Advances in Engineering Software 42 (2011) 435–443

Contents lists available at ScienceDirect

Advances in Engineering Software

journal homepage: www.elsevier.com/locate/advengsoft

Development of a parsimonious GA–NN ensemble model with a case study for Charpy impact energy prediction

Yong Yao Yang *, Mahdi Mahfouf, George Panoutsos

Institute for Microstructural and Mechanical Process Engineering: The University of Sheffield (IMMPETUS), Department of Automatic Control and Systems Engineering, The University of Sheffield, Mappin Street, Sheffield S1 3JD, UK

ARTICLE INFO

Article history:
Received 5 November 2010
Received in revised form 17 March 2011
Accepted 17 March 2011
Available online 9 April 2011

Keywords:
Neural networks
Genetic algorithms
Ensemble modelling
Charpy impact energy
Heat-treated steel
Model structure optimisation

ABSTRACT

A parsimonious genetic algorithm guided neural network ensemble modelling strategy is presented. Each neural network candidate model to participate in the ensemble model is structurally selected using a genetic algorithm. This provides an effective route to improve the performance of the individual neural network models as compared to more traditional neural network modelling approaches, whereby the neural network structure is selected through some trial-and-error methods or heuristics. The parsimonious neural network ensemble modelling strategy developed in this paper is highly efficient and requires very little extra computation for developing the ensemble model, thus overcoming one of the major known obstacles for developing an ensemble model. The key techniques behind the implementation of the ensemble model, include the formulation of the fitness function, the generation of the qualified neural network candidate models, as well as the specific definitions of the assemble strategies. A case study is presented which exploits a complex industrial data set relating to the Charpy impact energy for heat-treated steels, which was provided by Tata Steel Europe. Modelling results show a significant performance improvement over the previously developed models for the same data set.

1. Introduction

In the steel industry maintaining product consistency along with high quality has become a priority duo to increasing global competition and high customer demand. The accurate achievement of specified mechanical properties is becoming a crucial aspect of many modern steel manufacturers. There is an increased interest in 'right first time' production for critical mechanical and geometric properties through system engineering approaches. Mahfouf et al. [1] pointed out that two components are critical for 'right first time' production design, i.e., a reliable prediction model and an efficient optimisation paradigm. The mechanisms controlling mechanical properties are highly complex, depending on both the steel compositions and its microstructures, such as the grain size, phase fractions and precipitation. These microstructures are strongly influenced by the undergoing metal manufacturing route. As the availability of physical knowledge and understanding of the dynamic behaviours of mechanical properties are often either insufficient or unavailable to formulate a first-principle based physical model, data-driven models, such as artificial neural network and neural–fuzzy models, elicited from carefully assembled process data have become popular and represent ideal candidates [2,3]. The research presented here is focused on the development of reliable prediction models, via the combination of genetic algorithm (GA) guided neural networks (NN) structure optimisation, error back-propagation training for NN parameterisation, and ensemble modelling.

The NN model structure determination still remains an issue in data-driven modelling. It is common practice that the model structure be pre-determined in an ad hoc manner, and is largely based on expert knowledge or via a trial-and-error. While a reasonable model may be obtained at the end of tedious iterative modelling process, such an ad hoc approach often results in the model performance being 'sub-optimal'. To overcome such shortcoming, a GA-based systematic approach [4] is adopted in this paper to unify the NN structure optimisation with the model parameter optimisation. This leads to a more systematic search of the model structure space and a closer integration between the NN structure determination and the NN model parameterisation.

The aim of this present research is to develop an efficient strategy to implement the parsimonious genetic algorithm guided neural network (GA–NN) ensemble model paradigm. The key techniques towards realising such an ensemble modelling scheme include: the effective generation of the GA population (consisting of various NN models with different structures), the identification of candidate neural networks suitable for ensemble, and the

* Corresponding author.
E-mail addresses: y.y.yang@sheffield.ac.uk (Y.Y. Yang), m.mahfouf@sheffield.ac.uk (M. Mahfouf).

0965-9978/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.advengsoft.2011.03.012

development of an appropriate ensemble strategy to aggregate individual candidate neural networks into the final GA–NN ensemble model. Potential benefits from such GA–NN ensemble modelling paradigm include improved model performance in terms of both the prediction accuracy as well as the generalisation properties. The remainder of this paper is organised as follows. In Section 2 a brief description of the GA–NN model optimisation scheme is given. The GA–NN optimisation comprises two levels. At the top level, the NN structure is optimised via a GA-based evolution, while at the low level (after the NN structure and other critical NN model options have been setup), an error back-propagation based algorithm with early stopping mechanism is used to train the NN model parameters. Section 3 is devoted to the algorithmic extension of GA–NN model optimisation for ensemble modelling. A diversity index is introduced to measure the novelty of the individual neural networks generated in each GA generation, this helps to evolve the GA population with good structural diversities as well as a superior fitness as measured by the prediction accuracy. The development of a fitness-based ensemble algorithm is then presented. Section 4 describes a case study of the GA–NN ensemble modelling paradigm to predict the Charpy impact energy for heat-treated steels, with the modelling data assembled from an industrial database pertaining to Tata Steel Europe. Finally, concluding remarks and future work are outlined in Section 5.

2. GA–NN model optimisation

Neural networks and fuzzy logic-based models are two widely used model paradigms for data-driven modelling. In this study, a three-layer feed-forward neural network, as shown in Fig. 1, is adopted as the basic structure for GA-optimised NN modelling. It has previously been proved that such a simple network model possesses very powerful input–output mapping capability, and can approximate any continuous nonlinear function with arbitrary accuracy provided that enough hidden neurons have been used [5]. The output of the neural network can be expressed as follows:

$$y(k) = \sum_{j=1}^{N} w_j \phi_j \left(\sum_{i=1}^{n} w_{ji} x_i(k) + w_{j0} \right) + w_0 \tag{1}$$

where $x(k) = [x_1(k),\ x_2(k), \ldots, x_n(k)]^T$ is the input vector at time instant k, $y(k)$ is the corresponding output, w_j is the weighting coefficient of the jth hidden neuron to the output, w_0 is the output bias, w_{j0} is the input layer bias, w_{ji} is the weighting coefficients between the jth hidden neuron and input x_i with $i = 0$ dedicated for the associated hidden neuron bias, ϕ_j is the nonlinear activate function of the jth hidden neuron, and N is the total number of hidden neurons. It is worth noting that in this research there is no nonlinear activation function for the output neuron, and this is often the case when the neural network is used for regression.

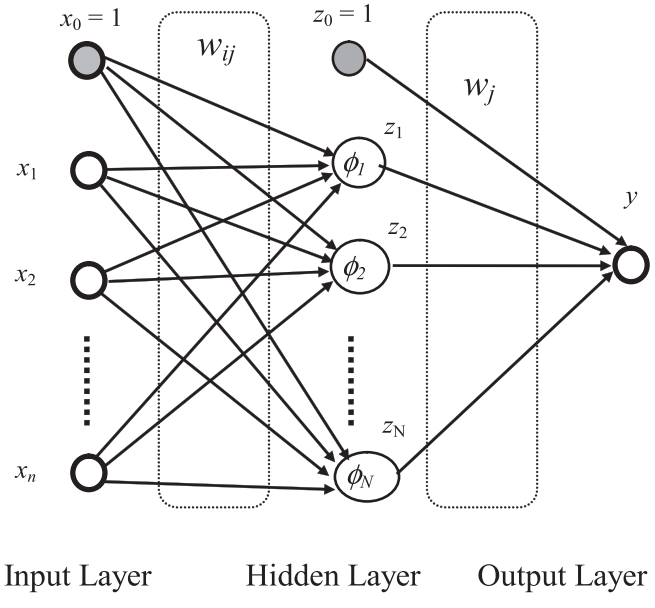

Fig. 1. Feed-forward NN with single hidden layer.

In NN modelling, the determination of the network structure still lacks a more systematic approach, compared to the training algorithms available for determining the weighting coefficients. For a three-layer feed-forward neural network shown in Fig. 1, the most 'critical' structural parameter is the total number of hidden neurons, N. Often, the determination of the total number of hidden neurons relies on individual modellers' preferences and experiences, with the help of some heuristic guidelines. Generally speaking, the optimal value of N should be positively correlated with the complexity of the process and the data to be modelled. While increasing the value of N enhances the neural network's ability of mapping complicated input/output behaviour, there is a risk of over-fitting if N becomes too large. The type of nonlinear activation functions of the hidden neurons also plays an important role towards the input/output behaviour of the NN.

Training algorithms also play a significant role in NN modelling. There exist considerable differences among different training algorithms in terms of algorithm efficiency, computational complexity, memory requirement, etc. The effectiveness and efficiency of different training algorithms are often compounded by the nature of the process to be modelled, the size of the modelling data, the dimension of input/output variables, and the underlying network structure. There exists no single training algorithm which is universally optimal across all attributes. In this research, training algorithms are limited to the category of error back-propagation, such as gradient descent, scaled conjugate gradient, or the Levenberg–Marquardt algorithm [6].

Data pre-processing is an important step [7] in the data modelling process, and normally includes data cleaning, normalisation, transformation, feature selection, etc. Data pre-processing prior to NN training may be desirable in order to simplify the tuning of the training algorithm, speed up the training convergence rate, and avoid training saturation [6]. In this paper, the focus is on the data normalisation, as other aspects of data pre-processing, such as data cleaning and feature selection, have already been dealt with in previous research work. Data normalisation aims at producing a data set where the variations across different inputs are more consistent, so that the numerical analyses in the modelling stage become easier to handle. Often, data normalisation is implemented via some kind of linear or nonlinear data scaling. One common data normalisation option for NN network with sigmoid activation functions is to scale all the inputs and outputs into a range of $[-1, 1]$. Another frequently used option is to transform the inputs/output variables into a new data set such that it has a zero-mean and a unit variance.

An integrated optimisation strategy is first proposed based on genetic algorithms [8,9], to determine the best combination of the neural network structure, the nonlinear activation function for the hidden neurons, the appropriate data pre-processing option, and the training algorithm for learning the weighting coefficients of the selected NN structure. To facilitate the discussion of the GA–NN optimisation strategy development, a GA decision vector $\boldsymbol{\beta}$ is introduced by aggregating all the decision parameters mentioned above, and $\boldsymbol{\beta}$ is given as follows:

$$\boldsymbol{\beta} = [N, F_{act}, T_{alg}, P_{pro}]^T \tag{2}$$