Viharos, Zs. J.; Kis K. B.: Increasing the artificial neural network based model building speed for supporting efficient technical diagnostics, 12th IMEKO TC10 Workshop on Technical Diagnostics, New Perspectives in Measurements, Tools and Techniques for Industrial Applications, Florence, Italy, June 6 - 7, 2013, pp. 90-95.

Increasing the artificial neural network based model building speed for supporting efficient technical diagnostics

Dr. Zsolt János VIHAROS[1], Krisztián Balázs KIS[1]

[1] *Computer and Automation Research Institute of the Hungarian Academy of Sciences, MTA SZTAKI, H-1111, Budapest, Kende u. 13-17., Phone: +36 1 279 6 195, Fax: +36 1 4 667 503, viharos.zsolt@sztaki.mta.hu, kis.krisztian@sztaki.mta.hu*

Abstract **-** The paper introduces an algorithm for increasing the speed for building up the general system model on Artificial Neural Network (ANN) basis. It finds that input/output configuration of the field analysed which realizes the most accurate estimation and explores the maximal number of dependencies among the related system parameters. The availability and the estimation capabilities of the needed system models are especially important for technical diagnostics in order to be able in differentiating between conform and non-conform situations. The performance of the novel solution is tested and evaluated under a field specific analysis applying the classical equations from the cutting theory. Experiments were done also for cutting diagnosis based on real measured parameters under varying conditions. These validations showed good empirical performance and practical applicability of the algorithm introduced. As result, the proposed algorithm increased significantly the speed of the model building stage.

I. Introduction

Reliable process models are extremely important in different fields of computer integrated manufacturing [1]. Difficulties in modelling manufacturing processes are manifold: the great number of different machining operations, multidimensional, non-linear, stochastic nature of machining, partially understood relations between parameters, lack of reliable data, etc. A number of reasons back the required models: the design, optimization, control and simulation of processes and the design of equipment. However, in spite of progress being made in fundamental process modelling, accurate models are not yet available for many manufacturing processes. Heuristic models are usually based on the rules of thumb gained from experience, and used for qualitative evaluation of decisions. Empirical models derived from experimental data still play a major role in manufacturing process-modelling [2].

The input-output configuration strongly influences the accuracy of the developed model especially if dependencies between parameters are non-invertable. In various stages of production (e.g. in planning, optimisation or control) different tasks arise; consequently, the estimation capabilities of the related applied models are different even if the same set of parameters is used. One of the main goals of the research to be reported here was to find a general model for a set of assignments, which can satisfy accuracy requirements.

A similar algorithm was already developed by the authors using ANNs as models [3, 4] but the necessity of technical diagnostics to increase the model building speed and the idea to introduce flexible learning step for the competitive, alternative system models drove us to redefine and adapt the original input-output search algorithm to a more flexible and quicker one. The introduction of this adaptation and the related test results in the field of technical diagnostics are the main aims of the paper.

II. Building up the general system model for technical diagnostics: input-output search

A lot of effort has been made to apply ANNs for modelling manufacturing operations (Monostori, et al., 1996). The assignments to be performed determined the input-output configurations of the models, i.e. the parameters to be considered as inputs and the ones as outputs. Considering the input and output variables of a given task together as a set of parameters, the ANN model estimates a part of this parameter set based on the remaining part. This partitioning strongly influences the accuracy of the developed model especially if dependencies between parameters are non-invertable. In different stages of production (e.g. in planning, optimisation, control or technical diagnostics) tasks are different, consequently, the estimation capabilities of the related applied models are different even if the same set of parameters is used. It should be stressed that in a classical application the input-output configuration of the applied model is determined by the given assignment, namely known parameters serves as inputs and unknown parameters serves as outputs. The estimation capabilities of the applied

ANN models are determined as results after the model building and testing stage. The breakthrough of the original, ANN based algorithm was not to determine the input-output configuration before the model building stage but the model building and training algorithm has to define the appropriate and best input-output configuration of the model automatically (Viharos, et al., 1999). By building up of this general model the algorithm does not have any regard to the given assignment of engineers its target is to satisfy accuracy requirements and build up the most useful system model, e.g. for control aspects. The automatic generation of the general process model, i.e. model, which is expected to work accurately enough in different assignments, consists of the following steps:

- Determination of the (maximum) number of output parameters (No) from the available N parameters which can be estimated using the remaining Ni = N -No input parameters within the prescribed accuracy.
- Ordering of the available parameters into input and output parameter sets having Ni and No elements, respectively.
- Training the network whose input-output configuration has been determined in the preceding steps.

For the realization of these steps a search algorithm combined with the ANN training is needed to select all the possible outputs from the given set of parameters with regard to the accuracy demands. From the user point of view the learning data set is given in the form of N dimensional vectors.

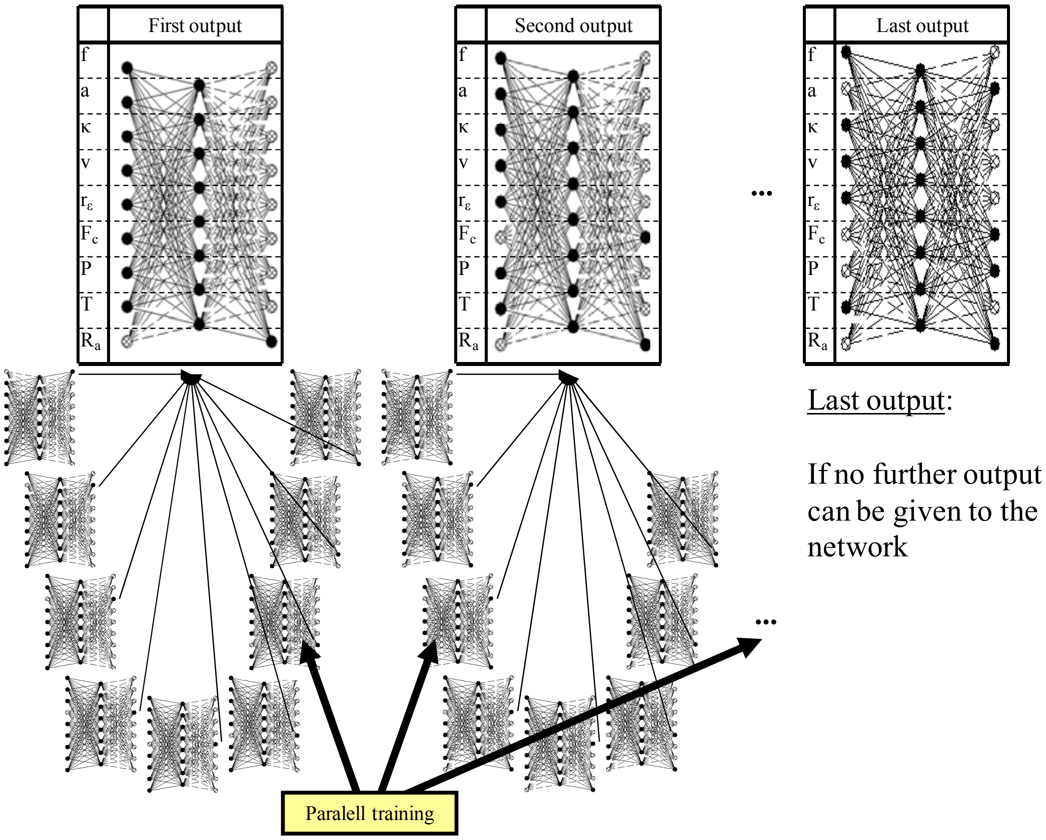

Figure 1. Model input-output search with parallel training

First, the search algorithm chooses only one parameter form the N parameters to be output of the model (Figure 1.). To select the first output parameter, N ANNs are generated, each having one different output and N-1 input parameters. *After generating the ANNs, learning begins by all ANNs, concurrently*. First, *each ANN performs M learning step*. The evaluation follows for checking whether the ANN with the smallest estimation error had reached the required estimation accuracy. If not, another learning phase is started with M epoch again. If yes, then this means that an output was found which can be estimated with the given accuracy based on the remaining input parameters. The next step of the algorithms is to order this variable to the output set of parameters and to select a further output parameter. This selection is realised by the same method as for the previous output(s). For searching the second output, N-1 ANNs are generated because one output is already fixed, consequently, there are N-1 different possibilities to add another output to the set of output parameters. The remaining N-2 parameters are used as inputs. After finding the second output, two outputs are fixed and a search starts to find a third output etc. This indicates that for adding a new output to the set of output parameters successful learning, consequently, finished model building is required. Learning is successful if an ANN configuration can learn the dependencies between input and output variables with a given accuracy. The algorithm stops if after a large number of learning steps, none of the ANNs, being in their learning stages can achieve the given accuracy. During this search algorithm the largest number of outputs can be found, the accuracy demands are satisfied and the ANN model is built up. It is called as the general model of the system because it is built up without considering any assignments related to the parameters analysed. Another algorithm was built up to solve various assignments using the same general model [4], e.g. different diagnosis assignments can be solved applying the same model.

A. Improvement of the model search algorithm through adaptive training steps

The input-output search method described in the previous paragraph realizes a search method where an addition of a new parameter to the set of output variables is based on a parallel training of a certain number of artificial neural network models. There are two cases in the training that cannot be avoided but need significant calculation times without improving the final model capabilities:

- In an unsuccessful, final case when no further models can be trained to reach the error limits. From the user point of view the last parallel training epoch is useless for the given application because it does not bring new knowledge to the model. It is only needed to decide that no further outputs can be estimated based on the remaining variables as inputs; consequently, this final step does not improve the incorporated knowledge or structure of the resulted system model. It means also that considering the model building speed and time it would be good to completely eliminate this final epoch.
- In successful, intermediate cases, when a certain number of ANN models are trained in a parallel way to reach the error limits. Only one model will be accepted and used in the next epochs, consequently, the training times of the other models are waste considering the computational and time efforts. It means also that considering the model building speed, it would be good to eliminate the training time of the not accepted models.

Unfortunately, the appointed two cases cannot be fully eliminated because they are need for the running of the algorithm; however, their processing time ratio inside the whole algorithmic run could be as small as possible. This motivation triggered the idea to decrease the processing time of the “waste” calculations. It is allowed by the novel methodology to estimate the error curve of the ANN training and by the introduction of a search algorithm to adapt the training step numbers of the models individually.
The training method of a single ANN model (improved backpropagation, in the current application case [5], however, it can be generalized for other training methods, too) results in usually decreasing model error curve. The analysis of the training curves indicated that they show typical forms, indicating that they can be approximated with an exponential curve based on the recent error values of the training batches. This is based not on the whole training curve but “only” on a certain number of the recent training steps, consequently, it is an adaptive solution during the training. The approximated curve results

- on one hand in an estimation for the *final model error* taking the tangential value of the exponential curve and
- on the other, when having a prescribed maximal allowed error threshold on the required accuracy of the system model, the *number of remaining training steps* can be calculated.

Based on these estimations the following improvements of the original algorithm were introduced:

- *Multistep default (adaptive step number)*: This improvement results changes in the concurrent training of a set of ANNs for finding an additional output of the system model. Previously, all individual ANN received equal calculation time, and except the winner, all the other calculations are waste times. Since there is not knowledge which model will be the winner, consequently, this waste time cannot be fully eliminated but can be decreased through the introduction of an adaptive distribution of training steps among the competitive ANN models: at a given training situation each competitive ANN models receive different numbers of training batches with an inverse relationship to their number of the remaining training steps. The reciprocals of the remaining training steps are calculated and a fixed number of training batches is distributed among the competitive models in proportion with this ratios. In our experimental case, for one model it is allowed to learn maximum 75 training batches (this number was determined as optimal by many experimental runs) but at least one training batch is needed, consequently, in one iteration the models are trained between 1 and 75 training steps where the distribution of the calculation times among the competitive models is proportional to the reciprocals of the remaining training steps of the individuals. Based on the recent behaviour of their training curves the allowed training steps are re-distributed at the next iteration. This adaptive solution enables to find the final system model through less calculation time than with equal distribution of the processing resources among competitive model candidates.

- *All exclude*: Final model error (the error of the output parameter having the worst estimation accuracy) is calculated at each training step according to the methodology above. It is an estimation for the final error of the model at theoretically endless training. If this final model error is above a prescribed threshold through a certain number of training iterations means that the model will never reach the required accuracy. Consequently, it is useless to train this model further, so, the introduced algorithm excludes such models from the further training epochs. As result, much less calculation is needed for the whole model building process.
- *Multistep A** search algorithm: In the next version of the improvements the distribution of the training batches were formed as an A* search algorithm [6] on the following basis:
    - *Search space*: a set of models that are trained in a competitive way as it is represented in the Figure 1. The decision point is at the end of the given training epoch and an ANN model has to be selected that is allowed to train a given number of training steps.
    - *Fitness of one point in the search space*: according to the A* methodology, the fitness of one model is the cumulative learning step that was already performed plus the estimated number of the remaining training steps to reach the prescribed error limits (that is calculated based on the estimated training curve).
    - *Termination criterion*: The algorithm is stopped if one model reaches the prescribed threshold in an intermediate step or at the end of the calculation when all of the competitive model candidates overstep the allowed maximal training step limit.

The *Multistep default (adaptive step number)* component helps to decrease the ratio of waste calculations during the intermediate steps when the search is successful and a new output parameter is found for the model. The *All exclude* solution helps to decrease the required calculations at the intermediate steps and also at the final step when no further outputs can be added to the model built up. The *multistep A** algorithm is an introduction of an appropriate search methodology beyond the heuristics of the *Multistep default (adaptive step number)* and incorporates the above described *All exclude* solution, too. The test results of these solutions for technical diagnosis oriented model building are presented in the next paragraphs.

III. Technical diagnostics oriented tests of the introduced solutions

Developments and trends in control and monitoring of machining processes shows the necessity of sensor integration, sophisticated models, multimodel systems, and learning ability [7]. The performance of the introduced novel solution is tested and evaluated under various conditions: field specific analysis was performed applying the classical equations from the cutting theory and experiments were done also for cutting diagnostics oriented model building based on real measured parameters under varying conditions. These validations showed good empirical performance and practical applicability of the algorithm introduced.

Tests were repeated in same circumstances using the I.: original model search algorithm without any adaptive step number (called as *multistep off*) and the novel solutions: II.: *multistep default*, III.: *all exclude* IV.: *multistep A**.

In the first test a non-linear case was analysed using mathematical equations for data generation. Such models are necessary in diagnosis [8] of manufacturing systems, e.g. to handle the cutting process efficiently. There are many variables describing a metal cutting process, the following parameters were selected for the analysis: Setting of the machine is handled through three machining parameters: depth of cut: a [mm], feed: f [mm/rev], speed: v[m/min]. The tool is presented by three tool parameters: cutting edge angle: χ [rad], corner radius: r_ε [mm], tool life: T [min]. Two monitoring parameters may be used for the turning operation: force: F_c[N] (main force component) and power: P [kW]. The customer demand is determined by the required roughness: R_a [mm]. In cutting theory and practice mathematical equations are applied with these parameters, they are published in catalogues of machine tool producers [9]. Data ranges and the applied equations are described in [4]. Figure 2. represents the required cumulative step numbers for building up the system models per algorithm variant and depending on the relative error limit that was prescribed as minimum model accuracy.

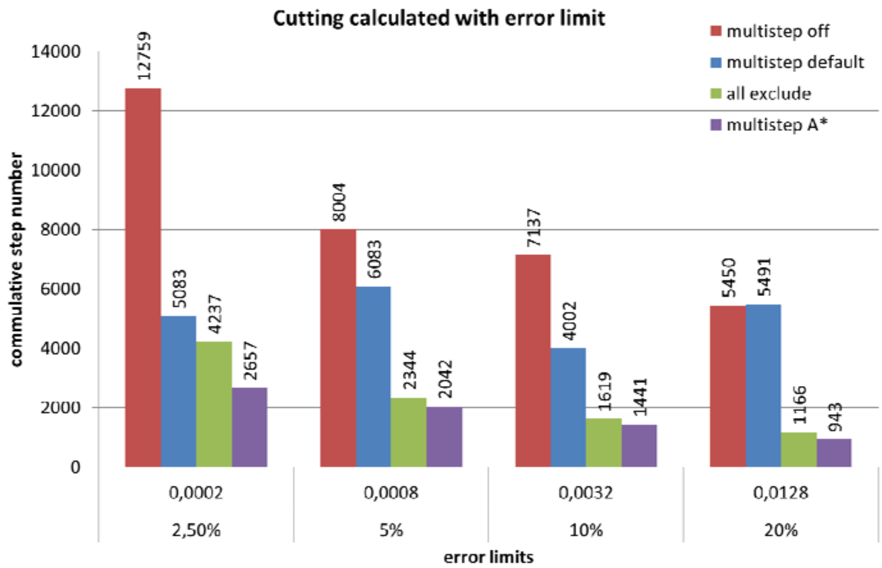

Figure 2. Required cumulative step number of the original and the introduced three novel solutions.

The second test was performed using real, measured data from the turning process. All of the machining parameters were varied and the roughness of the produced surface and the tool wear was measured during these cutting experiments, moreover the temperature (T) of the cutting was measured in process, through changes of the resistance of the machine-tool-material-machine loop. When the tool wear condition reached the wear-out phase, it was replaced by a new tool. Circumstances of cuttings were: material: 42CrMo4, machine: NC, Voest-Alpine, Nr. 085064, type: WNC500S/1, tool: CNMG12040843, cp 3, 1820091, p15, k20, radius: 0.8 mm, without cooling. The speed was varied form 2.12 to 4.89 m/s, the depth of cut form 0.25 to 1.75 mm and the feed from 0.1 to 0.45 mm/revolution. All of the combinations of the parameters were set-up, the parameter ranges was split equally in these intervals. 5 different feed values were selected together with 4 depth of cut and 6 speed set-ups, consequently, data from 120 measurements are available. The data vectors consist of 7 parameters: a f v VB ΔVB Ra T. VB and ΔVB mean the tool wear and the changes in the tool wear during one turning experiment respectively and T is the measured cutting temperature at the edge of the tool. Figure 3. represents the required cumulative step numbers for building up the system models per algorithm variant and depending on the relative error limit that was prescribed as minimum model accuracy.

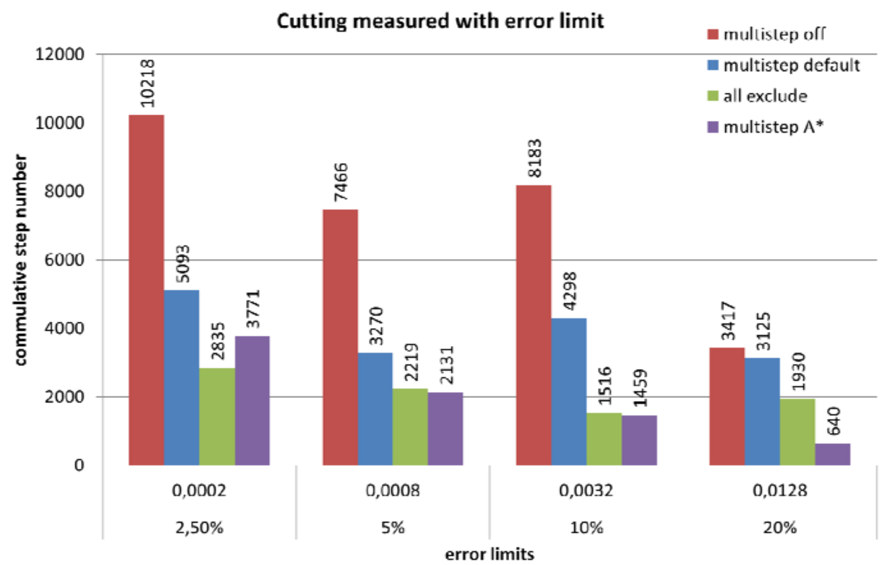

Figure 3. Required cumulative step number of the original and the introduced three novel solutions.

Figure 4. illustrates the average calculation step numbers required to build up the system model in general. Average was calculated on various accuracy thresholds of the required accuracies and on the two analysed cases.

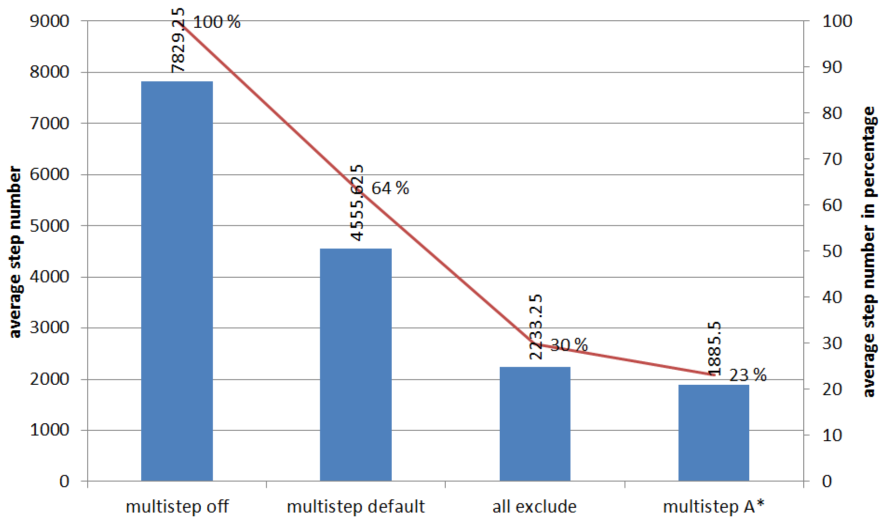

Figure 2. Average training step numbers of the original and the introduced three novel solutions.

The Figure 4. clearly represents that the *multistep default* solution is around 36% quicker than the original one, the *all exclude* methodology is three times quicker, moreover, with the *multistep A** solution the model building time can be decreased to around a quarter in comparison to the original algorithm. Considering the sizes of the problems these improvements result that systems having significantly higher number of parameters can be handled in the same calculation time.

IV. Conclusions

The paper introduced an algorithm for increasing the speed of building up the general system model on ANN basis. It finds that input/output configuration of the field analysed which realizes the most accurate estimation and explores the maximal number of dependencies among the related system parameters. The performance of the solution is tested and evaluated under a field specific analysis applying the classical equations from the cutting theory, moreover, experiments were done on real turning measurements under varying conditions. As result, the proposed algorithm increased the speed of the model building stage significantly; the required model building time can be decreased to around a quarter in comparison to the original algorithm running without adaptive training steps. This is an especially important feature in the field of technical diagnosis where changes and the required related reactions are typically quick and time critical.

References

[1] M. E. Merchant, "An interpretive look at 20th century research on modelling of machining operations", *Proc. of the CIRP International Workshop on Modelling of Machining Operations*, pp. 27–31, 1998.
[2] S. Yerramareddy, S.C. Lu, K.F. Arnold, "Developing empirical models from observational data using artificial neural networks, *Journal of Intelligent Manufacturing*, Vol. 4., pp. 33-41, 1993.
[3] Zs. J. Viharos, L. Monostori, "Automatic input-output configuration of ANN-based process models and its application in machining", In: *Lecture Notes of Artificial Intelligence*, pp. 659-668, 1999.
[4] L. Monostori, Zs. J. Viharos, S. Markos, „Satisfying various requirements in different levels and stages of machining using one general ANN-based process model", *Journal of Materials Processing Technology*, pp. 228-235, 2000.
[5] T. Tollenaere, „SuperSAB: Fast adaptive backpropagation with good scaling properties", *Neural Networks*, Vol. 3, pp. 561-573, 1990.
[6] P.E. Hart, N. J. Nilsson, B. Raphael, „A Formal Basis for the Heuristic Determination of Minimum Cost Paths", *IEEE Transactions on Systems Science and Cybernetics,* Vol 4., No. 2, pp. 100–107, 1968.
[7] H. K. Tönshoff, J. P. Wulsberg, H. J. Kals, W. König, C. A. van Luttervelt, „Developments and trends in monitoring and control of machining processes", *CIRP Annals*, Vol. 37., No. 2., pp. 611-622, 1988.
[8] L. Ciani, M. Catelani, „Photovoltaic Inverter: Thermal Characterization to Identify Critical Components", *International Measurement Confederation (IMEKO), XX. IMEKO World Congress, Metrology for Green Growth*, 9-14. September, Busan, Republic of Korea, paper 534, 2012.
[9] F. Krupp Gmbh, *Widia-Richtwerte für das drehen von Eisenwerkstoffen.* Book, Fried, Krupp Gmbh., Germany, Essen, *(machine tool producer catalogue)*, 1985.