Int J Adv Manuf Technol (2006) 28: 688–693
DOI 10.1007/s00170-004-2429-4

ORIGINAL ARTICLE

Z.W. Zhong · L.P. Khoo · S.T. Han

Prediction of surface roughness of turned surfaces using neural networks

Received: 3 July 2004 / Accepted: 21 September 2004 / Published online: 23 November 2005

Abstract In this study, the prediction of surface roughness heights R_a and R_t of turned surfaces was carried out using neural networks with seven inputs, namely, tool insert grade, workpiece material, tool nose radius, rake angle, depth of cut, spindle rate, and feed rate. Coated carbide, polycrystalline and single crystal diamond inserts were used to conduct 304 turning experiments on a lathe, and surface roughness heights of the turned surfaces were measured. A systematic approach to obtain an optimal network was employed to consider the effects of network architecture and activation functions on the prediction accuracy of the neural network for this application. The reliability of the optimized neural network was further explored by predicting the roughness of surfaces turned on another lathe, and the results proved that the network was equally effective in predicting the R_a and R_t values of the surfaces machined on this lathe as well.

Keywords Neural network · Prediction · Surface roughness · Turning

1 Introduction

The manufacture of dimensionally accurate, closely fitting parts is essential in interchangeable manufacturing. The accuracy and wearability of mating surfaces is directly proportional to the surface finish produced on the part. Good surface finish also contributes to the aesthetic appeal of the product. Though many new cutting tools and methods have evolved, much work remains to be done before all the factors contributing to the surface finish and tool life can be controlled [1].

Surface finish is an important attribute of quality in any machining operation. Researchers have studied the influence of various factors that can improve the surface finish of a workpiece, but results are not very satisfactory, especially in terms of the complex interactions of the various factors [2].

The theoretical arithmetic average surface roughness (mm), R_a, is given by [3, 4]

$$R_a \approx 0.032\frac{f^2}{R} \tag{1}$$

where f = feed rate (mm/rev), and R = tool nose radius (mm). This means that surface roughness increases with increasing feed rate, and a large tool nose radius reduces surface roughness of the workpiece.

However, when aluminum and copper alloys were machined using ultra precision turning processes, R_a decreased with decreasing tool nose radius, contradictory to the theoretical prediction. Moreover, a significant trend in the surface roughness with increasing depth of cut could not be found, which differs from the traditional expectation that a greater depth of cut results in greater roughness. One reason for this might be the presence of voids, impurities, swelling and recovery of the workpiece materials that might affect the trend from occurring. However, the trend that surface roughness decreased with increasing spindle rate was observed [5].

Many theoretical models have concluded that the effect of cutting speed on surface finish is insignificant. In practice, however, cutting speed has been found to be a significant factor. It was found that surface finish improves with the increase of cutting speed for most metals such as aluminum alloy 390, medium carbon leaded steel 10L45 and ductile cast iron, and surface finish deteriorates for Inconel 718 with an increase in cutting speed. The speed also has mixed effects on the surface finish of steel workpieces. An intermediate region of deterioration on surface finish due to the formation of built-up edge was discovered, but this trend was not observed with ceramic tools [6].

Depth of cut has been found to be an important factor affecting the finish of machined surfaces. The critical depths of cut for performing ductile-mode turning of aluminum-based metal matrix composites reinforced with SiC and Al_2O_3 particles were found to be 0.2 and 1 μm, respectively. At these depths of cut, there was almost no subsurface damage [7].

Z.W. Zhong (✉) · L.P. Khoo · S.T. Han
School of Mechanical & Production Engineering,
Nanyang Technological University,
50 Nanyang Avenue, Singapore 639798, Singapore
E-mail: mzwzhong@ntu.edu.sg

Choice of tool insert grades is vital in producing good surface finish. Some of the most commonly used grades are carbides, ceramics, cubic boron nitride (CBN), and diamond [8].

$TiC + TiN$, $TiC + Al_2O_3 + TiN$ and $TiC + Al_2O_3$ coatings give improved performance in crater wear compared with uncoated carbides. The cutting conditions dictate the coating combination best suited for an application. Polycrystalline diamond (PCD) tools provide maximum tool life at high cutting speeds due to their high hardness, and they are recommended for machining aluminum alloys and hypereutectic aluminum-silicon alloys. They are also often used for machining copper and copper alloys at cutting speeds of over 500 m/min. Diamonds are not commonly used for machining of steel because tool wear is very rapid. A single crystal diamond (SCD) is able to produce mirror surface finish given the proper cutting parameters [9].

The sources of machining errors can be divided into three groups: geometric, thermal and cutting force. Among these, thermal errors can comprise 40%–70% of the workpiece errors in precision cutting [10]. The use of a neural network to predict thermally induced errors of the workpiece was conducted [11], its validity was confirmed, and it was found that thermally induced errors increased steadily with time.

Neural networks are models intended to imitate some functions of the human brain using its certain basic structures. Biological networks can process millions of input stimuli in milliseconds even though the processes are electrochemical in nature, and therefore propagate at relatively slow millisecond rates. Although digital computers can perform operations in picoseconds, they fall far short of the performance exhibited by biological systems in their processing capabilities. Researchers have been inspired to set up artificial neural networks (ANNs) based on our knowledge of biological nervous systems. ANNs have been shown to be effective as computational processors for various associative recall, classification, data compression, combinational problem solving, adaptive control, modeling and forecasting, multisensor data fusion and noise filtering [12].

Back propagation is a method of minimizing a squared error by back propagating the error from the output layer to the hidden layers and then back to the input layer, and has the capability to obtain complex non-linear relationships between the input and output layers of an ANN. This process updates the weights and trains the network. The process is repeated for each pattern in the training set until the total output error converges to a minimum, or until some limit is reached. The network is then ready to be tested by inputting new data not used yet to test the accuracy [13].

There are no exact solutions to the numbers of layers and units required for particular applications. The ability to generalize is a function of the numbers of hidden layers, units in the hidden layers, and the training data. In general, if a network has too few hidden units, it cannot effectively learn the training data. Failure to generalize also occurs when too many hidden units are used. Therefore, it is often appropriate to use the minimum size of hidden layers for the required task. Removal of redundant units has no serious ill effects on the network's ability to generalize. For complex mappings, two hidden layers give better generalization performance than a single layer. One method of choosing suitable network architecture is by trial and error [14].

For a network to be able to generalize, it should have fewer parameters than the data points in the training set. It is often desired to use a smaller (but still suitable and effective) network [15].

Trial and error on the selection of network architecture with four inputs and two outputs was conducted [16] with 30 models of varying hidden layers and units. Tests were carried out to obtain the optimal architecture for on-line prediction of surface finish and dimensional deviation in turning with emphasis on obtaining the lowest predicted error and the fastest learning speed. It was found that a 4–3–2 network was optimal for the application.

Back-propagated neural networks are often applied in the field of metal cutting. The effects of thermal induced errors on a two-axis turning lathe were studied by creating a radial-basis-function neural network [12]. This was able to predict the thermally induced errors with an error within 15%.

On-line prediction of surface finish was performed using neural network based sensor fusion [17]. The surface finish was assessed with an error varying from 2% to 15% under different cutting conditions, whereas predicted dimensional deviations predicted varied from -20 to $+20$ μm, with an average error of 6 μm.

The effects of cutting speed, feed rate and depth of cut were studied using HSS tools containing 8% cobalt, and rolled steel bars with 0.35% carbon [18]. Prediction of the surface roughness had -18.21% to $+8.05\%$ errors.

Proper estimation of surface roughness has been the study focus of a number of researchers. However, some have conducted only experimental studies, while others have validated prediction models with limited experiments. The developed models are often valid for a particular domain of parameters, but for different domains of parameters the models have to be redeveloped by means of carrying out a large number of experiments. This limits their application on shop floors.

Surface finish in turning is influenced by many factors, such as feed rate, work materials, unstable built-up edge, cutting speed, depth of cut, tool nose radius, tool angles, machine tools, cutting fluids, etc. It is impossible to consider most of these factors without the help of computers. Current technology is still unable to reach the goal of performing conclusive prediction of surface roughness using artificial intelligence.

In this study, seven factors contributing to surface finish in turning were investigated. A neural network model was established to predict surface roughness R_a and R_t given a set of cutting conditions. Turning was performed under varying cutting conditions. To ensure an accurate prediction using the neural network, 304 sets of data were collected. Of these, 274 data were used for training of the network, and 30 were used to test the accuracy of the network.

2 Turning experiments

Aluminum and copper were chosen as the work materials for the turning experiments. Rods of 19 mm diameter were turned using

Table 1. Tool inserts used

Tool insert	Tool nose radius (mm)	Rake angle (deg)
TiAlN coated carbide 1	0.8	+5
TiAlN coated carbide 2	0.4	+5
TiAlN coated carbide 3	0.2	+5
TiAlN coated carbide 4	0.8	+15
TiAlN coated carbide 5	0.4	+15
PCD	0.8	0
SCD	0.8	0

a lathe. Five sets of titanium aluminum nitride (TiAlN) coated carbide inserts with three tool nose radii and two rake angles were used to turn the workpieces, and three cutting parameters – depth of cut, feed rate, and spindle rate – were varied. PCD and SCD inserts were also used for mirror surface finishing. The tool inserts used are shown in Table 1.

The cutting parameters selected for the rough turning throughout the experiments were depth of cut = 1 mm, spindle rate = 1000 rpm, and feed rate = 0.1 mm/rev. The cutting parameters selected for the finishing turning varied as follows: (depth of cut = 0.005–0.5 mm, spindle rate = 500–2500 rpm, and feed rate = 0.001–0.2 mm/rev). As such, when the datasets were presented to the neural network, there was a high chance of predicting accurate surface roughness over a wide range of cutting parameters.

Surface topography can be presented by various two-dimensional (2D) and three-dimensional (3D) surface roughness parameters, many of which are redundant [19]. It may be also characterized using the fractal concept; but this method is a new approach and its effectiveness has not completely established [20]. The surface roughness heights (parameters) considered in this study are two 2D surface roughness parameters: arithmetic average surface roughness (R_a) and the maximum peak to valley height (R_t). These parameters were selected because they are the most commonly used. In total, 304 turning experiments and measurements of surface roughness were carried out, and therefore, 304 datasets were collected.

3 Prediction of surface roughness

3.1 Neural networks

The popular multi-layer architecture of feed-forward neural networks was employed to produce a system that was able to predict two surface roughness parameters given the seven cutting parameters shown in Fig. 1. This architecture has been proven to be an excellent universal approximation of non-linear functions. Its ability to map complex input-to-output relationships with an acceptable error best demonstrates its suitability.

We collected 304 sets of data in the turning experiments, and measurements of the surface roughness were scaled into the range [−1, 1]. 30 datasets were picked randomly to form testing data to check the network accuracy, and the remaining 274 sets were allocated to training data for training the network.

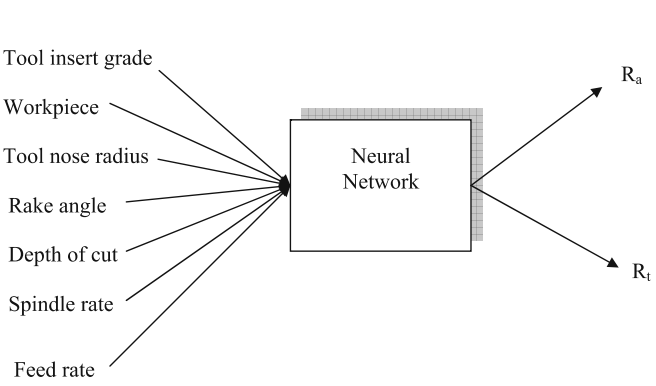

Fig. 1. Inputs and outputs of the network

Table 2. Networks generated for ANOVA and the least absolute percentage errors

Network architecture	Testing		Training	
	R_a error (%)	R_t error (%)	R_a error (%)	R_t error (%)
Hyperbolic tangent function				
7–3–2	45.47	55.57	44.02	57.48
7–4–2	48.75	61.52	45.81	64.29
7–10–2	48.08	31.22	38.12	57.73
7–15–2	43.79	50.50	40.44	52.37
7–3–6–2	34.59	51.71	39.97	50.10
7–4–8–2	41.87	34.67	33.20	45.78
7–10–20–2	21.18	18.73	11.14	15.98
7–15–30–2	25.81	21.54	7.31	11.15
Sigmoid function				
7–3–2	58.67	63.04	47.88	68.09
7–4–2	40.58	60.46	44.42	64.63
7–10–2	47.24	61.77	44.64	65.90
7–15–2	28.63	53.75	31.60	47.99
7–3–6–2	55.42	55.21	46.41	56.56
7–4–8–2	32.81	41.41	33.56	48.78
7–10–20–2	24.52	20.65	11.03	16.06
7–15–30–2	20.10	18.45	8.07	12.04

Variables in the neural network, such as activation functions, and the number of hidden layers and neurons, had to be chosen and applied before the network could be generalized. Three- and four-layer networks were considered. Using the Widrow Rule of Thumb [14], and the Theorem of Kolmogorov [16], the number of hidden neurons was calculated to be between 3.04 and 15. Thus, the number of neurons in the first hidden layer used in an analysis of variance (ANOVA) was set at 3, 4, 10, and 15. For the four-layer networks, because we had no idea of the exact number of hidden neurons in the second hidden layer, it was set to twice the number of neurons in the first hidden layer for simplicity. A total of 16 networks were generated and are shown in Table 2, which also displays the least absolute percentage errors obtained from individual networks for ANOVA.

3.2 Results of ANOVA

ANOVA was performed to evaluate the significance of the effects of three- and four-layer networks with sigmoid and hyper-

bolic tangent activation functions. The ANOVA revealed that the network layers were highly significant. The four-layer network produced less percentage errors compared to the three-layer networks. The need for a four-layer network was due to the complexity of the relationships between the input and output layers. A three-layer network tended to have a minimum percentage error (30%–70%). No matter how many hidden neurons were in the hidden layer, a three-layer network could not learn the relationships between the input and output layers accurately. With an extra layer of hidden neurons, the calculation of the relationships could be further disintegrated to that extra layer, improving the ability of the network to learn precisely.

However, the ANOVA revealed that the choice of activation functions and the joint effect of three- and four-layers and the activation functions were not significant. Hyperbolic tangent and sigmoid activation functions produced comparable results. This means that the squashing of the outputs of individual layers that resulted in ranges of [0, 1] or [−1, 1] did not make any considerable differences. Thus, either one would be suitable for this application.

3.3 Effect of the number of hidden neurons

Trial networks were further generated to obtain the optimal number of neurons in the hidden layers. The two activation functions were used separately to generate two sets of R_a-prediction-error plots, as shown in Figs. 2 and 3, where the number of the hidden neurons in the first layer is plotted in the x-axis, and the number of the hidden neurons in the second layer is plotted in the y-axis.

As shown in Figs. 2 and 3, the minimum prediction percentage error could be obtained using a neural network with 10–15 neurons in the first hidden layer and 15–20 neurons in the second hidden layer. Therefore, more networks were generated in these ranges using both sigmoid and hyperbolic tangent activation functions to search for the optimal network. The results obtained were tabulated in Table 3. It was found that a 7–14–18–2 network using the hyperbolic tangent activation function produced the minimum prediction percentage error.

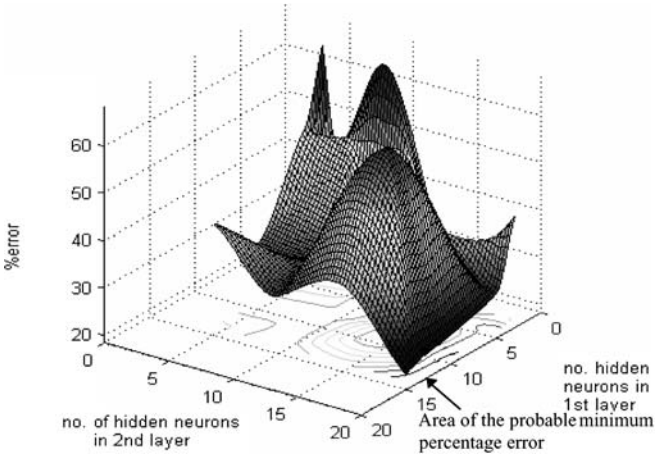

Fig. 2. Percentage errors of predicting of R_a using neural networks with the sigmoid activation function

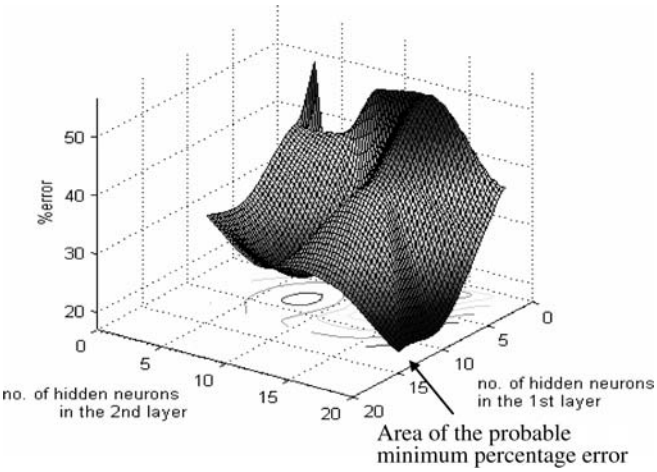

Fig. 3. Percentage errors of predicting of R_a using neural networks with the hyperbolic tangent activation function

Table 3. Prediction percentage errors of R_a and R_t using 4-layer neural networks

Network architecture	Testing		Training	
	R_a error (%)	R_t error (%)	R_a error (%)	R_t error (%)
Hyperbolic tangent function				
7–10–18–2	30.86	41.01	11.90	18.59
7–11–16–2	22.20	22.22	12.73	16.61
7–12–18–2	19.80	19.03	11.62	13.89
7–13–16–2	34.18	28.52	13.32	18.40
7–13–17–2	26.22	26.37	11.11	15.64
7–14–16–2	30.06	26.30	11.31	14.40
7–14–18–2	15.27	18.63	11.62	13.89
Sigmoid function				
7–10–18–2	35.12	30.17	12.59	17.08
7–11–16–2	25.81	26.26	11.50	17.19
7–12–18–2	40.39	34.24	13.16	16.22
7–13–16–2	23.95	27.93	14.70	19.15
7–13–17–2	18.42	19.24	10.28	13.18
7–13–18–2	36.09	39.62	10.94	14.13
7–14–16–2	22.14	19.89	10.11	12.66
7–14–18–2	29.56	22.54	9.40	13.25

3.4 Prediction of surface roughness using the 7–14–18–2 network

Table 3 and Figs. 4 and 5 illustrate that the 7–14–18–2 network was capable of predicting the surface roughness R_a and R_t with an average error of 15.27% and 18.63%, respectively. The theoretically estimated R_a values were calculated using the 30 sets of testing data and Eq. 1. The results are shown in Fig. 5. Here, the measured values obtained using the surface roughness tester can be compared to the predicted values obtained using the network.

As shown in Fig. 5, the theoretical equation did not provide a good estimation of R_a values. Theoretically, R_a approaches

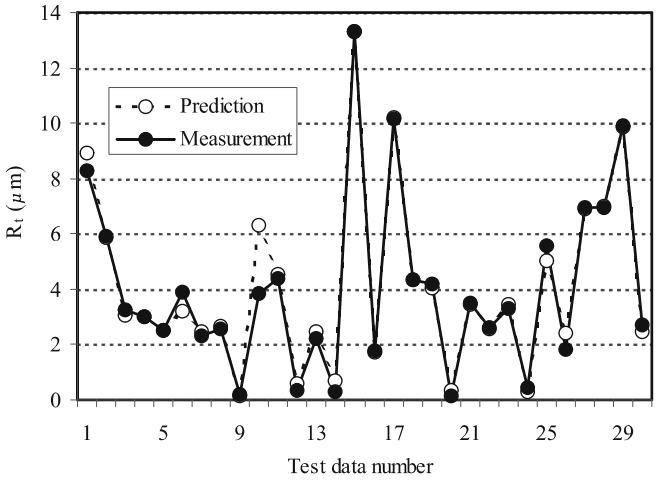

Fig. 4. Comparison of R_t values obtained from actual measurements, and prediction using the network

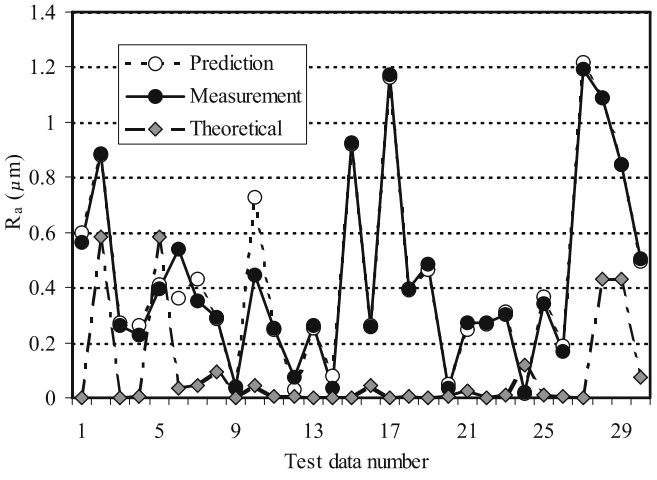

Fig. 5. Comparison of R_a values obtained from actual measurements, theoretical calculation, and prediction using the network. Average absolute % error of the prediction = 15.27%; average absolute % error of the theoretical calculation = 100.81%

zero when feed rate approaches zero. But in reality, this is not possible due to other factors, such as depth of cut, tool insert geometry and material, cutting speed, sharpness of the cutting edge, vibration and accuracy of the machine tool, which are not included in the theoretical equation. On the other hand, the network produced an average of 15.27% error, and Fig. 5 clearly shows that the network was capable of predicting R_a values. In fact, the network is about seven times more accurate than the theoretical equation. This proves that neural networks are able to provide accurate estimations of the surface roughness values, given the necessary input.

3.5 Reliability of the network

The reliability of the network in predicting surface roughness was investigated. 18 sets of surface roughness R_a and R_t values were predicted using the network, and 18 turning experiments were carried out on a smaller tabletop lathe. The surface roughness heights (R_a and R_t) of the 18 new samples turned were measured and compared with the predicted surface roughness R_a and R_t values.

Table 4. Comparison of predicted and measured roughness values of samples turned on the second lathe

Predicted R_a (μm)	Measured R_a (μm)	Prediction Error (%)	Predicted R_t (μm)	Measured R_t (μm)	Prediction Error (%)
0.0422	0.0681	38.10	0.2412	0.3597	32.94
0.0425	0.0725	41.39	0.1937	0.4658	58.43
0.0522	0.0719	27.36	0.2824	0.5257	46.28
0.0422	0.0544	22.48	0.2412	0.4191	42.44
0.0413	0.0716	42.29	0.3559	0.3711	4.09
0.0449	0.0548	18.05	0.3408	0.3206	6.31
0.0379	0.0489	22.43	0.1671	0.2729	38.75
0.0411	0.0467	11.85	0.2000	0.3032	34.04
0.0495	0.0643	23.07	0.2865	0.2919	1.86
0.0558	0.0586	4.82	0.2009	0.2924	31.31
0.0530	0.0691	23.34	0.3232	0.4863	33.53
0.0370	0.0378	2.17	0.1427	0.3069	53.51
0.0598	0.0553	8.26	0.1621	0.2748	40.99
0.0592	0.0911	35.03	0.2043	0.4751	56.99
0.0569	0.0642	11.35	0.3251	0.2769	17.43
0.0410	0.0696	41.07	0.1773	0.3883	54.32
0.0436	0.0917	52.48	0.2036	0.4318	52.84
0.0495	0.0571	13.30	0.2672	0.3092	13.58

Table 4 demonstrates the network's ability to predict surface roughness parameters R_a and R_t with small errors. It had an average absolute percentage error of 24.38% and 34.42% with respect to surface roughness parameters R_a and R_t, respectively. Compared to the network's performance with respect to the testing data obtained from the first lathe, it showed a slightly higher error in both R_a and R_t values. This was due to the difference between the two machines used. The first lathe used was a larger, precision machine commonly used for production. The second machine used was a tabletop machine used for training and education. It was not as rigid as the first machine and had more vibrations and other errors. Therefore, the predicted roughness values of the samples turned on the second machine had more errors. However, this difference in errors can be deemed to be small. As such, the network can be said to be effective in predicting the surface roughness produced on different machines.

4 Conclusion

We predicted the surface roughness heights R_a and R_t of turned surfaces using networks with seven inputs. These included tool insert grade, workpiece material, tool nose radius, rake angle, depth of cut, spindle rate, and feed rate. Coated carbide, PCD and SCD inserts were used to conduct 304 turning experiments, and surface roughness heights of the turned surfaces were measured. A systematic approach to obtain an optimal network was employed, and the effects of the network architecture and hyperbolic tangent and sigmoid activation functions on the accuracy

of the neural network were considered. The reliability of the optimized neural network was further explored by predicting the roughness of surfaces turned on another lathe. The results proved that the network was able to predict the R_a and R_t values of the surfaces machined on this lathe as well.

References

1. Krar SF, Oswald JW (1990) Technology of machine tools, 4th edn. McGraw-Hill, New York
2. Groover MP (2002) Fundamentals of modern manufacturing. Wiley, New York
3. Whitehouse DJ (1994) Handbook of surface metrology. Institute of Physics Publishing, Bristol, UK
4. Hung NP, Zhong ZW, Lee KK, Chai CF (1999) Precision grinding and facing of copper-beryllium alloys. Precis Eng 23(4):293–304
5. Cheung CF, Lee WB (2000) Study of factors affecting the surface quality in ultra-precision diamond turning. Mater Manuf Process 15(4): 481–502
6. Mital A, Mehta M (1988) Surface finish prediction models for fine turning. Int J Prod Res 26(12):1861–1876
7. Zhong Z, Hung NP (2000) Diamond turning and grinding of aluminum-based metal matrix composites. Mater Manuf Process 15(6):853–865
8. Trent EM, Wright PK (2000) Metal cutting, 4th edn. Butterworth-Heinemann, Boston
9. Rahman M, Seah KHW, Goh TN, Lee CH (1992) A study on the appropriate usage of different tool materials. J Mater Process Technol 29:119–132
10. Bryan J (1990) International status of thermal error research. Ann CIRP 39(2):645–65
11. Li X (2001) Real-time prediction of workpiece errors for a CNC turning centre, part 2, modeling and estimation of thermally induced errors. Int J Adv Manuf Technol 17(9):654–658
12. Davalo E, Naim P (1991) Neural networks, 1st edn. Macmillan, London
13. Aleksander I, Morton H (1990) An introduction to neural computing. Chapman and Hall, London
14. Patterson DW (1996) Artificial neural networks: theory and applications. Prentice-Hall, New York
15. Hagan MT, Demuth HB, Beale M (1996) Neural network design. PWS Publishing Company, Boston
16. Azouzi R, Guilloti M (1997) On-line prediction of surface finish and dimensional deviation in turning using neural networks based sensor fusion. Int J Mach Tools Manuf 37(9):1201–1217
17. Mackay D (2003) Bayesian methods for neural networks:theory and Applications. Available online at http://www.inference.phy.cam.ac.uk/mackay/BayesNets.html, accessed December 2003
18. Risbood KA (2003) Prediction of surface roughness and dimensional deviation by measuring cutting forces and vibrations in turning process. J Mater Process Technol 132:203–214
19. Zhong ZW, Lu YG (2002) 3D characterization of super-smooth surfaces of diamond turned OFHC copper mirrors. Mater Manuf Process 17(2):269–280
20. Zhong ZW, Lu YG (2003) Fractal roughness structure of diamond-turned copper mirrors. Mater Manuf Process 18(2):219–227