish a primary model selection. Network itself chooses the model what is the result of the nature of neural networks. Neural network tolerates large, unsorted data collections. It has a higher number of freedom degrees than the classical static models.

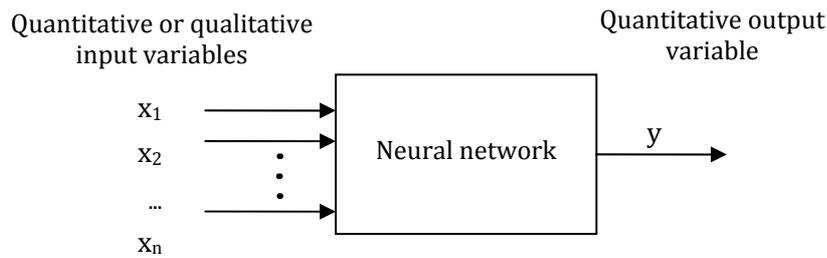


Figure 3. Use of neural regression model for representation of relations between input (explanatory) variables and the output (explained) variable
(source: self study)

Usefulness of neural models is determined by their significant features. Using the relations that are present in neural networks does not require formulation of assumptions, which are difficult to validate. They characterize with the ability to approximate any non-linear relations. They enable the generalization of learning results of training data into new data. Creation of neural model is based on the analysis of available historical data. This way the main relations of the researched event are estimated in the descriptive model. Such models are most useful in case when the evaluator does not know the rules characterizing the creation of analyzed relations. They are especially useful in description of changing, complex social and economical events.

Use of neural networks requires a proper preparation of considerable amount of historical data, according to the character of variables and the used type of network. It is connected with relatively large financial resources necessary for the determination of the neural network. It is usually assumed that the number of learning samples should be 10 to 20 times greater than the number of weights in the network [1]. There is no empirical evidence on a direct and unambiguous relationship between the number of samples from the learning set and the precision of the results. However, in extreme cases of little data (similar to the number of weights), network is not able to generate correct results.

Mathematical model is the projection of a fragment of social and economical events characteristics and their linear and non-linear relations. It modifies its parameters according to the information coming from the environment. It consists of many interrelated processing elements – the neurons. Each of the neurons can be a multivariable function, which computes one output value, based on many input values. Input network in-

formation is introduced into the entrances of successive neurons. This process lasts until the signal processed by the network reaches the output layer or when the process finishing criterion is met.

The following networks types are the most common:

- unidirectional, when neurons are placed in layers – one can distinguish the input, tacit and output layer through which the signal is transferred (starting through input, next through tacit and finishing at output layer),
- recurrent, when neurons placed in layers, as mentioned above, transfer signals two ways, due to feedbacks,
- cellular, when neurons are not grouped into layers and transfer information only to neighboring neurons.

Deliberations, in this point of the article, are limited to the methods connected with the possibility the primary type of neural networks in the data analysis – unidirectional networks. Research using four types of neural networks will be presented, in the following part of the article.

Effective functioning of neural networks is dependent on the proper value of its parameters. Successive determination of these parameters, with the method of successive approximation, is the learning process of the neural network. This process is based in the use of gathered information, which is the learning set. One can distinguish the basic iterative stages of network learning process:

- stimulation of the network with signals coming from the environment,
- stimulation effect, expressed through the changing parameters of the network,
- transformation of network relations into information, coming from the environment, which results in network parameters' value changes.

Effectiveness of neural networks in the data analysis process is expressed through the possibility of description and analysis of any relations and their generalization. Neural network can provide an answer, based on input information processing, to the formulated problem, due to this fact. Therefore, regularities originating from a particular reality – the environment of the researched project – are in the disposal. The functioning of the model is based on an assumption that gathered information (research sample) is the typical representatives of the whole population of data. Desirable feature of the neural network functioning model is the ability of proper formulation of answers for input values, which are not a part of current data collection. Usefulness of unidirectional neural networks in the process of European project evaluation results from a fact that social and economical events are of non-linear character. These networks are able to approximate any non-linear relations and to their generalization. They have an adaptive character and can support the description of relations that change with time. That is why, in the existence of new information, process of further learning of the network takes place. Sometimes it can have the character of a small correction, which includes the transformations of the real system. Especially the unidirectional neural networks can be useful to evaluate the effects of the project.

For example, they can be used to solve a regressive issue, which relies upon the determination of the percentage of positive effects, which can be included in the results of the evaluated project effects. It is usually solved with the use of classical regression equations.

5 Neural model's research

The aim of the research is to determine the possibility to use neural networks in the process of European project evaluation. Neural models can be a significant supplement for other data analysis methods hitherto used in project evaluation process.

Any neural network, which accepts the real value vectors as input and creates signals with real values at output, can be used in the process of regressive data analysis, after the data gathering stage.

The research includes the following network types:

- linear – equivalent of the linear regression function,
- unidirectional three-layer – perceptrons MLP (Multilayer Perceptron),

- RBF (Radial Basis Function),
- GRNN (Generalized Regression Neural Networks).

The simplest neural network has only one output and one or more inputs. Regardless of the type of network, the number of input and output neurons is the same for given conditions. Usually the output layer consists of one neuron (which corresponds to the explained variable). Number of neurons in the input layer depends on the set number of explanatory variables (input). Doubts connected with the selection of the architecture apply only to tacit layers of the network (e.g. the number of layers and neurons present in these layers).

It is not necessary to determine the network architecture for linear networks, because there is no tacit knowledge in them. RBF and GRNN networks have strictly determined number of layers. RBF has always three layers and GRNN respectively four layers. Perceptrons can practically be built of any number of layers. Usually they consist of three layers.

Regressive model built with the use of selected network type, which has a particular architecture, should be subjected to a learning process. Regressive statistics can be used to evaluate taught network models. Only the taught model with positive evaluation can be later used in the data analysis for the evaluation of European projects.

Empirical data collection (result of data gathering stage) used for data analysis should be divided (e.g. through drawing of lots) into three of the following sub-sets:

- Learning

Statistical measures determined for the subset allow evaluating the ability of the set to approximate. Network functioning indicators, for the learning subset, determine the ability of the network to properly recall of input data and suitable output variable, which are both presented during the learning process. However, too high ability to approximate is not negative. It leads to, so called, “over-learning” of the network. Over-learned network does not generalize knowledge, but instead tries to adjust everything to prior data. The essence of the problem is the proper preparation of the regression model, which will aid the data analysis in the evaluation process for different projects (not only the projects presented during the learning stage). The negative over-learning of the network (too high approximation ability) can be reduced with the determination of statistical measures for the next validation subset.

- Validating

Statistical indicators counted for this subset allow monitoring the learning process of the network. They support the search of proper moment to break the learning process. Increase of error number of the validating subset usually means the decrease of the ability of the network to generalize (the probability of over-learning increases).

- Testing

Statistical measures for this subset assure the finishing research which validates the practical usefulness for given neural model (or the lack of it). It is tested in order to check the reactions of the taught and validated network reacts on input data that is different than the one presented in the beginning of the learning process.

Therefore, the comparison of statistical measures for particular sets allows evaluating the quality of elaborated neural regression model. Network is taught with the data from the learning set. This learning process is broken due to the simultaneous process of network validation. Evaluation of network's ability to generalize knowledge, with the use of testing data, is performed in the end of the learning process.

Measures that determine the quality of neural regression model are determined for three subsets according to presented rules. For example, in Statistica, software designated for statistical analysis with the use neural networks, the following parameters are designated to evaluate the taught models (regression statistics) [13]:

- Data Mean – determined on the basis of real values of dependent variable (output, explained) – used in evaluation of European project,
- Standard Deviation of Data - Data S.D., calculated for real values of dependent variable,
- Error Mean – average value of the result of subtraction of two values of explained variable: real and calculated,
- Error S.D. – standard deviation of errors – for the explained variable,
- Abs. E. Mean - average absolute error – arithmetic mean determined for the absolute value of subtractions of: real values of explained variable and its values at the output of the model,
- S.D. Ratio: data standard deviation ratio – standard deviation of errors and standard deviation of real data (both parameters determined above),
- Correlation: the standard Pearson-R correlation – determined between real and output values of the model (dependent variable).

The highest significance in the evaluation of neural models' quality has: the quotient of standard deviations and real and output values correlation. The first one (non-negative) parameter for created models should have the values of 0,1 ... 0,2. The smaller its value is the better. The number of deviations close to zero determined the good quality of the model. If it is higher than one (or close to it), designed model can be rejected (standard deviation of the errors is comparable to the standard deviation of real data).

It is difficult to estimate in case of deviation's quotient in the range of: 0,3 ... 0,7. Finally the quality of the model depends on whether the results reached with it are acceptable in particular case. Correlation of real and output values in the model takes values from the 0 to 1 range. It is better when the value is close to one (the closer the better).

Example of a regression problem, which can be attempted to solve with the use of neural regression model, is the relations study of input variables characterizing positive effects (observed during data gathering) after the end of realization of Sectoral Operational Programme: Improvement of the competitiveness of enterprises.

Essence of the presented example is the demonstration of possibilities of neural regression model. Previously mentioned Statistica Neural Networks software was selected for the usefulness analysis of data in European project evaluation process.

Selected project deals with capital investments in enterprise's fixed assets, equipment, new technologies and training. Relatively technologically similar enterprises (or groups of enterprises) were subjected to the research, located in a distinguished local or regional market. Analyzed enterprise collective consisted of 100 cases. Explained variable was determined as: median of number of new workplaces in the first year after the finishing of the project.

Dependent variable was set as the median of number of new workplaces, because this parameter is actually an average (median is usually called the middle value or second fractal).

Median is the most resistant to outstanding elements, Contrary to other types of mean (e.g. arithmetic, geometric, winsdor etc.). In most cases this feature, which is a considerable immunity to disturbances, is said to be an advantage. However, in some cases, the lack of influence on its value can cause major errors, even for

large and extreme observations. In analyzed case study, the lack of influence on the final result of extraordinary (outstanding) data can be treated as an advantage.

In the analyzed case, six exemplary independent variables, connected with the realization of the project, were distinguished:

- number of employees, who have finished trainings (cofinanced by the project) that increase their professional competences,
- number of new introduced technologies,
- number of employees with secondary and higher education,
- percentage of project participants who evaluate the project positively,
- increase of enterprise income in given sector (percentage),
- number of new products launched into market.

Works on neural regression model were started with the research of the usefulness of linear network. In general, if it is possible to describe the relations with the six, above mentioned, independent variables and one dependent variable with the simplest linear regression function, it is not necessary to search for more complex models. However the lack of possibility to gain proper description of these relations between variables with linear network indicates that they have a non-linear character. 6-1 architecture linear network experiment results (six neurons in the input layer and one in the output layer) can be treated as a reference point for further research with other types of networks useful to solving regression issues.

Statistica Neural Network software can determine regression statistics, which allow evaluating the linear regression model describing relations between six explanatory variables and one explained variable, when this tool is used to model neural networks.

One can deduct, basing on the reached statistical measures values results, for example that the relation is of non-linear character. What especially indicates that fact is e.g. high value of standard deviation quotient (higher than 0,5 – 0,6) and relatively low correlation value (lower than 0,7 – 0,8). In such case it is necessary to start the creation of non-linear models, which base e.g. on RBF and GRNN perceptrons.

Relatively low value of standard deviations quotient (of 0,2 - 0,3) and proper correlation value (higher than

0,8) also should not discourage to further experiments for the linear model. One should check the parameter values for non-linear networks.

The following research programme, connected with the construction of neural models, was realized (see Fig. 4):

- network selection (one of the three mentioned above),
- determination of model's architecture (number of tacit layers and neurons in these layers),
- performing of the learning, validation and network testing processes,
- evaluation of reached regression model and decision making about the finishing of the search or another beginning from the network selection stage and its architecture,
- use of built model for the regression analysis.

Number of neurons in the input layers of all networks that can be used in this analysis equals six. It is equal to the number of explanatory variables. Output layer consists of one neuron (corresponding to the explained variable). Issues connected with the network structure selections concern only the tacit layers of the network. RBF has always three layers and GRNN respectively four layers. Perceptrons are usually built of three layers.

Network type and structure can be selected intuitively after gaining of some experience in this matter (using the experience from previously realized projects). However this method can lead to creation of models with insufficient number of layers and neurons, with unsatisfactory approximation abilities. On the other hand using networks with too high number of neurons can cause the network to have limited ability to generalize the knowledge gathered in the learning process.

It is possible to use the automatic tools for automatic network parameters searching in order to avoid the listed problems with the network type and structure selection. For example, the Statistica Neural Networks software enables the option of automated network set browsing with different types and structures (so called Intelligent Problem Solver). This option is especially useful in the construction of the regression model later used in data analysis for project evaluation processes.

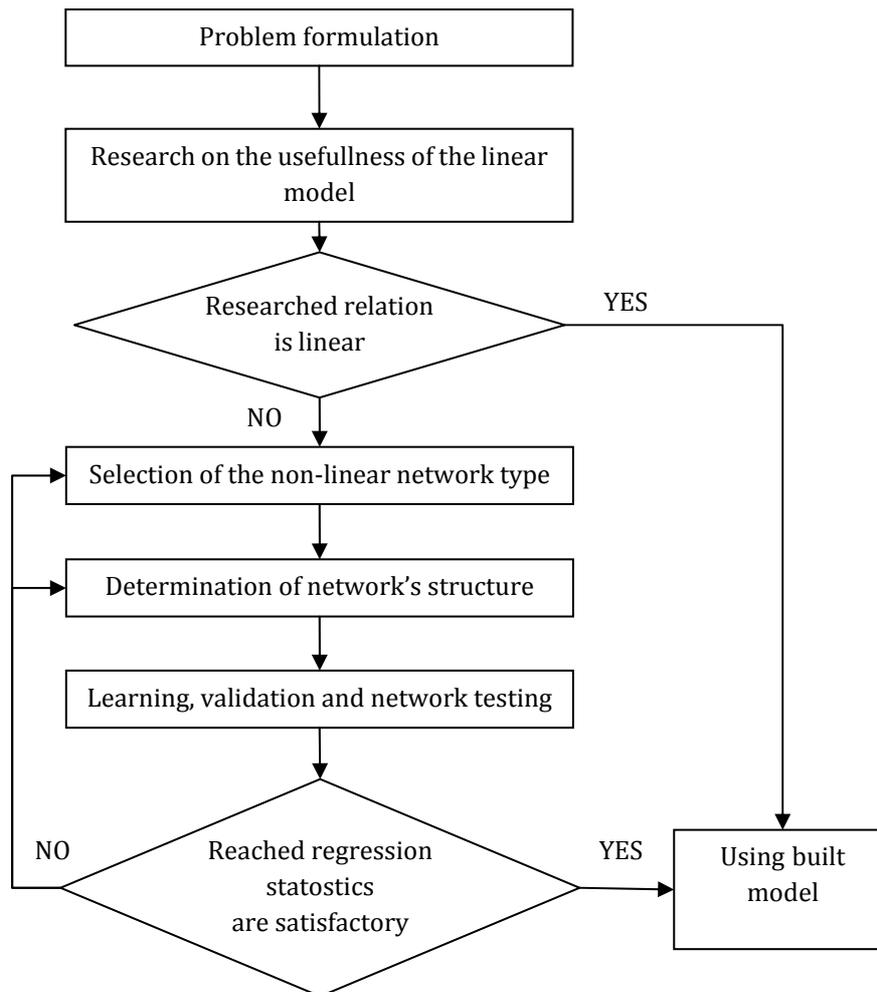


Figure 4. Construction of neural regression model
(source: self study)

Intelligent Problem Solver gives numerous solutions without the necessity for long-term and time-consuming research that require an number of decisions, such as network type and architecture selection, input and other variables selection. Process of construction of separate neural models is tiresome, time-consuming and requires considerable experience from the person performing the experiment. Relatively easy finding of a proper solution, in comparison with the linear statistical approach, is not possible. Successive experiments (with other types and neural architectures) not always lead to the revealing of new, better results.

Mentioned tool is definitely helpful in the network architecture selection process realization. Often it provides better results in comparison with the results reached by the network designers, which only use their intuition.

6 Experiment results

Researched possibilities of using the neural networks in European project evaluation process were realized during the solving of ex-post project realization in the Sectoral Operational Programme: Improvement of the competitiveness of enterprises (six input variables and one output variable). It is a regression problem, because the output variable usually takes continuous numerical values. The assumption is to use linear and non-linear regression techniques. It is not necessary to search for the functional form of regression curve. It is one of the basic advantages of neural networks. They can be treated as “black boxes” with six inputs and one output. Result form is decided by the number of neurons and connections between them. This number should correspond to the complexity of non-linear regression curve. The higher the number is, the more complicated form the curve has.

Fig. 5 presents three windows from Statistica NN 6.0 software: spreadsheet window (with data), interface of Statistica NN (with indicated problem type – regression) and the active window of variable selection (with one output variable highlighted – dependent and six input variables - independent).

After the determination of variables the network types were selected, which would later be taken into consideration in the regression model construction process. Author decided to use the linear, GRNN, RBF and three-layer perceptron networks. Others regression analysis parameters were as follows:

- number of maintained networks: five,
- network maintenance selection criteria: “maintain the balance between the error and diversity of the network”,
- time of optimization – number of tested networks: 10,

- maximal number of tacit neurons: for the RBF network - 25 and for the second perceptron – 10,
- linear activation function for the perceptron,
- communicates: “current information about the better network”.

After the finishing of construction process and the automatic selection of five best (for this particular analysis) neural networks, one can review the “Summary model report” (see Fig. 6) and the worksheet that includes descriptive (regression) statistics of created and recorded models (see Fig. 7). Summary model report is presented in the form of table with columns that include indicators of the, so called, “quality” of learning, validation and testing as well as errors determined for the three data subsets. This “quality” is calculated as the quotient of standard deviation of the remainder and standard deviation for dependent variable (in analyzed case: median of new workplaces number).

The screenshot displays the Statistica NN 6.0 interface. The background is a spreadsheet window titled "STATISTICA - [Dane: 5_Wzrost_konkurencyjności_przedsiębiorstw (7 zmn. * 100 prz.)]". The spreadsheet has 7 columns: 1-Prac szkol, 2-Nowe tech, 3-Prac wyksz, 4-Interesariusze, 5-Wzrost przych, 6-Nowe produkty, and 7-Miejsca pracy. The data rows are numbered 1 to 40. Overlaid on the spreadsheet are two dialog boxes. The top dialog box, titled "STATISTICA Sieci neuronowe: 5_Wzrost_konkurencyjności_przedsię...", shows the "Typ problemu" set to "Regresja" and "Typ analizy" set to "Automatyczny projektant". The bottom dialog box, titled "Wybierz zmienne wejściowe (niezależne), wyjściowe (zależne) i selekcyjną", shows the selection of input variables (1-6) and the output variable (7). The spreadsheet data at the bottom of the image is as follows:

	1	2	3	4	5	6	7
	Prac szkol	Nowe tech	Prac wyksz	Interesariusze	Wzrost przych	Nowe produkty	Miejsca pracy
1	780	17	1120	30,0	9,9	8	2
2	998	14	4480	38,4	8,4	10	22
3	1014	14	3680	39,0	8,4	10	15
4	1045	16					
5	983	14					
6	718	17					
7	1076	16					
8	967	11					
9	920	16					
10	718	18					
11	952	15					
12	936	14					
13	1014	15					
14	874	13					
15	1014	15					
16	905	14					
17	1061	16					
18	796	17					
19	889	14					
20	967	17					
21	1201	19					
22	983	17					
23	1045						
24	1186						
25	764						
26	858						
27	1045						
28	1092						
29	998						
30	952						
31	749						
32	920						
33	858						
34	983						
35	998						
36	811						
37	764						
38	842						
39	1232						
40	686	16	1040	26,4	9,6	7	2

Figure 5. Variable selection window in Statistica Neural Networks 6.0
(source: self study with the use of Statistica software)

Ind.	Typ	Jakość ucz.	Jakość wal.	Jakość test.	Błąd ucz.	Błąd walid.	Błąd test.	Uczenie/składn.	Uwagi	Wejść	Ukryta(1)	Ukryta(2)
1	Liniowa 6:6-1:1	0,266360	0,302721	0,305166	0,106101	0,121109	0,132375		PI	6	0	0
2	MLP 4:4-4:1:1	0,221121	0,253703	0,299442	0,077583	0,076519	0,102152		BP100,CG20,CG15b	4	4	0
3	GRNN 6:6-50-2:1:1	0,239080	0,412636	0,394948	0,029064	0,043871	0,044677		SS	6	50	2
4	RBF 6:6-4:1:1	0,261148	0,404825	0,417589	0,031714	0,042108	0,047130		KM,KN,PI	6	4	0
5	RBF 6:6-9:1:1	0,215615	0,324947	0,353572	0,026184	0,033861	0,040264		KM,KN,PI	6	9	0

Figure 6. Summary report for the work results of the automatic designer
(source: self study with the use of Statistica software)

	Miejsca pracy.1	Miejsca pracy.2	Miejsca pracy.3	Miejsca pracy.4	Miejsca pracy.5
Srednia	11,99000	11,99000	11,99000	11,99000	11,99000
Odch. std.	7,65832	7,65832	7,65832	7,65832	7,65832
Sredni błąd	-1,54018	-0,21688	0,25414	-0,05335	-0,13151
Odch. błędu	2,20514	1,92340	2,47559	2,57433	2,13351
Sr. bł. bezwz.	2,01110	1,51350	1,96762	1,98722	1,60436
Wzrost odch.	0,28794	0,25115	0,32326	0,33615	0,27859
Korelacja	0,95854	0,96817	0,94641	0,94185	0,96043

Figure 7. Regression statistics of five created neural models
(source: self study with the use of Statistica software)

In case of resignation from using the used neural regression model predicted values of dependent variable would be based on the arithmetical average from the values in the learning set. In such case the mentioned quotient (“quality”) would be equal to 1. Remainder would be the result of division of output variable value and the average value. Numerator and denominator of the quotient of deviations would have the same value.

Summing up, the value of deviation quotient close to unity (or higher) disqualifies given neural model. It indicates that the model, determined with a considerable amount of time, is similar (or even worse in case of values exceeding 1) to the determined average of dependent variable value.

Mean error (subtractions) between real and determined values has a relatively low value (from -0,05 to 0,25 for four models). Absolute values were not calculated from these subtractions.

Standard deviations from errors’ series (approximately 1,9 ... 2,5) indicate the insignificant deviations of errors from the mean value. This parameter is significant from the point of view of the most important measure of the model – deviation quotient.

Average absolute error was determined (contrary to the error described above) with the calculation of absolute values from the subtractions between theoretical values (data) and the values at the output of the model. Relatively low averages (approximately 1,5 ... 2,0) determined from subtraction differences indicate that the model was properly elaborated.

Correlation indicators are in the last row of regression statistics table (previously described). Indicators, reached in an exemplary analysis, close to one (approximately 0,94 – 0,96) indicate that the models were properly created. Deviation quotient is also at satisfactory level (0,25 – 0,34). The lower its value is the better.

In the selection of the neural network from the five reached models it is necessary to bear in mind the linear network parameters as a reference point for the comparative analysis. All non-linear networks reach lower values of errors in comparison with a linear network. It indicates the non-linear character of the analyzed regression issue.

7 Summary

Review of currently used European project evaluation methods was performed in the process of the research realization. Special attention was drawn to the use of regression analysis in this application. Examples of a multi-criterion European project evaluation problem solving were presented. Moreover the use of modern computing technologies for the construction of regression models was proposed. Elaboration focused in the use of learning systems, which base on neural networks, in the project evaluation.

Models, which use neural networks, have many desired and significant features. They are non-linear, non-parametrical and do not assume the outline of the input and output relation modeling function. They do not require assumptions on forms and parameters of random variable distribution. Models are resistant to disturbances present in real-life systems. They allow gathering additional knowledge sufficient and necessary for regression data analysis as well as the selection of data relevant for evaluation purposes with simultaneous elimination of irrelevant factors. In case of classical statistical methods it is necessary to determine the relations between the explanatory variables and the explained variable as well as justification of these relations. However, it is not necessary to use e.g. neural networks. Formulated feature is at the same time certain limitation of the neural network application possibilities, because the result reaching process cannot be justified with particular relations.

Method feature analysis (advantages and disadvantages) that are included in artificial intelligence (especially neural networks) leads to a conclusion that the methods can be significant supplement for other (classical) data analysis methods (e.g. regression equation).

Example presented in the article shows examples of neural network usage in the construction of regression model, indicating that they are useful in regression analysis, similar to the classic single-equation regression equations. In case of classical methods usage that bases on neural networks both methods can complete each other. There are no problems in using these methods simultaneously.

Results reached from the data analysis method, which bases on neural networks, create empirical basis for the formulation of project evaluation, thus the realization of the final stages of European project evaluation.

Author used the network type and structure intuitively, using previous experience. However, this approach led to the creation of networks with not-satisfactory parameters (e.g. improper number of neurons or layers). Wrongly selected models can include e.g. too little number of neurons. This leads to insufficient approximation ability. In other cases it can result in relatively low ability of the network for generalization of the knowledge gathered in the learning process.

IT tools from the Statistica Neural Networks software proved to be a significant improvement in the network type and architecture selection process. These tools are especially useful in the construction of regression model used in evaluation process of EU cofinanced projects. Use of such tools usually enables reaching better regression parameters in comparison with the results gained with intuition or trial and error.

Results in the scope of neural model implementation look promising and can be used in the European project evaluation systems. Further experiments, with the use of artificial intelligence methods, in this field are planned. Authors interests are especially focused on the knowledge bases built with the use of the rough set theory.

8 References

- [1] Duch W., Korbicz J., Rutkowski L., Tadeusiewicz R. - *Sieci neuronowe, w Biocybernetyka i inżynieria biomedyczna 2000*. PAN, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2000.
- [2] Grzeszczyk T.A. - *Application of Neural Networks for Prior Appraisal of Structural Funds Project Proposals* [in] Proceedings of the Sixth International Conference on Enterprise Information Systems. Vol. 2, Portucalense University, Porto, Portugal 2004, pp. 501-504.
- [3] Grzeszczyk T.A. - *Artificial Intelligence Applied for Forecasting in Enterprise Decision Support*. Publishing House of Institute of Production Systems Organization Warsaw University of Technology, Warszawa 2005.
- [4] Grzeszczyk T.A. - *A Rough Set Method for Knowledge Management in the Process of Structural Funds Projects Preparation* [in] Journal of Tung Fang Institute of Technology, No. 25, Taiwan 2005, pp. 257-271.
- [5] Grzeszczyk T.A. - *Ocena projektów europejskich 2007-2013*. Placet, Warszawa 2009.

- [6] *Guide to the Evaluation of Socio-Economic Development*. www.evaled.info/frame_about.asp, 2007.
- [7] Kolenda K., Kolenda M. - *Analiza i prognozowanie szeregów czasowych. Program komputerowy*. Wydawnictwo Placet, Warszawa 1999.
- [8] Lula P. - *Jednokierunkowe sieci neuronowe w modelowaniu zjawisk ekonomicznych*. Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków 1999.
- [9] *MEANS Collection, Volume 1, Evaluating socio-economic programmes, Evaluation design and management*. European Commission Structural Funds, Luxembourg 1999.
- [10] *MEANS Collection, Volume 3, Evaluating socio-economic programmes, Principal evaluation techniques and tools*. European Commission Structural Funds, Luxembourg 1999.
- [11] Pszczołowski T. - *Mała encyklopedia prakseologii i teorii organizacji*. Zakład Narodowy im. Ossolińskich, Wrocław 1978.
- [12] Sokołowski A, Zaleski J. - *Niezbędne warunki umożliwiające wdrożenie dla Polski oceny sumarycznego efektu funduszy strukturalnych i funduszu spójności za pomocą modeli ekonometrycznych*. www.europa.edu.pl, 2004.
- [13] *Statsoft: statystyki regresyjne*. www.statsoft.com/textbook/glosr.html, 2007.
- [14] Stefanowicz B. - *Metody sztucznej inteligencji i systemy eksperckie*. Monografie i opracowania, SGH, Warszawa 1993.