A Fuzzy-Logic and Multi-Layer Perceptron Hybrid Model Applied to a Real World Steel-Making Problem

Yannick Lallement (1) and Frédéric Alexandre (2)
(1) Carnegie Mellon University, Human-Computer Interaction Institute
5000 Forbes Avenue, Pittsburgh PA 15213, USA (*yannick@cmu.edu*)
(2) Crin-Inria Lorraine, BP 239, Campus Scientifique
54506 Nancy-Vandoeuvre, France (*falex@loria.fr*)

Abstract: *One of the most interesting properties of neuro-symbolic hybrid models is that they allow for the combination of knowledge-based and data-based approaches. We have used such a strategy in a large, complex steel-making problem. We present this problem together with one of the hybrid models that have been tested on it. We choose to describe this hybrid in details because it is representative of such data-knowledge integration.*

1 Introduction

From 1994 to 1997, the European project MIX has been working on neuro-symbolic hybrid models. Its goal was to better understand this approach, investigate strategies and create tools for a better implementation of neuro-symbolic coupling. This work was assessed on three real world applications, including a steel-making application. Several hybrid models have been tested on this latter application and a combination of these hybrids has been chosen and implemented on-line in the steel plant. In this paper, we first present the application and two previously existing models that implement it. Then we fully describe one specific hybrid model, in order to underline the performance gain due to the hybridization and give an example of successful integration of data-based and knowledge-based approaches.

2 The application

The application we are interested in [6] consists of presetting the *rolling force* that is applied to a sheet of steel in a steel rollmill (Figure 1). The value of that force will determine the characteristics of the sheet when it leaves the rollmill. When the rollmill reaches its full speed, external regulations set the rolling force to its optimal value so that the band has the desired characteristics at the output of the rollmill. The optimal value for a given band can therefore always be known, but only after full speed is reached by the rollmill.

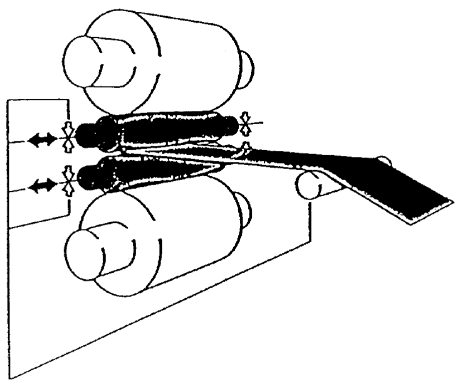

Figure 1: Picture of a steel rollmill

It is necessary to preset the value of the rolling force that will be applied during the transitory period, before full speed is reached. A poor presetting of the force will lead to bad characteristics of the beginning of the steel sheet, that will be wasted, and to unwanted oscillations when the automatic regulations start. For economic reasons, it is important to minimize those two effects; this can be achieved by presetting the rolling force to a value as close as possible to the optimal value at full speed.

This problem has 12 parameters (listed in table 1) related to the input characteristics of the sheet, and its desired output characteristics.

Number of passes	Width	Thickness
Lubrication	Input tension	Output tension
Roll finish	Length	Diameter
Elongation	Temperature	Resistance

Table 1: The 12 parameters of the application

Cases	Atypical domains
1	Pass number 1
2	Width $<$ 800 mm or $>$ 1700 mm
3	Thickness $<$ 0.6 mm or $>$ 2.0 mm
4	Elongation $<$ 0.4% or $>$ 2.0%
5	Resistance $<$ 160 Mpa
6	Input tension stress $<$ 1.6 T
7	Output tension stress $<$ 2.8 T
8	Lubrication type 2
9	Roll finish = 16/17/19/3E
10	Roll finish = 10 without lubrication

Table 2: Atypical domains

3 Previous models

Before designing our hybrid model, we had two different existing models of the application at our disposal: one using a knowledge-based approach (the physical model) and one using a data-based approach (a multi-layer perceptron).

3.1 The physical model

Presetting the rolling force can be done by a model describing the physical processes that take place in the rollmill [8]. This model is a set of equations based on several areas of physics (especially mechanics and thermodynamics). Using 8 of our 12 parameters, the model leads to a value for the rolling force. This value is obtained by following a set of well defined and meaningful steps, as described in [8]; this is clearly a knowledge-based approach to solve the problem.

Unfortunately this model is not very accurate (with an average error rate as defined in section 6.2 of 25%) and its accuracy depends on the region of 12-dimensional space on which it is applied. In order to alleviate this problem, one can divide the space into subspaces in which the error of the physical model is almost constant, and, for each subspace, apply an error-correcting α coefficient to the output of the model to determine the final force:

$$rolling\_force = \alpha \times physical\_model's\_force$$

Under this condition, the physical model can give acceptable results. Unfortunately, dividing the space optimally is a complex problem.

3.2 A multi-layer perceptron

Presetting the rolling force can also be done by a multi-layer perceptron (MLP). A 12x12x6x1 MLP was designed and trained on this task [5]. The inputs of the MLP are the normalized values (in the [0..1] interval) of the parameters listed in table 1. The output of the MLP is the normalized value of the rolling force. The learning algorithm is back-propagation. The MLP has been trained for 1000 epochs using the training set described below. Unlike the physical model, this model does not use any knowledge but only data to achieve its performance.

Previous tests [5] showed that the MLP model was accurate for most data points, but performed poorly on certain specific points. We call a point typical if the MLP gives an accurate reply. If not, the point is called atypical. If a point is in a region of the input space in which the density of examples is high, it is likely to be typical (because the MLP will have learned on a sufficient number of examples in that region), and vice-versa. The atypical points are situated in the atypical domains listed in table 2 [5] and correspond to uncommon customer orders. The atypical domains have been empirically determined by examining the error distribution of the MLP.

4 Architecture of the hybrid system

In order to maximize the performance of the MLP, our hybrid system uses the MLP only to solve typical points. To solve atypical points, the physical model is used. In order to maximize the accuracy of the physical model, we use a fuzzy logic engine to compute the error-correcting α coefficients. This component is based on the FuzzyCLIPS [3] engine. FuzzyCLIPS is an extended version of the CLIPS expert system shell developed by the NASA that allows fuzzy and normal terms to be freely mixed in the rules and facts of an expert system. The rules can be defined by an expert, or learned from a data set, which is the ap-

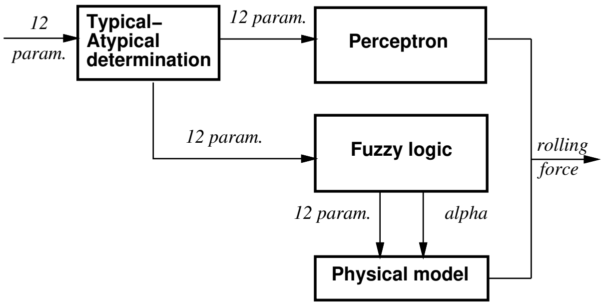

Figure 2: Information flow in the system

proach adopted here. The rules were generated by Iglesias *et al.* [2] using the C4.5 decision tree algorithm [7] and the training set described below. This data-based fuzzy-logic module is used to tune a knowledge-based model.

Once trained, the fuzzy logic engine is fed with the 12 parameters constituting a point, and returns an α value. The fuzzy rules define a precise partition of the input space; each of them has its zone of influence. It can happen that a new point is not in any zone of influence; in this case the fuzzy system will return no α.

To select the model to use for a given point (MLP or physical model tuned by fuzzy logic), we examine whether the point is in a typical domain by looking up table 2. The resolution algorithm is described in the next section.

5 Resolution algorithm

In order to determine the rolling force associated to a data point, the following algorithm is executed (illustrated in figure 2):

- Determine whether the case is typical.
- If the case is typical, it is passed to the MLP, and the MLP's answer is returned to the user.
- If the case is not typical, it is passed to the FL engine. The resulting α is multiplied by the response of the physical model to produce the returned answer.
- If the case is passed to the FL engine, and the FL engine has no rule that applies, then the case is passed to the MLP: it is better to have a degraded solution than no solution at all.

6 Evaluation of the system

6.1 Data sets

We used two data sets: a training set and a test set. The same training set is used both to train the MLP and to generate the fuzzy rules of the FL engine. It contains 5549 normalized vectors of 13 components (the 12 original components and the optimal rolling force). We use the same training set in order to easily compare raw performance of each module to the hybrid system; better performances are expected if we train the MLP only on typical points (the replies should be more accurate) and the FL engine only on atypical points (less rules should be generated). The test set is used to test either only the MLP, or only the physical model tuned by the FL engine, or the hybrid system. It counts 1309 normalized vectors of 13 components.

6.2 Error computation

For each test vector, the error (called *relative error*) is computed by:

$$E_r(n) = \left(\frac{u_d(n) - u(n)}{u_d(n)}\right)^2 \qquad (1)$$

where $u_d(n)$ is the desired rolling force for the band n (known when the rollmill reaches full speed) and $u(n)$ is the rolling force given by the system for this vector. We use this relative error criterion because an error in the estimation of the rolling force in low values causes more characteristic degradations of the band than if the force is high; this error computation reflects economical and quality issues and is therefore more important for the steelmakers than the absolute error.

6.3 Results

Table 3 shows the performance of the system and its components on our test set. The first line presents the results we obtained using only the physical model tuned by the fuzzy logic engine. When the fuzzy logic engine had no applicable rule, we did not include this point for the error computation (this happened in 17 cases, which is why only 1292 are counted in that line). The second line is the same experiment using only the MLP. The third line shows the results obtained with the whole system: in that case, 333 vectors are passed to the physical model tuned by the fuzzy logic engine, and 976 to the MLP.

Model	# vectors passed to FL+physical	# vectors passed to MLP	Relative Error	Standard Deviation
FL+physical alone	1292	0	18.0%	0.145
MLP alone	0	1309	17.7%	0.148
Hybrid system	333	976	15.9%	0.118

Table 3: Performance of the system and its components on the steelmaking application

These results are positive in two different ways: First, the relative error is less for the hybrid model than for each of two single models. The improvement is 1.8 percentage point, representing more than 10%, compared to the best of the two models (the MLP). Second, the standard deviation is better in the hybrid model than for each of the two single models, which is another noteworthy gain brought by hybridization.

The results of the hybrid system are better than the results of each system specifically because the purely data-based MLP shows a better performance when it is used only for typical points; conversely, the performance of the data- and knowledge-based fuzzy logic and physical model combination remains good on atypical points only. The global performance of the system is thus enhanced.

7 Conclusion and ongoing work

In this paper, we chose to concentrate on a specific hybrid model to illustrate the coupling of data- and knowledge-based approaches in a practical case. It is also important to come back to the whole MIX project and to the real-world application.

This hybrid model plus another component evaluating the rolling force directly using fuzzy logic has been implemented on-line in the steel plant. The error rate of the whole system has been measured on-line as low as 12.5%. As a consequence, the steelmakers found that 3,500 tons of steel that used to be returned each year by customers will not be returned any longer because of the higher quality of the steel. A new version of that system, using a Kohonen network [4] to select the model to use instead of a look-up table leads to a 10.5% error rate [1]. An on-line implementation of this new version is currently under study. These hybrid systems are an application of a more general framework for hybridization that we proposed in the MIX project that integrates the notions of knowledge- and data-based approaches, of partial hybrid models and of model selection [1].

Acknowledgments

Thanks to Stéphane Vialle and Prem Janardhan for their helpful comments. This work was partly supported by the Commission of the European Communities under the ESPRIT Basic Research Project MIX 9119. This research was also partly supported by the ONR, Cognitive Science Program, Contract Number N00014-89-J-1975N158. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the ONR or the US Government.

References

[1] F. Alexandre. Tools and experiments for hybrid neuro-symbolic processing. In *International Conference on Tools with Artificial Intelligence*, 1997.

[2] C. A. Iglesias, J. C. González, J. R. Velasco, and K. Eder. Prototype of a fuzzy-neural hybrid model. Tech. report, Universidad Politécnica de Madrid, 1996.

[3] Knowledge Systems Laboratory - Institute for Information Technology - National Research Council Canada. *FuzzyCLIPS 6.02A User's Guide*, 1994.

[4] T. Kohonen. Self-Organizing Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, 43(1), 1982.

[5] N. Pican. Approches statique et dynamique de la modulation de l'efficacité synaptique dans les réseaux de neurones, 1995. Thèse de doctorat de l'Université de Nancy I.

[6] N. Pican, F. Alexandre, P. Bresson, J.P. Haton, and E. Couriot. A optimized backpropagation learning algorithm to preset a temper mill machine: Neuroskin. In *International Conference on Neural Networks and their Applications. NeuroNimes*, 1993.

[7] J.R. Quinlan. *C4.5*. Morgan Kaufmann Publishers, 1992.

[8] W. L. Roberts. An approximate theory of temper rolling. *Iron and steel engineer*, October 1972.