

Modelling Dental Milling Process with Machine Learning-Based Regression Algorithms

Konrad Jackowski, Dariusz Jankowski, Héctor Quintián,
Emilio Corchado and Michał Woźniak

Abstract Control of dental milling processes is a task which can significantly reduce production costs due to possible savings in time. Appropriate setup of production parameters can be done in a course of optimisation aiming at minimising selected objective function, e.g. time. Nonetheless, the main obstacle here is lack of explicitly defined objective functions, while model of relationship between the parameters and outputs (such as costs or time) is not known. Therefore, the model must be discovered in advance to use it for optimisation. Machine learning algorithms serve this purpose perfectly. There are plethoras of competing methods and the question is which shall be selected. In this paper, we present results of extensive investigation on this question. We evaluated several well-known classical regression algorithms, ensemble approaches and feature selection techniques in order to find the best model for dental milling model.

Keywords Dental milling process · Machine learning · Regression · Ensemble of predictors · Feature selection

K. Jackowski (✉) · D. Jankowski · M. Woźniak
Department of Systems and Computer Networks, Wrocław University
of Technology, Wyb. Wyspińskiego 27, 50-370 Wrocław, Poland
e-mail: konrad.jackowski@pwr.edu.pl

D. Jankowski
e-mail: dariusz.jankowski@pwr.edu.pl

H. Quintián · E. Corchado
Department. de Informática y Automática, Universidad de Salamanca,
Salamanca, Spain
e-mail: hector.quintian@usal.es

E. Corchado
e-mail: escorchado@usal.es

1 Introduction

Over recent years, there have been a high increase in the use of artificial intelligence and soft computing methods to solve real-world problems [5, 7, 16].

It is known that the complexity inherited in most of the new real-world problems increases with the computing capabilities [18]. Higher performance requirements with a lower amount of data examples are needed due to the costs of gathering new instances, especially in those processes where new technology arises.

The optimisation process of machine parameters could significantly help to increase companies' efficiencies and substantially contributes to cost reductions in preparation and setting machines processes and it also helps in the production process using new materials. Nevertheless, the variables and parameters setting processes are a well-known problem that has not been fully resolved yet. Several different techniques are proposed in the literature. In [8] the influence of operating parameters of ultrasonic machining is studied using Taguchi and F-test method.

Also, ANN has been used to find relationships between the mechanical properties of different real-world problems [2, 7] and they have been also applied for identification of the parameters for operating conditions [10]. Conventional methods can be greatly improved through the application of soft computing techniques [6], by generating soft computing models with high accuracy that are used to optimise machine parameters. This research is focused on creating the best model of a dental milling machine applying several soft computing techniques, comparing all these techniques in terms of RMSE (Root Mean Square Error).

This study is a consequence of a preliminary study [17], where several models were generated using ANN; however, these models did not get the expected accuracy, so in these research other softcomputing techniques have been applied using a bigger dataset than in previous study, in order to get models with lower errors.

The rest of this paper is organised as follows. Section 2 introduces all the algorithms used in this study for modelling. The experiments and commented results are presented in Sect. 3. Finally, conclusions are drawn and future work outlined.

2 Machine Learning-Based Regression

One of the method for creating model of given process is the application of regression algorithms. Their objective is to create the best possible mapping between inputs and outputs. In regression tasks, it is assumed that the output \hat{y} is the realisation of real variable which depends on input X [1]. Task of finding optimal input–output mapping (1) means that one have to select model of function F , which reflect real relationships between \hat{y} and X , and choose the function parameters β .

$$\hat{y} = F(X, \beta). \quad (1)$$

The task is not easy as there are several issues which have to be addressed. Let us discuss the most important ones. First, in many real situations we do not have any information on model of relation (1) or its parameters. In this case one can select the model arbitrarily in hope that it allows for acceptable approximation of the relation. Alternatively, several models can be chosen and assessed experimentally by comparison of their accuracies achieved on available testing set. Lists of available models are wide. Among the other one can list linear or polynomial regression, non-parametric regression–smoothing algorithms [9]. Alternatively, one can try to employ machine learning algorithms [11], which prove their usefulness for regression and classification tasks. Advantage of this sort of systems is that most of them are very adaptable and can create mapping model which is able to adjust to current tasks in more flexible manner than regression algorithm based on mapping described in previous paragraph. Some of the machine learning-based options are described later on in Sect. 1. Second, input variables can have different types and formats, which causes problems with their unification. Nonetheless, in many cases, their transformation into numerical values can be easily done. Some problems can appear while casting nominal variables. Usually, this process requires designing tailored distance measures which are very often used by machine learning algorithms. Nonetheless, without losing ability to generalisation, in further consideration we assume that input to the system consists of the set of numerical variables only. Next, regardless which model of F is selected, its parameters have to be adjusted to form appropriate mapping. For this purpose a learning set is used (2). It consists of set of sample, i.e. pairs of input and output.

$$LS = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}. \quad (2)$$

There is one essential issue related with forming the learning set. Its content has to be representative, i.e. distribution of collected samples should reflect real distribution of population in feature space. It is hard to meet this requirement as it requires collecting relatively large number of samples, what would be costly, time consuming or impractical. Therefore, usually we have in hand only limited learning set and which does not guarantee that resulting regression model approximates the real relationships acceptably. Finally, one has to select a measure for model evaluation. The same measure can be also used for regression algorithm training procedure which aims at setting optimal model parameter values. In case of regression tasks, mean root squared error measure (3) can be used for that purpose:

$$RMSE(F, LS) = \sqrt{\frac{1}{N} \sum_{n=1}^N (F(x_n) - y_n)^2}. \quad (3)$$

Formula (3) can be also used for comparative analysis of different regression algorithms. Due to lack of knowledge on model of dental milling process we decided to test possible wide range of regression algorithms. For our purposes we decided to select several classical regression algorithms and some ensemble methods.

2.1 Classical Algorithms

2.1.1 Linear Regression

This is supposedly the simplest predictive model [1]; nonetheless, it features surprisingly high accuracy in many practical problems. It uses simple weighted combination of input variables to calculate the output. The weights are set in a course of training procedure aiming at minimising RMSE. More advanced version of linear regression named Pace regression was proposed by Wang [19].

2.1.2 Multilayer Neural Network

This is the implementation of classical multilayer neural network [1] with nodes arranged in sequentially organised layers. Layers are fully interconnected and input signal is transferred forward from the first to the last layer. Each node performs weighted fusion of its input signals and maps the response with sigmoid transfer function. There is only one output neuron which returns the system response. Entire system is trained using back-propagation algorithm.

2.1.3 SMOreg

SMOreg stands for sequential minimal optimisation for support vector machine (SVM) for regression. SVM is one of the most popular machine learning algorithms proposed by Smola [14]. SVM computes linear regression in a high-dimensional feature space. Input data are mapped using a non-linear kernel function optimisation of the SVM that is performed by algorithm which is an extension of the SMO algorithm proposed by Platt [12].

2.2 Feature Selection

In many real situations, results obtained by classical algorithms are not satisfactory or even acceptable. Usually, it is hard to find the reason, as there are many factors which affect the results. One can list wrong selection of regression model, and poor representation of samples in learning set. There are few options how to improve the performance. One of them is feature selection. It is based on the assumptions that there can exist input variables which are not relevant to outputs. Such inputs presented to the system can negatively affect overall performance.

Removing such an irrelevant or noised feature can improve the accuracy. There are many possible ways how to do that. The optimal one is exhaustive selection, i.e. comparative evaluation of all possible combinations of features. This method

is feasible only for limited number of inputs due to computational complexity. More effective (from this point of view) are heuristic algorithms such as backward or forward feature selection, or selection based on genetic approach.

2.3 Ensemble Techniques

The last presented approach to improve regression accuracy covers a wide range of ensemble techniques. All of them share the same principle, i.e. it is assumed that there is a set $\Pi = \{F_1, F_1, \dots, F_K\}$ which consists of several elementary (classical) regression algorithms F_k (predictors).

The predictors in Π are trained independently. Ensemble system fuses responses of all elementary predictors while the fusion function varies depending on the selected model. In the most simplest form result can be obtained by calculating the average of responses (4)

$$\hat{F}(x) = \frac{1}{K} \sum_{k=1}^K F_k(x). \quad (4)$$

The model (4) is very intuitive and can be calculated straightforward as there are no additional parameters which have to be computed in a course of ensemble learning procedure. On the other hand, this model does not take into consideration the quality of the elementary predictors. It means that all the predictors contribute to final response in the same degree which can lead to spoiling results by weak or irrelevant predictors. Much better results can be obtained by incorporating weighting which reflects the quality of the predictors (5):

$$\hat{F}(x) = \frac{1}{K} \sum_{k=1}^K w_k F_k(x). \quad (5)$$

While selecting model (5) one must remember that $\sum_{k=1}^K w_k = 1$. Important issue that must be addressed is the method of weight selection. In the simplest implementation, the weights shall be counter proportional to RMSE of the given elementary predictor. In more sophisticated case, they can be adjusted in the course of learning process which aims at minimising RMSE of the ensemble according to (3).

2.3.1 Diversity of the Ensemble

Collecting elementary predictors which shall form ensemble system is not a trivial task. Common sense tells that set of the same or similar predictors cannot help to improve quality of regression as all the responses are almost the same. Therefore, there is a necessity to inject some diversity into ensemble. There are several methods possible to do this. According to Brown et al. [4] diversity can be ensured using

implicitly or explicitly diversity maintaining algorithms. In the following paragraphs, we provide some insight into selected diversity maintaining methodologies.

2.3.2 Bagging

Bagging was originally proposed by Breiman [3]. In this method ensemble consists of predictors of the same type, i.e. having the same mapping model. Diversity is generated by partitioning empirical dataset used for training individual predictors in the ensemble. Each of them is trained with its own subset created using bootstrapping, i.e. drawing sample randomly from original dataset with replacement.

2.3.3 Boosting

More sophisticated approach is used in boosting introduced by Schapire [13]. Forming homogeneous ensemble is an iterative process. At each repetition recently created predictor is weighted according to its accuracy. The better one gets higher weights in the ensemble. Samples in learning set are also weighted proportionally to accuracy of their estimation. Smaller prediction error means that weights of sample are decreased in next iteration.

2.3.4 Heterogeneous Ensemble

The last ensemble method used in our researches is creating heterogeneous ensemble, i.e. such that consists of predictors which have different mapping models. There are many techniques which can be used for setting fusion weights (5). In our tests we decided to use evolutionary algorithms for that purpose. In this method, weights are encoded in a form of chromosome (i.e. vector containing weights stacked together). Setting their values is optimisation process which aims at minimising regression error (3). Population of possible solution (weights values) are processed using standard genetic operators (i.e. mutation and crossover), which introduce some diversity into population of solutions and exchange the parts of chromosomes between selected individuals.

3 Experimental Evaluation

We set the following objectives for our research and experiments:

1. Examining wide range of classical machine learning-based regression algorithms,
2. Investigating possibility of increasing of the regression by application of ensemble methods such as bagging, boosting and heterogeneous ensembles.

3.1 *Experimental Framework*

The following regression algorithms were implemented and tested:

1. Linear Regression (*Linear Reg.*),
2. Pace Regression (*Pace Reg.*),
3. Multilayer Neural Network with one hidden layer consisting of 5 neurons (*MLP Reg.*),
4. SVM for regression optimised with SMO algorithm (*SMO Reg.*).

Additionally, several listed below techniques were also tested.

1. Feature selection algorithm which uses genetic algorithms. Methods were applied to all elementary predictors.
2. Homogeneous ensembles. Four ensembles were created based on four listed classical predictors, respectively. Two ensemble creating techniques were implemented:
 - a. Bagging,
 - b. Boosting.
3. Heterogeneous ensemble which consisted of four different classical predictors. Three methods were tested for calculating output of the system:
 - a. Simple mean of predictors outputs,
 - b. Weighted mean with weights set proportional to predictor accuracy,
 - c. Weighted mean with weights set by evolutionary algorithm which minimised RMSE of the ensemble.

Classical predictors, bagging and boosting algorithms were modelled in KNIME (an open source data mining framework [15] available at¹) using WEKA nodes.² MATLAB optimisation toolbox was for our own implementation of evolutionary ensemble.

3.2 *Dataset*

In this paper we use empirical material collected during manufacturing of dental pieces (see Fig. 1). A dynamic high-precision machining centre with five axes was applied in this research. This real industrial use case is described by a dataset of 218 samples obtained by a dental scanner in the manufacturing of dental pieces characterised by 12 input variables. The input variables (see Table 1) are the following: the type of work, the thickness, the number of pieces, the radius of the tool, the revolutions of the drill, the feed rate in each of the dimensions (X, Y and Z), the initial tool

¹<https://www.knime.org/>.

²<http://www.cs.waikato.ac.nz/ml/weka/>.

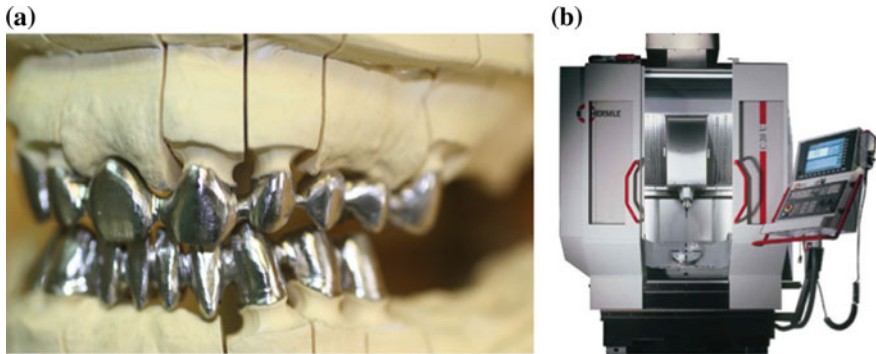


Fig. 1 Metal pieces drilled by the drilling machine. **a** Metal pieces manufactured by a dynamic high-precision machining centre with five axes. **b** Milling centre of HERMLE type-C 20 U (iTNC 530)

Table 1 Different features from the process, their units and ranges

Variable (Units)	Number of feature	Range of values
INPUTS		
Type of work	1	1–7
Thickness (mm)	2	8–18
Number of pieces	3	1–4
Radius (mm)	4	0.15–2
Tool	5	1–4: toric, spherical, plain, drill
Revolutions per minute (RPM)	6	7,500–39,000
X-axis feed rate (mm per minute)	7	0–3,000
Y-axis feed rate (mm per minute)	8	0–3,000
Z-axis feed rate (mm per minute)	9	50–2,000
Initial diameter tool (mm)	10	91.0608–125.56
Initial temperature (°C)	11	21.6–31
Estimate work time (s)	12	6–6,318
OUTPUTS		
Time error for manufacturing (s)	1	–682–1017

diameter, the initial temperature and the estimated duration of the work. The main parameter to estimate is the time error for manufacturing, which is the difference between the estimated time by the machine itself and real work time—negative values indicate that real exceeds estimated time. All the variables and their ranges are presented in Table 1.

3.3 Results and Discussion

Table 2 presents results obtained in test procedure. For elevating test confidence, 10-fold cross-validation methods were used. Therefore, the table shows average of RMSE (3) obtained for 10 repetitions of the experiments for 10-folds, respectively. Target signal was normalised before tests to get a value from range $< 0, 1 >$. Analysis of the results from Table 2 allows to make the following observations:

1. All elementary predictors obtained very similar results. The only one that stands out is SMO algorithm. Three other got almost the same RMSE. Therefore, it is hard to firmly state that any one can be nominated as the winner.
2. Feature selection did not bring any advantage as the results for respective algorithms remain also almost the same.
3. Bagging and Boosting did not help neither. Only small improvement can be noticed for SMO predictor as its RMSE was reduced from 0.000977 to 0.000447 by application of boosting techniques.
4. No improvement was get by ensemble techniques which use heterogeneous set of predictors. However, it is worthy to mention that simple average, which is probably one of the most popular techniques, spoiled the quality of the system as it increased the error to 0.007888. More advanced weighted fusion methods preserve the same quality. Nonetheless, as it can be seen, they are useless, as there is no error reduction.

Finally, no one ensemble technique allows to bring any significant error deduction. All tested systems got very similar results. Therefore, it can be clearly seen that there are cases for which classical regression algorithms are the best options and their improvement cannot be done by more advanced and sophisticated ensemble techniques. Nonetheless, it has to be underlined that this conclusion cannot be generalised.

Table 2 RMSE for tested regression algorithms

		Linear Reg.	MLP Reg.	Pace Reg.	SMO Reg.
Single one predictor	All inputs	0.000425	0.000424	0.000430	0.000977
	Selected features	0.000428	0.000428	0.000426	0.000949
Homogeneous ensembles	Bagging	0.000419	0.000420	0.000421	0.000693
	Boosting	0.000425	0.000411	0.000430	0.000447
Heterogeneous ensembles	Mean fusion	0.007888			
	Weighted fusion	0.000434			
	Evolutionary based fusion	0.002079			

Authors are conscious that for any other cases results could be quite different. Even more, we would like to underline that this conclusion cannot be done without making all the reported tests.

4 Conclusion

Generation of dental milling models with high accuracy allows to optimise machine parameters and a reduction in costs and time. The extensive investigation done in this paper about the modelling of dental drilling machine let us to conclude that all classical regression techniques used in this paper are the best options for modelling the dataset used in this research.

Results obtained with the soft computing techniques applied in this paper are better than results obtained in previous studies; however, the reason is not the usage of more complex modelling techniques. Based on the current results and previous one obtained in other studies, it can be concluded that the dataset used in previous studies was smaller and not so much informative than this one; so for this reason results are better using simpler softcomputing techniques.

Future lines of research include modelling the temperature difference and the erosion difference (between diameters of the tool before and after the manufacturing), which helps to measure the accuracy of the dental milling process. Additionally, it will investigate about optimisation algorithms in order to optimise the dental milling machine parameters.

Acknowledgments This work was supported by the Polish National Science Centre under the grant no. DEC-2013/09/B/ST6/02264.

References

1. Alpaydin, E.: Introduction to Machine Learning, 2nd edn. The MIT Press, Boston (2010)
2. Bosch, J., López, G., Batlles, F.: Daily solar irradiation estimation over a mountainous area using artificial neural networks. *Renew. Energy* **33**(7), 1622–1628 (2008). <http://www.sciencedirect.com/science/article/pii/S0960148107002881>
3. Breiman, L.: Bagging predictors. *Mach. Learn. Boston* **24**, 123–140 (1996)
4. Brown, G., Wyatt, J., Harris, R., Yao, X.: Diversity creation methods: a survey and categorisation. *J. Inf. Fusion* **6**, 5–20 (2005)
5. Calvo-Rolle, J.L., Casteleiro-Roca, J.L., Quintián-Pardo, H., del Carmen Meizoso-Lopez, M.: A hybrid intelligent system for PID controller using in a steel rolling process. *Expert Syst. Appl.* **40**(13), 5188–5196 (2013). <http://dx.doi.org/10.1016/j.eswa.2013.03.013>
6. Chang, H.H., Chen, Y.K.: Neuro-genetic approach to optimize parameter design of dynamic multiresponse experiments. *Appl. Soft Comput.* **11**(1), 436–442 (2011). <http://www.sciencedirect.com/science/article/pii/S1568494609002567>
7. Diz, M.L.B., Baruque, B., Corchado, E., Bajo, J., Corchado, J.M.: Hybrid neural intelligent system to predict business failure in small-to-medium-size enterprises. *Int. J. Neural Syst.* **21**(4), 277–296 (2011). <http://dx.doi.org/10.1142/S0129065711002833>

8. Kumar, V., Khamba, J.S.: Statistical analysis of experimental parameters in ultrasonic machining of tungsten carbide using the taguchi approach. *J. Am. Ceram. Soc.* **91**(1), 92–96 (2008). <http://dx.doi.org/10.1111/j.1551-2916.2007.02107.x>
9. Liero, H., Härdle, W.: Applied nonparametric regression (Biometric society monographs no. 19) Cambridge University Press 1990, p. 333. *Biom. J.* **33**(6), 704–704 (1991). <http://dx.doi.org/10.1002/bimj.4710330610>
10. Liu, Y.H., Liu, C.L., Huang, J.W., Chen, J.H.: Neural-network-based maximum power point tracking methods for photovoltaic systems operating under fast changing environments. *Sol. Energy* **89**(0), 42–53 (2013). <http://www.sciencedirect.com/science/article/pii/S0038092X12004082>
11. Mitchell, T.M.: *Machine Learning*, 1st edn. McGraw-Hill Inc, New York (1997)
12. Platt, J.: Probabilistic outputs for support vector machines and comparison to regularize likelihood methods. In: Smola, A., Bartlett, P., Schoelkopf, B., Schuurmans, D. (eds.) *Advances in Large Margin Classifiers*, pp. 61–74. MIT Press, Cambridge (2000). <http://citeseer.ist.psu.edu/platt99probabilistic.html>
13. Schapire, R.E.: The boosting approach to machine learning: an overview. In: *Proceedings of the MSRI Workshop on Nonlinear Estimation and Classification* (2001)
14. Scholkopf, B., Smola, A.J.: *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge (2001)
15. Silipo, R., Mazanetz, M.P.: *The KNIME Cookbook: Recipes for the Advanced User*. KNIME Press, Switzerland (2012)
16. Torreglosa, J., Jurado, F., García, P., Fernández, L.: PEM fuel cell modeling using system identification methods for urban transportation applications. *Int. J. Hydrog. Energy* **36**(13), 7628–7640 (2011). <http://www.sciencedirect.com/science/article/pii/S0360319911007233> hysy-days
17. Vera, V., Corchado, E., Redondo, R., Sedano, J., Garcia, A.E.: Applying soft computing techniques to optimise a dental milling process. *Neurocomputing* **109**, 94–104 (2013). <http://dx.doi.org/10.1016/j.neucom.2012.04.033>
18. Villar, J.R., González, S., Sedano, J., Corchado, E., Puigpinós, L., Ciurana, J.: Meta-heuristic improvements applied for steel sheet incremental cold shaping. *Memetic Computing* **4**(4), 249–261 (2012). <http://dx.doi.org/10.1007/s12293-012-0100-4>
19. Wang, Y., Witten, I.H.: Modeling for optimal probability prediction. In: *Proceedings of the Nineteenth International Conference in Machine Learning*, pp. 650–657. Sydney, Australia (2002)