

REGRESSION ANALYSIS AND NEURAL NETWORKS AS METHODS FOR PRODUCTION TIME ESTIMATION

Predrag Čosić, Dragutin Lisjak, Dražen Antolić

Original scientific paper

An experienced process planner usually makes decisions based on comprehensive data without breaking it down into individual parameters. So, as the first phase it was necessary to establish a technological knowledge base, define features of the 2D drawing (independent variables), possible dependent variables, size and criteria for sample homogenization (principles of group technology) for carrying out analysis of variance and regression analysis. The second phase in the research was to investigate the possibility for easy automatic, direct finding and applying 3D features of an axial symmetric product to the regression model. The third phase in the research was to investigate the possibility for the application of neural networks in production time estimation and to compare the 224 results between the regression models and neural network models. The most important characteristic of our approach presented in this paper is estimation of production times using group technology, regression analysis and neural networks.

Keywords: *group technology, knowledge base, neural networks, production time, stepwise multiple linear regression, Total Cost Estimation*

Regresijska analiza i neuronske mreže kao metode za procjenu proizvodnog vremena

Izvorni znanstveni članak

Iskusni planeri tehnoloških procesa donose odluke temeljene sveobuhvatnim podacima bez bavljenja pojedinim parametrima. Tako, u prvoj je fazi bilo potrebno kreirati tehnološku bazu podataka, definirati značajke 2D crteža (nezavisne varijable), moguće zavisne varijable te veličinu i kriterije za homogenizaciju uzorka (načela grupne tehnologije) za provođenje analize varijance i regresijsku analizu. Druga faza istraživanja bilo je istraživanje mogućnosti jednostavnog i direktnog traženja i primjene 3D značajki aksijalno simetričnog proizvoda u regresijskom modelu. Treću fazu istraživanja činilo je proučavanje mogućnosti primjene neuronske mreže u procjeni proizvodnog vremena proizvoda te usporedba 224 rezultata između regresijskog modela i neuronske mreže. Najznačajnija karakteristika našeg primijenjenog pristupa, prezentiranog u radu, je procjena proizvodnih vremena primjenom grupne tehnologije, regresijske analize i neuronskih mreža.

Ključne riječi: *grupna tehnologija, baza znanja, neuronske mreže, vrijeme proizvodnje, postupna višestruka linearna regresija, ukupna procjena troškova TCE Total Cost Estimation*

1

Introduction

An experienced process planner usually makes decisions based on comprehensive data without breaking it down into individual parameters. So, as the *first phase* it was necessary to establish a technological knowledge base, define features of 2D drawings (independent variables), possible dependent variables, size and criteria for sample homogenization (principles of group technology) for carrying out analysis of variance and regression analysis. The *second phase* in the research was to investigate the possibility for easy automatic, direct finding and applying 3D features of an axial symmetric product to the regression model. The defined requirements resulted in the development of the procedure for retrieval of parameters from the 3D model with a low level of subjectivity, a very fast and reliable process via CAD report to the regression model [12]. The *third phase* in the research was to investigate the possibility for the application of neural networks in production time estimation and to compare the results between the regression models and neural network models. As it can be seen from the list of references, different approaches are used for data retrieval from AD (STEP) [1], integration of CAPP, CAD/CAM and business activities [2], development of database system of mechanical components [3, 4], and integrated product engineering [5] for costs estimation and rapid cost estimation [6], application of neural networks in estimation of production times [7], connection from CAPP, CAD, CAM; DFX to DFA through product development [8], etc. The most important characteristic of our approach presented in this paper is estimation of production times using group technology, regression analysis and neural

networks [7, 9, 10, 13].

2

Drawing features and technological database for production time estimation

Very frequently (especially in the case of Small and Medium Enterprises - SMEs) it is necessary to respond quickly to some important requests for offers, generated for individual or batch production, for example in the case of:

- 1) a great number of requested offers for manufacturing of products at once,
- 2) small batches that are very rarely repeated,
- 3) frequent changes of priorities during production,
- 4) short deadlines for delivery,
- 5) market demands for bringing prices of individual or batch production close to the prices of mass production, etc.

It must be noted that technological knowledge and speed of process planning are often more important than the technological level of equipment, skills and knowledge of people who implement the technology. So, very often in practice we can be faced with the following:

- a) A great amount of time spent on planning of the technological process for a product without any specific contract being made concerning the order for manufacturing of the product,
- b) Signing of a contract without estimated precise production times/costs necessary for product manufacturing and realization in accordance with contracted production.

Technological processes are basically based upon

Table 1 Explanation of the meaning of used symbols

Symbol	Physical unit	Meaning of the symbol
f_{ea}	–	Features of 3D
K	–	Coefficient of time
K_s	–	All dimension lines
R^2	–	Index of determination
t	min	Machining time
x_1	IT	Order of tolerance outside diameter
x_2	mm	Outside diameter of material
x_4	mm	Width of material
x_6	mm	Length of material
x_8	Class h	Roughness of open areas
x_9	HRC	Hardness of product
x_{10}	mm	Outside diameter of product
x_{11}	mm	Inside diameter of product
x_{15}	–	Number of product perspectives
x_{16}	–	Number of descriptions of product
x_{18}	–	Number of location marks (geometry)
x_{19}	–	Number of dimension line tolerances
x_{20}	–	Number of special dimension lines
x_{21}	–	Number of usual dimension lines
x_{22}	1/class	Roughness request Ra
x_{23}	1/mm	Location request (geometry)
x_{24}	1/mm	Dimension request
x_{25}	1/IT	Diameter request
x_{26}	mm ²	Area of sketch
x_{29}	N/mm ²	Ultimate tensile strength of material
x_{30}	m ²	Requested area of sketch
x_{31}	–	Mass strength of material
x_{32}	mm	Thickness wall of products
x_{33}	–	Ratio of diameter and length
x_{39}	–	Number of all dimension lines
x_{40}	–	Product complexity
x_{42}	Class h	Difference in roughness
x_{43}	dm ²	Difference in superficial areas of material
x_{44}	cm ³	Volume of material
x_{45}	kg	Mass of material
x_{46}	mm	Difference in outside diameters
x_{47}	mm	Difference in outside diameter of products
x_{49}	mm	Difference in thicknesses
x_{50}	mm	Difference in lengths
Y	min	Production time

product drawings with adequately defined dimensions, tolerances (dimensional and geometrical), surface roughness, batch size, shape and kind of material, heat treatment, requested delivery, disposable equipment, tools, etc. At the same time, process plans are primarily the result of the planner's experience, intuition and decision support. A process planner can establish possible connections between drawing features and necessary production times for products manufacturing. The fundamental idea in the approach [10, 13] to production time estimation is investigation of the existence of some kind of relationship between the shape and data from the drawing and the process type, process sequencing, primary process, way of clamping, selection of tools, machine tools, production times, etc. As one of the *first steps in our project research*, we defined possible shapes of raw material and 30 potential basic technological processes.

3

Development of stepwise multiple linear regressions

A desirable level of generalization in regression analysis will be an important indicator for the quality of regression equation. One of the most important problems was the process of homogenization of the sample of products. Adequate method for this action was one of the methods of group technology. For the sample of real products (420 parts) and considered features, we created, as a result of our investigation and step multiple linear regression (in previous research and papers), 8 regression equations for different groups of parts with different number and kind of independent variables. So, we can see for different values of parts' features (independent variables), the values for the estimation of production times (dependent variables). However, the results of process optimization for 8 regression equations by genetic algorithm cannot be applied in real production. Logical operators during query process in database Access were very helpful in the process of homogenization of the sample of products. As the result of previous research, sample homogenization, classifier selection and stepwise multiple linear regression, we obtained: 7 independent selected variables, basic sample of 320 parts, constraints for data parts, 8 regression equations, percentage of explained effects, relative error (7-30 %), etc. (Tab. 1). The lowest relative error 8,01 % (Tab. 2, for ground discs, AC102 No. 5) and the highest index of determination $R^2 = 0,9851$ for the ground discs group are the consequence of the simultaneous action of logical operators (round bars, discs and fine machining – i.e. diameter tolerance better than IT7). Thus, with the simultaneous action of several operators, a lower scattering of production time values has been achieved, i.e. better homogeneity of the created group. Since there was too great subjective influence of workers in the process of filling in the values of independent variables, we continued with investigation in the 3D area. The question was how to get automatically the 3D features from CAD application (CATIA, PRO/E) in the application for developed regression equations and avoid thus this subjectivity factor. The *second phase* in the research was the investigation of the possibility for easy automatic, direct retrieval of 3D features of the considered axial symmetric product into the regression model. The defined requirement resulted in the development of the process for the transfer of parameters from 3D models with a low level of subjectivity. It is a very fast and reliable process via CAD report to the regression model [11]. As a possibility to improve our precision in estimating the production times of 'new unknown products', in the next phase of our research we chose a neural network to compare the validity of the two methods: linear multiple stepwise regression and neural network model.

4

Neural network model

Artificial neural networks (ANN) are inspired by the biological neural system and its ability to learn through examples. Instead of following a group of well defined rules specified by the user, neural networks learn through intrinsic rules obtained from presented samples. The most commonly used ANN architecture is the multilayer *backpropagation neural network*. Backpropagation was created by generalizing the *Widrow-Hoff* learning rule to multiple-layer networks and nonlinear differentiable

Table 2 Presentation of created regression equations

No.	Shape of product representative of product group	Regression equations	Index of determ. R^2	Relative error / %	Comment on regression equation
1	Whole sample A000	$t = -11,69 + 16,95 \cdot x_{45} + 1,22 \cdot x_{40} + 0,54 \cdot x_{47} + 127,47 \cdot x_{22} - 3,24 \cdot x_{18} + 0,15 \cdot x_{32} + 0,03 \cdot x_6$	0,736552	30,74	Model is developed with procedure in advance. Three independent variables are omitted x_8 , x_{19} and x_{33} .
2	Round bars A00B1	$t = 55,47 + 22,43 \cdot x_{45} + 1,162 \cdot x_{40} + 0,43 \cdot x_{11} + 1,61 \cdot x_{50} - 5,41 \cdot x_8 - 3,26 \cdot x_{18} + 1,78 \cdot x_{42}$	0,74285	30,95	Model is developed with procedure in advance. Two independent variables are omitted x_1 and x_{26} .
3	Shafts AB101	$t = 6,13 + 0,83 \cdot x_2 + 1,27 \cdot x_{39} - 3,30 \cdot x_8 + 5,51 \cdot x_{46} - 6,86 \cdot x_{18} + 0,09 \cdot x_6 + 124,33 \cdot x_{22}$	0,807626	25,90	Model covers a narrower field of rotational parts. It gives better results than No. 2.
4	Discs AB1C1	$t = -5,17 + 0,73 \cdot x_{47} + 0,93 \cdot x_{40} + -5,25 \cdot x_{20} + 0,52 \cdot x_{24} + 139,11 \cdot x_{30} + 0,23 \cdot x_{32} - 0,51 \cdot x_{33}$	0,809405	24,24	Similar results as in No. 3.
5	Discs-with fine machining AC102	$t = -60,78 + 0,59 \cdot x_{47} + 0,47 \cdot x_9 + 0,74 \cdot x_1 + 0,25 \cdot x_{10} + 0,84 \cdot x_{39} + 291,07 \cdot x_{25} + 5,9 \cdot x_{15}$	0,985057	8,01	Model covers a narrower field of rotational parts. It gives better results than all the previous models.
6	Rotational parts AB103	$t = -37,11 + 0,94 \cdot x_{40} + 0,03 \cdot x_{29} + 319,22 \cdot x_{26} + 0,13 \cdot x_{23} + 114,67 \cdot x_{43} - 80,98 \cdot x_{45} - 0,46 \cdot x_6$	0,893321	27,06	Model is better than No. 2 as a result of higher degree of homogenization of data. Solution is better with omitted variables x_2 and included variables x_6 , x_{23} , x_{43} and x_{45} .
7	Flat bars A0004	$t = -10,96 + 0,58 \cdot x_{40} + 34,50 \cdot x_{45} + 218,42 \cdot x_{22} - 5,48 \cdot x_{50} + 185,03 \cdot x_{26} + 0,39 \cdot x_9 - 0,50 \cdot x_{49}$	0,900332	15,92	Constraints are greater for all variables so results are better. Narrow field of homogenization.
8	Sheet metals A0005	$t = 0,47 + 1,27 \cdot x_{40} + 37,45 \cdot x_{45} - 13,23 \cdot x_{43} - 0,70 \cdot x_{43} + 0,28 \cdot x_4 + 0,05 \cdot x_6 + 3,91 \cdot x_{16}$	0,900823	24,04	Model is characterized with the presence of complex variables x_{40} , x_{43} , x_{45} .

transfer functions [14]. Input vectors and the corresponding target vectors are used to train the network until it can approximate a function, associate input vectors with specific output vectors. Standard backpropagation is a gradient descent algorithm, as is the *Widrow-Hoff* learning rule, in which the network weights are moved along the negative of the gradient of the performance function. The term *backpropagation* refers to the manner in which the gradient is computed for nonlinear multilayer networks. *Backpropagation* neural networks often have one or more hidden layers of sigmoid neurons followed by an output layer of linear neurons. Multiple layers of neurons with nonlinear transfer functions allow the network to learn nonlinear and linear relationships between input and output vectors. There are numerous variations of the basic algorithm that are based on other standard optimization techniques, such as conjugate gradient and Newton methods. The one used in this paper is the *feedforward backpropagation* training algorithm designed to minimize the *mean square error (MSE)* between the actual (estimation) output (a , A) and the desired (target) output (d , T).

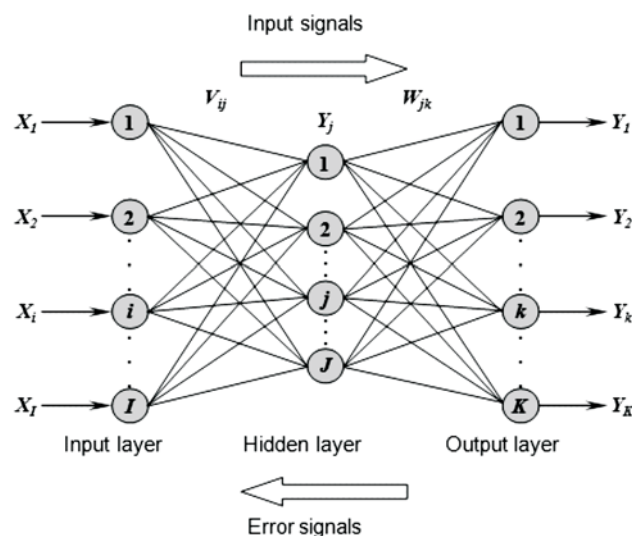


Figure 1 Principle of the feedforward backpropagation training algorithm

Fig. 1 shows the principle of the *feedforward backpropagation* training algorithm, where: V_{ij} – weight

between the input layer and the hidden layer, W_{jk} – weight between the hidden layer and the output layer, X_i – input signals, i – number of neurons of the input layer, I – number of inputs of neuron j in the hidden layer, Y_j – output of the hidden neurons, j – number of neurons of the hidden layer, J – number of inputs of neuron k in the output layer, Y_k – output signals, k – number of neurons of the output layer. For the estimation of performance of the learning algorithm in solving the specified task, performance index was defined. Performance index enabled comparison of the applied neural network algorithm with other learning algorithms. The most frequent performance index is the *normalized root mean square error – NRMSE*.

Where: N is the total number of patterns, d_n is the desired (target, T) outputs, a_n is the actual (estimation, A) outputs, σ_{dn} is the standard deviation.

$$NRMSE = \frac{\sqrt{\frac{\sum_{n=1}^N (d_n - a_n)^2}{N}}}{\sigma_{dn}} \tag{1}$$

$$\sigma_{dn} = \sqrt{\frac{1}{N} \sum_{n=1}^N (d_n - \bar{d})^2}$$

$$\bar{d} = \frac{1}{N} \sum_{n=1}^N d_n$$

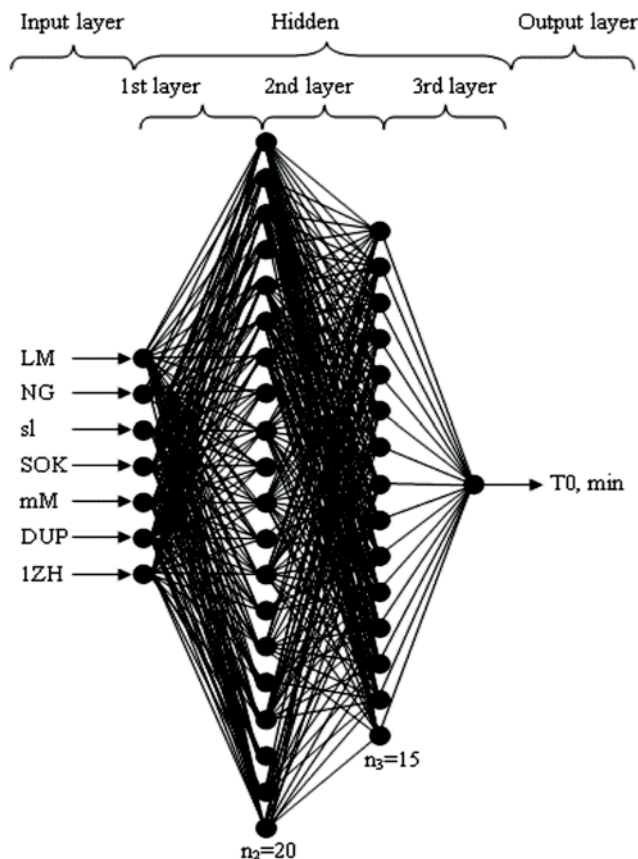


Figure 2 Simplified model of the used neural network of A0000 model

5 Experimental results

As a better method for solving the problem of production time estimation, we proposed a three-layer backpropagation neural network the simplified structure of which is shown in Fig. 2.

Presented input parameters (LM-lengths of material, NG-number of position, sl-thickness of the workpiece, SOK-complexity of the workpiece, mM-mass of material, DUP-internal diameter, 1ZH-outside diameter) refer to the model A0000. Output parameter (TO) is the estimate of time in minutes. Parameters $n_2=20$ and $n_3=15$ represent the of neurons in the second and third layer of the network. Between the layers the following transfer functions are applied: *tansig-tansig-purelin*. Data important for neural network training are: Performance goal: 0,0001, Learning rate-0,01, Ratio to increase learning rate-0,5, Maximum performance increase: 1,04, Maximum performance gradient: 1e-10, Momentum constant: 0,9, Number of layers: 3, Number of neurons: 20-15-1, Transfer functions: *tansig-tansig-purelin*, Number of epoch to train: 15000. For neural network training the available experimental data are divided in three sets: training set (70 %), validating set (15 %), and testing set (15 %). The same model of experimental data division is applied to all models. The following parameters are selected as *key performance indexes* of the neural network model (NNM) in relation to the regression model (RM): R (correlation coefficient), R^2 (index of determination), $RMSE$ (root mean square error) and $NRMSE$ (normalized root mean square error). In Fig. 3 through Fig. 6 for each experimental model the graphical presentation of parameter R values and tabulated values of parameters R , R^2 , $RMSE$ and $NRMSE$ are given.

6 Conclusion

According to the presented results (Fig. 3 through Fig 6), we can conclude that the assumption on the use of a neural network for the production time estimation in relation to a classical robust regression model is justified. For all experimental models (A0000, A00B1, AB1C1, AB101, AC102, AB103, A0004, A0005) the applied backpropagation neural network gives better values of key performance indexes (R , R^2 , $RMSE$, $NRMSE$). The biggest difference between the key performance indexes for NNM and RM estimation models is in the case of model A0005 (input set of 35 data), and lowest in the case of model AC102 (input set of 25 data). The next differences in key performance indexes of individual models ranged from the highest to the lowest values are as follows: AB103, A0004, AB1C1, A00B1 and AC102. The lowest difference between the NNM and RM estimation model in AC102 (finely machined discs) follows from the nature of the model independent variables and their values that are from a relatively narrow range. The reverse is true for the biggest difference in A0005 (sheets), because of the relatively wide range of independent variables. The key performance indexes in NNM estimation models are significantly better than those in all proposed RM models, especially in the case of A0005. The reason for this is the proper selection of transfer functions (*tansig - tansig - purelin*) within the backpropagation neural network layers which provide approximation of linearities and nonlinearities within

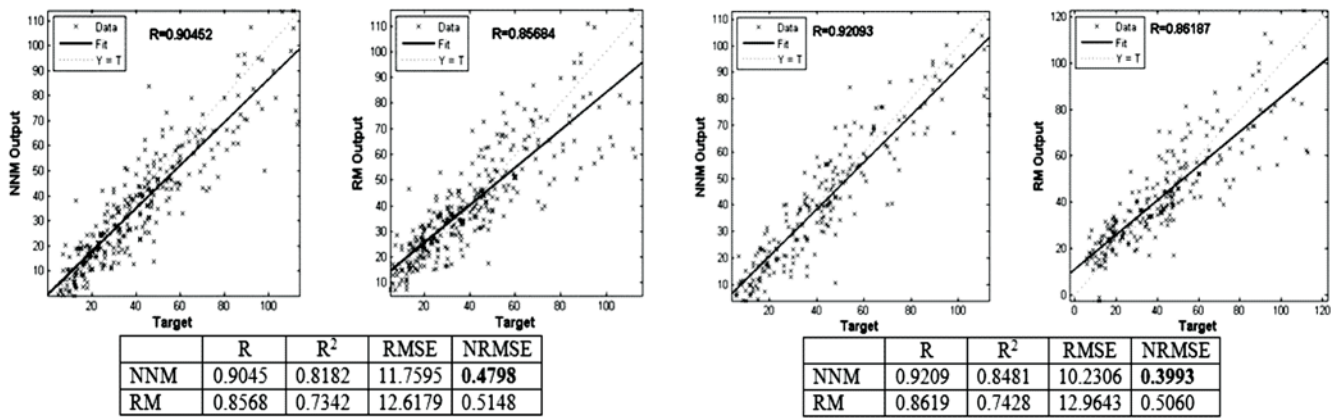


Figure 3 Models: A0000 and A00B1

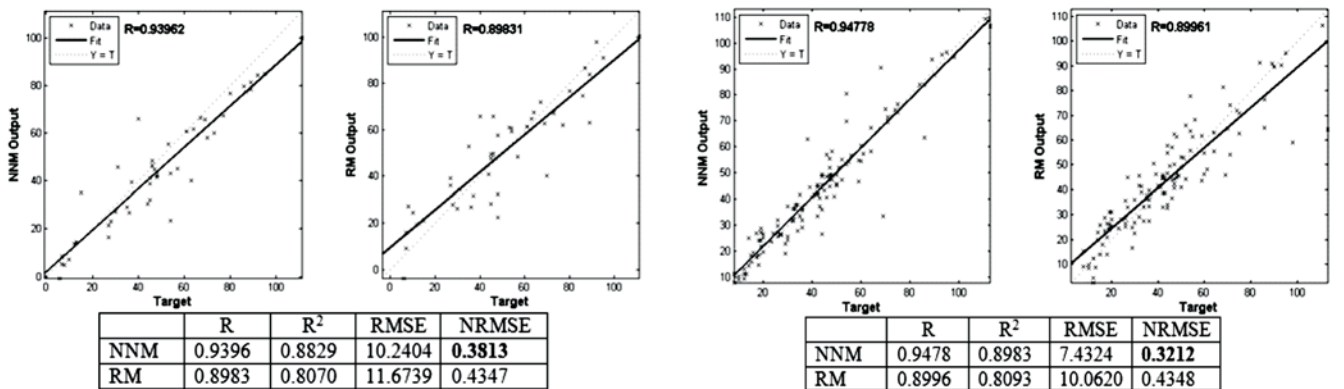


Figure 4 Models: AB101 and AB1C1

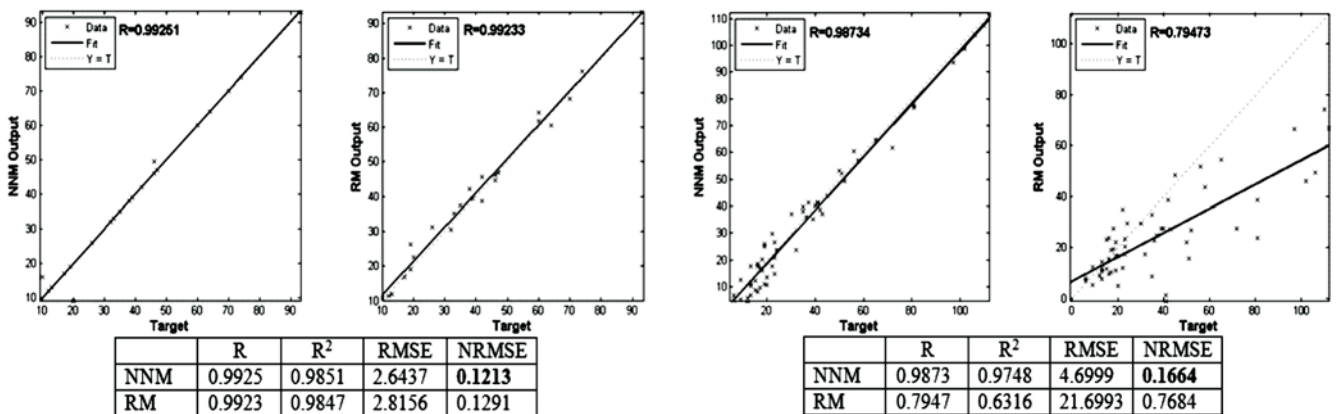


Figure 5 Models: AC102 and AB103

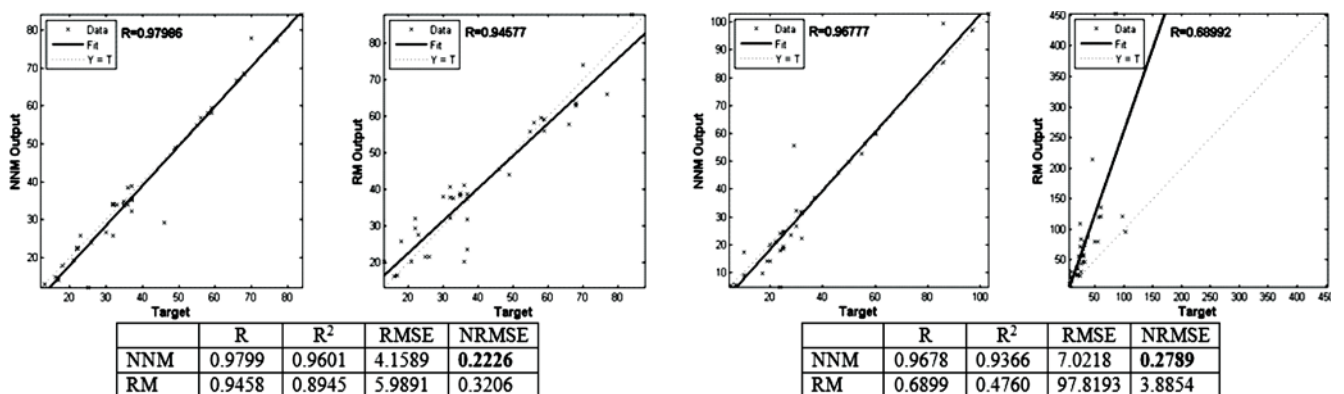


Figure 6 Models: A0004 and A0005

independent variables, as opposed to the regression model whose approximation is only linear. It should be also noted that the estimation by the *NNM* model would be even better

if the experimental data had not been divided in three sets (training set, validating set, and testing set), and thus the estimation made by *NNM* model is based on 70 % of the

presented data in contrast to the *RM* model where for the obtaining of the regression function all experimental data are always presented.

Acknowledgements

This paper is a part of the scientific projects (2007-2011) titled *Process Production Impacts to the Competitive and Sustainable Development*, 120-1521781-3116 (Main Researcher: Predrag Cosic, PhD, Full Professor) financed by the Ministry of Science Education and Sports of the Republic of Croatia. We express our gratitude for the financial support for these projects.

7

References

- [1] Han, J.; Kang, M.; Choy, H. STEP – based recognition for manufacturing cost optimization, // *Computer-Aided Design*, 33, (2001), pp.671 – 686.
- [2] Culler, D. E.; Burd, W. A framework for extending computer aided process planning to include business activities and computer aided design and manufacturing (CAD/CAM) data retrieval, // *Robotics and Computer-Integrated Manufacturing*, 23, (2007), pp. 339 – 350.
- [3] El-Melahalawi, M.; Miller, R. A. A database system of mechanical components based on geometric and topological similarity. Part I: representation, // *Computer-Aided Design*, 35, (2003), pp. 83 – 94.
- [4] El-Melahalawi, M.; Miller, R. A. A database system of mechanical components based on geometric and topological similarity. Part II: indexing, retrieval, matching, and similarity assessment, // *Computer-Aided Design*, 35, (2003), pp. 95 – 105.
- [5] H'mida, F.; Martin, P.; Vernadat, F. Cost estimation in mechanical production: The cost entity approach applied to integrated product engineering, // *Int. J. Production Economics*, 103, (2006), pp. 17 – 35.
- [6] de Cos, J.; Sanchez, F.; Ortega, F.; Montequin, V. Rapid cost estimation of metallic components for the aerospace industry, // *Int. J. Production Economics*, doi: 10.1016/j.ijpe.2007.05.016. (2007)
- [7] Simunovic, G.; Saric, T.; Lujic R. Application of Neural Networks in Evaluation of Technological Time, // *Strojniški Vestnik-Journal of Mechanical Engineering*, 54, 3(2008), pp. 179-188.
- [8] Kostanjevec, T.; Balic, J.; Polajnar, A. Multiple Criteria Analysis of a Large-Scale Product with the DFX Method, // *Advances in Production Engineering & Management*, 4, 3(2009), pp. 105-114.
- [9] Verlinden, B.; Duflou, J.; Collin, R.; Cattrysse, D. Cost Estimation for sheet metal parts using multiple regression and artificial neural networks: A case study, // *Int. J. Production Economics*, 111, (2008), pp.484-492.
- [10] Cosic, P.; Antolic, D.; Milic, I. Web Oriented Sequence Operations, //19th International Conference on Production Research, ICPR-19, July 29-August 2, 2007, Valparaiso, Chile, on CD.
- [11] Cosic, P.; Milicic, D.; Kovacic, I. Production time estimation as the part of collaborative virtual manufacturing, // *International Conference on Industrial Logistics, ICIL, 2008, TelAviv: ICIL, 2008., pp.93-100.*
- [12] Antolic, D. Estimation of production times by regression models (in the Croatian language), Master's thesis, University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture (FSB), 2007.
- [13] Kovacic, I. An overview of fast estimation of production times and delivery deadlines, Graduation thesis, University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture (FSB), 2007.
- [14] Hagan, T. M.; Demuth, B. H.; Beale, H. M. *Neural Network Design*; PWS, Publishing Company: Boston, 1995.

Authors' addresses

Predrag Čosić

University of Zagreb
Faculty of Mechanical Engineering and Naval Architecture
Ivana Lucića 5
10000 Zagreb, Croatia
Tel. 01/61 68 340
E-mail: predrag.cosic@fsb.hr

Dragutin Lisjak

University of Zagreb
Faculty of Mechanical Engineering and Naval Architecture
Ivana Lucića 5
10000 Zagreb, Croatia
Tel. 01/61 68 377
E-mail: dragutin.lisjak@fsb.hr

Dražen Antolić

ADPP, Ilirski trg 5
10000 Zagreb, Croatia
Tel. 01 /4851 524
E-mail: drazen@adpp.hr