Simulation Reduction Models Approach Using Neural Network

Philippe Thomas, Denise Choffel, André Thomas

Simulation Reduction Models Approach Using Neural Network

Philippe THOMAS, Denise CHOFFEL, André THOMAS
Centre de Recherche en Automatique de Nancy (CRAN-UMR 7039),
Nancy-University, CNRS
ENSTIB 27 rue du Merle Blanc, B.P. 1041
88051 Epinal cedex 9 France
philippe.thomas@cran.uhp-nancy.fr

Abstract

Simulation is often used for the evaluation of a Master Production Schedule (MPS). Also, the goal of this paper is the study of the design of a simulation model by reducing its complexity. According to theory of constraints, we want to build reduced models composed exclusively by bottleneck and, in order to do that, a neural network, particularly a multilayer perceptron, is used. Moreover, the structure of the network is determined by using a pruning procedure. This approach is applied to a sawmill flow shop case.

1. Introduction

The planning or scheduling scenario evaluation by simulation is very useful for the decision makers. Indeed, simulation highlights the evolution of the machines states, the WIP (work in process), and the queues. This information is useful in order to perform a "Predictive scheduling" [13] which is the function that concerns the MPS initially established with the Manufacturing Planning and Control System (MPCS), as opposed to "Reactive scheduling" which gives a new MPS established after significant events occur during the concerned period. Effectively, once the MPS is released, daily events may require it to be brought into question: this is the problem of rescheduling. The real time systems performing manufacturing checks (production reporting) lead to have current follow up information very quickly into the management system [11]. However, it is difficult to use this huge amount of information in order to make decision [19, 20]. At this level of planning, load/capacity equilibrium is obtained via the "management of critical resource capacity" function or

Rough-Cut Capacity Planning (RCCP) which essentially concerns bottlenecks [27]. Goldratt and Cox, in "The Goal" [6] put forward the Theory of Constraints (TOC), which proposes to manage the whole supply chain by bottlenecks control. For this, the use of dynamic discrete events simulation of material flow is helpful [22]. In fact, simulation models of actual industrial cases are often very complex and the modellers encounter problems of scale [17]. Also, many works have highlighted the interest to use simplest (reduced/aggregated) models of simulation [1, 3, 15, 18, 28]. In addition, neural networks have proved there abilities to extract performing models from experimental data [26]. Also the use of neural networks appears recently as an interesting approach within the framework of the supply chain [2, 21].

According to the theory of constraint concepts, the main goal of this paper is to prove the interest of using neural network in order to reduce simulation models useful for the decision making process in re-scheduling problem. In the next part, the proposed approach of reduction model and the multilayer perceptron are presented. The third part will be devoted to the presentation of an industrial application which is a sawmill flow shop case. In order to evaluate the proposition, the obtained reduced simulation model of the sawmill flow shop case is compared to the complete one in the last part.

2. The model reduction

2.1. The algorithm

Amongst various authors, Zeigler was the first to deal with this problem [29]. In his view, the complexity of a model is relative to the number of elements, connections and model calculations. He

distinguished four ways of simplifying a discrete simulation model, in replacing part of the model by a random variable, coarsening the range of values taken by a variable and grouping parts of a model together. Innis et al. [10] first listed 17 simplification techniques for general modelling. Their approach was comprised of four steps: hypotheses (identifying the important parts of the system), formulation (specifying the model), coding (building the model) and experiments. Leachman [12] has proposed a model that considers cycle times in production planning models, especially for the semi-conductor industry which uses cycle time as an indicator. Brooks and Tobias [1] suggest a "simplification of models" approach for those cases where the indicators to be followed are the average throughput rates. Other cases have been studied [8, 9]. The reduction algorithm proposed is an extension of those presented by Thomas and Charpentier [23]:

1. Identifying the structural bottleneck (work center (WC) which for several years has been mainly constrained in capacity).
2. Identifying the conjectural bottleneck for the Manufacturing Order (MO) portfolio of the MPS under consideration.
3. Among the WC not listed in 1 and 2, identify the one (synchronisation WC) satisfying these two conditions:
   - Present at least in one of the MO using a bottleneck,
   - widely used considering the whole MO.
4. If all MO have been considered go to 5 if not go to 3.
5. Use of neural networks for modelled the intervals between the entire WC find during preceding steps.

2.2. The multilayer perceptron (MLP)

The works of Cybenko [4] and Funahashi [5] have proved that a multilayer neural network with only one hidden layer using a sigmoïdal activation function and an output layer using a linear activation function can approximate all non linear functions with the desiderated accuracy. This result explains the great interest of this type of neural network which is called multilayer perceptron. In this research work, our hypothesis lies in the fact that a part of the modelized production system could be approximate by a non linear function obtained thanks to a MLP. The structure of the multilayer perceptron is recalled here. Its architecture is shown in figure 1.

The i-th neuron in the hidden layer (figure 1) receives n_0 inputs $\{x_1^0, \cdots, x_{n_0}^0\}$ with associated weights $\{w_{i1}^0, \cdots, w_{in_0}^0\}$. This neuron first computes the weighted sum of the n_0 inputs:

$$z_i^1 = \sum_{h=1}^{n_0} w_{ih}^1 . x_h^0 + b_i^1 \qquad (1)$$

where b_i^1 is a bias or threshold term. The output of the neuron is given by a so called activation function of the sum in (1):

$$x_i^1 = g(z_i^1). \qquad (2)$$

where g(.) is chosen as an hyperbolic tangent:

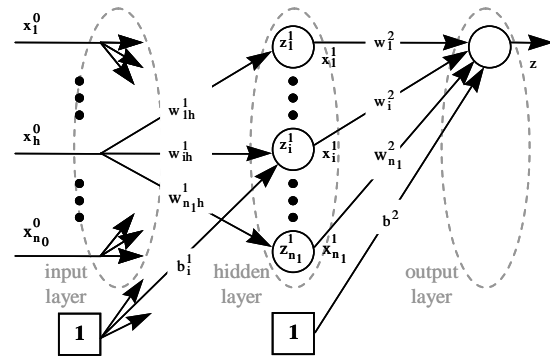$$g(x) = \frac{2}{1+e^{-2x}} - 1 = \frac{1-e^{-2x}}{1+e^{-2x}} . \qquad (3)$$

Figure 1. Architecture of the multilayer perceptron

The neuron in the last layer simply performs the following sum, its activation function being chosen linear:

$$z = \sum_{i=1}^{n_1} w_i^2 . x_i^1 + b^2 \qquad (4)$$

where w_i^2 are the weights connecting the output of the hidden neurons with the output neuron and b^2 is the threshold of the output neuron.

Now, only the number of hidden neurons is always unknown. In order to determine it, the learning starts from an overparametrized structure. A weight elimination method is used to remove spurious parameters. The pruning algorithm used is the Optimal Brain Surgeon (OBS) [7].

The learning of the MLP is performed in three steps:

- Initialisation of the weights and biases of an oversized structure by using the Nguyen Widrow algorithm [16].
- Learning of the parameters by using Levenberg-Marquard algorithm with robust criterion [24].
- Weights elimination by using the Optimal Brain Surgeon algorithm with a robust criterion [25].

3. The overview of the sawmill

At the time of the study, the sawmill SIAT had a capacity of 270.000 m3 / year, a 52 million euros turnover and 300 employees.

The physical industrial production system is composed of sequential work centers (kockums saw, trimmer, sorter,…) and queues or conveyors (RQM4, RQM5, RQM7, …). The log enters the system in RQM1 then is steered to RQM4 or 5 according to the cutting pattern. Following that it passes to the cutting machine (Canter). It then enters the edger. After this phase, the log is transformed into main and secondary products. The final operation is the cross cutting which consists in cutting up products to length.

4. The simulation models

4.1. The complete model

Preceding works [22, 23] have permit to construct the complete model of the sawmill process. This model is presented figure 2. It permits to simulate the log arrival. The second module, "input sorting" permits to drive the log into RQM4 or RQM5 in function of its characteristics. The conveyors RQM4, RQM5 and RQM7 serve of input inventory for the canter line.
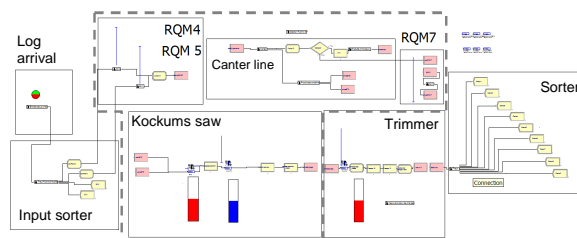

Figure 2. The complete model

Two other modules permit to model the canter line and the passage on the RQM7 conveyor. The modelisation of the canter line needs two submodels which permit to differentiate main and secondary products. Two other modules permit to model the kockums saw and the trimmer.

4.2. The reduced model

As we can see, the design of a complete model for the simulation of a workshop is a difficult task which leads to a complex model. Moreover, the bottleneck of this line is the trimmer [22]. And industrial objectives

are optimising the use of the bottleneck work station in order to maximise its utilisation [6, 14].

Consequently and within this framework, the modelisation of the functioning of inventories RQM4 RQM5 and RQM7, and of the canter line is unnecessary. Also, all the parts surrounded by grey dashed line on the figure 2 give no direct and useful information for the evaluation of a MPS. In fact, only the arrival times of the products in trimmer queue are useful in order to simulate the work load of this bottleneck. And this is why a multilayer perceptron is used to replace all the parts surrounded by grey dashed line on the figure 2. Then the neural network uses the available shop floor information.

For this, we need to collect the available data in input of the process. First, each log holds numerous information which are collected by a scanner in input of the canter line. This information are relative to the product dimension, as length (Lg) and three values for timber diameter (diaPB ; diaGB ; diaMOY). These variables serve to drive the log to RQM4 or RQM5 queues which is additional information (RQM).

In addition of this dimensional information, we have to characterise the process variables at the time of the log arrival. Particularly, the input stock of the trimmer (Q_trim), the utilisation rate of the trimmer (U_trim) and the number of log present in the process between the inputs of RQM4 and RQM5 and the output of the canter line (Q_RQM).

The last type of information is related to the cutting plan of the logs. In fact, each log will be cut into n main or secondary products. In our application, the cutting plan divides the log into 7 products:

- 2 secondary products resulting from the first step of the canter line cutting process on the saw CSMK,
- 2 secondary products resulting from the second step of the canter line cutting process on the saw CSMK after going through on the RQM7 queue,
- 3 main products resulting from the third step of the canter line cutting process on the saw MKV.

These two saws (CSMK and MKV) belong to the canter line. These 7 products can be classified into three categories according to the location (CSMK or MKV) and the time during the cutting process (first or second cutting). This information is given by the variable (T_piece). The last information is the thickness (in mm) of the product which is also the reference. In our case, we are taking into account only two references: main products 75; secondary products 25 (ref).

Consequently, the neural networks input variables are: Lg ; diaGB; diaMOY ; diaPB ; ref ; T_piece ;

Q_trim ; U_trim ; Q_RQM ; RQM. In our application 12775 products are simulated.

Our objective is to estimate the delay (ΔT) corresponding to the throughput time for the 12775 products. ΔT is measured between the process input time and the trimmer queue input time. In practice ΔT is the output of the neural network:

$$\Delta T = \sum_{i=1}^{30} w_i^2 . g\left(\sum_{h=1}^{10} w_{ih}^1 . x_h^0 + b_i^1 \right) + b^2 \qquad (5)$$

The learning of the network is supervised. So, it is necessary to divide the database into two datasets, the learning one and the validation one. The number of hidden neurons should be determined. In order to determine it, the learning starts from an overparametrized structure and a weight elimination method is used to remove spurious parameters [25].
So, the learning begins with a structure using 30 hidden neurons (5) which correspond to 361 parameters. Then, the spurious parameters are removed one after the other and the sum square of the error (SSE) is calculated for the two data sets. Figure 3 presents the SSE for the learning data set (grey circle) and for the validation data set (black cross) in function of the number of parameters remaining.
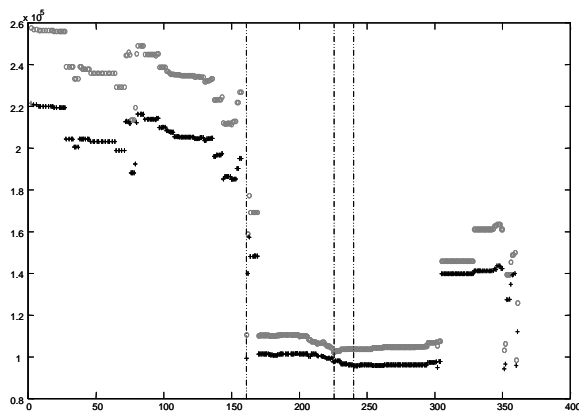

Figure 3. SSE for learning (grey circle) and validation datasets (black cross) function of parameters number

Different choices can be performed concerning the best structure. The structure with 161 parameters is the smallest which satisfies SSE. The structure with 225 parameters gives the lowest SSE for the learning dataset and those with 240 parameters gives the best SSE for the validation dataset. However, these three structures are very close together. And, in these three structures only 4 hidden neurons have been removed. The selected structure, with 240 parameters, uses 26 hidden neurons. It is presented figure 4.
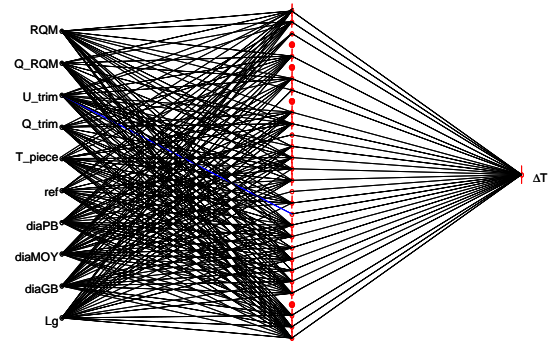

Figure 4. Structure of the network

4.3. The results

Now, the results obtained with the reduced and the complete models will be compared. As explained in the preceding part, the database has been divided into two datasets for learning and validation. The division of the database has been performed randomly and the learning dataset is constituted by 6365 products and the validation dataset by 6410 products.
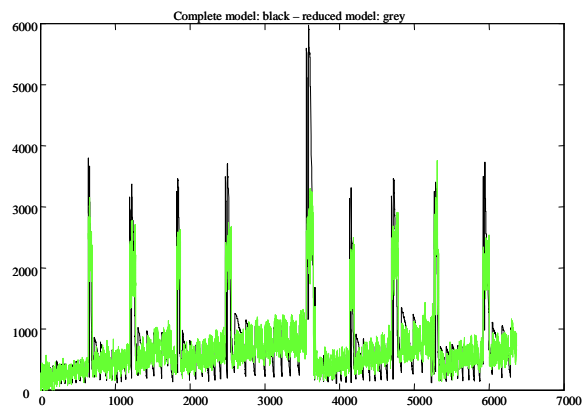

Figure 5. ΔT – learning set

Firstly, the results obtained on the learning data set are presented. Figure 5 shows the delay ΔT (ordinate) used by the 6365 products (abscissa) of the learning dataset. The black line presents the delay for the complete model and the grey line for the reduced model. The times are given in seconds. We can see that the major part of the pieces uses less than 1000 seconds in order to enter in the input stock of the trimmer. However, some of these products use more than 3000 seconds to perform the same course. Figure 5 shows that the times given by the two models are very similar except nine peaks where we can notice important differences. Figure 6 which presents the error performed between the two models corroborate these first remarks. This fact can be explained by the

lack of some variables which can have a particular incidence on the data, as, for example, the percentage of machine-gunned logs, or the repartition of logs into the different RQM queues, …
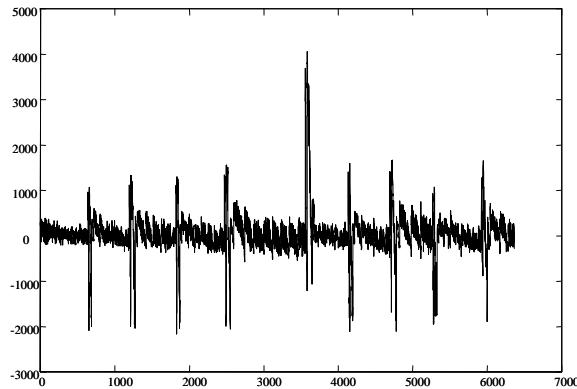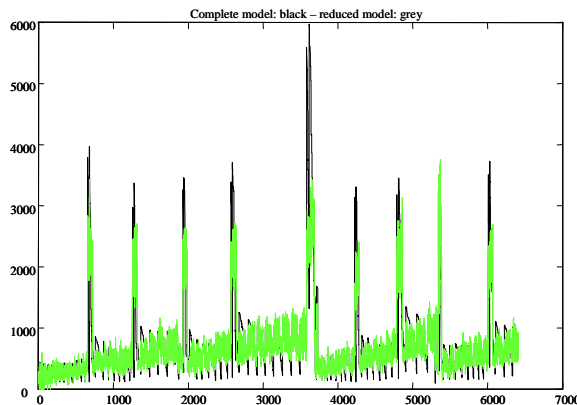

Figure 6. Error – Learning set

Figure 7. ΔT – Validation set

The error mean performed on the learning data set is small (7.9 seconds) but with an important standard deviation (430) which is due to these nine error peaks (figure 6). If these peaks are not considered, the error mean becomes 2.9 seconds. These results obtained on the learning data set prove that the use of a neural network permits to estimate the delays. However, it is necessary to verify that these results can be extended for other data. Figure 7 shows the delay (ordinate) used by the 6410 products (abscissa) of the validation dataset. The black line presents the delay for the complete model and the grey line for the reduced model. The times are given in seconds. These results are very similar to those obtained on the learning dataset. So, no overfitting problem occurs, and the neural network can estimate the delay for other data as those of the learning algorithm.

In the same way, the error mean performed on the validation data set remain small (9.4 seconds) and with an important standard deviation (448) for the same reasons as those mentioned above. If peaks are not considered, the mean error becomes 2.1 seconds.
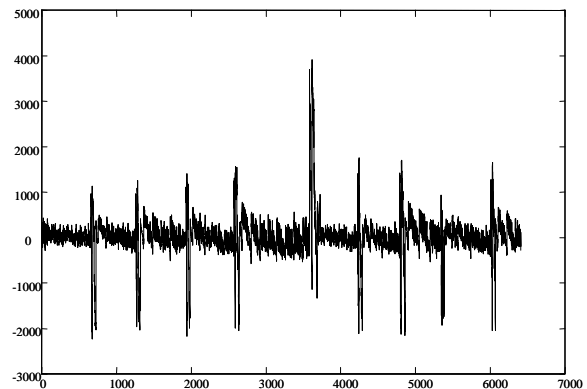

Figure 8. Error – validation set

The figure 8 shows the error performed on the validation data set and confirms the remarks performed during the study of the figure 6.

5. Conclusion

A new approach of simulation model reduction is presented here. This approach uses a neural network, and more particularly, a multilayer perceptron. The aim is to model the functioning of the part of the process which is not constrained in capacity. This approach has been applied to a sawmill flow shop case. The results have shown that:
- the two sets of data present similar results,
- the average of the error (less than 10s) is small according to the process time scale (2000s).

That means that it is very interesting to use neural network to model a part of the process instead of computing with the complete model.It is quicker and easier. It allows the decision making system to focus on the management of the bottlenecks.

Our perspectives are to investigate the structure determination of the network, and particularly, the choice of the inputs of it, and the validation of this approach on different application cases.

6. References

[1] Brooks R.J., and A.M. Tobias, "Simplification in the simulation of manufacturing systems", *Int. J. Prod. Res.,* 38(5), 2000, p. 1009-1027.

[2] Chiu M., and G. Lin., "Collaborative supply chain planning using the artificial neural network approach", *J. of Manufacturing Technology Management,* 15(8), 2004, p. 787-796.

[3] Chwif L., R.J. Paul, and M.R. Pereira Barretto, "Discret event simulation model reduction: A causal approach", *Simulation Modelling Practice and Theory,* 14, 2006, p. 930-944.

[4] Cybenko G, " Approximation by superposition of a sigmoïdal function", *Math. Control Systems Signals,* 2(4), 1989, p. 303-314.

[5] Funahashi K, "On the approximate realisation of continuous mapping by neural networks", *Neural Networks,* 2, 1989, p. 183-192.

[6] Goldratt E., and J. Cox, *The Goal : A process of ongoing improvement*, North River Press; 2ⁿᵈ Revised edition, Great Barrington, USA, 1992.

[7] Hassibi B., and D.G. Stork, "Second order derivatives for network pruning: optimal brain surgeon", *Advances in Neural Information Processing Systems*, S.H. Hanson, J.D. Cowan and C.L. Gilles (Eds.), Morgan Kaufmann, San Mateo, CA, 1993, Vol. 5, pp. 164-171

[8] Hung Y.F., and R.C. Leachman, "Reduced simulation models of wafer fabrication facilities", *Int. J. Prod. Res.,* 37, 1999, p. 2685-2701.

[9] Hwang J.S., S. Hsieh, and H.C. Chou, "A Petri net based structure for AS/RS operation modeling", *Int. J. Prod. Res.,* 36, 1999, p. 3323-3346.

[10] Innis G.S., and E. Rexstad, "Simulation model simplification techniques", *Simulation,* 41, 1983, p. 7-15.

[11] Khouja M., "An aggregate production planning framework for the evaluation of volume flexibility", *Production Planning and Control,* 9(2), 1998, p. 127-137.

[12] Leachman R.C., *Preliminary design and development of a corporate level production planning system for the semi conductor industry*, Eds Optimization in industry, Chichester, UK, 1986.

[13] Lopez P., and F. Roubellat, *Ordonnancement de la production*, Hermès, Paris, 2001.

[14] Marris P., *Le Management par les contraintes en gestion industrielle. Trouver le bon déséquilibre*, Editions d'Organisation, Paris, 1994.

[15] Musselman K.J., "Guideline for simulation project success", *Proc. of the 1993 Winter Simulation Conference,* 1993, p. 58-64.

[16] Nguyen D., and B. Widrow, "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptative weights", *Proc. of the International Joint Conference on Neural Networks IJCNN'90,* 3, 1990, p. 21-26.

[17] Page E.H., D.M. Nicol, O. Balci, R.M. Fujimoto, P.A. Fishwick, P. L'Ecuyer, and R. Smith, "An aggregate production planning framework for the evaluation of volume flexibility" *Proc. of the 1999 Winter Simulation Conference, 1999,* p. 1509-1520.

[18] Pidd M., "Five simple principles of modeling", *Proc. of the 1996 Winter Simulation Conference,* 1996, p. 721-728.

[19] Pritsker A., and K. Snyder, "Simulation for planning and scheduling", *APICS,* August 1994.

[20] Roder P., "Visibility is the key to scheduling success", *APICS Planning and Schedulling,* August, 1994.

[21] Shervais S.; T.T. Shannon, and G.G. Lendaris, "Intelligent supply chain management using adaptive critic learning", *IEEE Trans. On Systems, Man and Cybernetics, Part A Systems and Human,* 33(2), 2003, p. 235-244.

[22] Thomas A., and Charpentier P., "De la pertinence de modèles réduits pour la prise de decision en réordonnancement", *Proc. of the 2ⁿᵈ International Conference on Integrated Design and Production CPI'01,* Fès, Maroc, 2001.

[23] Thomas A., and Charpentier P., "Reducing simulation models for scheduling manufacturing facilities", *European Journal of Operational Research,* 161(1), 2005, p. 111-125.

[24] Thomas P., and G. Bloch, "From batch to recursive outlier-robust identification of non-linear dynamic systems with neural networks", *Proc. of the IEEE Int. Conf. on Neural Networks ICNN'96,* Washington D.C., USA, 1, 1996, p. 178-183.

[25] Thomas P., and G. Bloch, "Robust pruning for multilayer perceptrons", *IMACS/IEEE Multiconference on Computational Engineering in Systems Applications CESA'98,* Nabeul-Hammamet, Tunisia, 1998 p. 17-22.

[26] Thomas P., G. Bloch, F. Sirou, and V. Eustache, "Neural modeling of an induction furnace using robust learning criteria", *J. of Integrated Computer Aided Engineering,* 6(1), 1999, p. 5-23.

[27] VOLLMANN T.E., W.L. BERRY and D.C. WHYBARK, *Manufacturing, Planning and Systems Control*, The Business One Irwin, 1992.

[28] Ward S.C., "Argument for constructively simple models", *J. of the Operational Research Society,* 40(2), 1989, p. 141-153.

[29] Zeigler B.P., *Theory of modelling and simulation*, Wiley, New York, 1976.