Engineering Applications of Artificial Intelligence 26 (2013) 2270–2286

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Multilayer perceptron for the learning of spatio-temporal dynamics—application in thermal engineering

Matthias De Lozzo[a,b,*], Patricia Klotz[b], Béatrice Laurent[c]

[a] EPSILON Ingénierie, Portes Sud – Bâtiment 3, 12 rue Louis Courtois de Viçose, 31100 Toulouse, France
[b] ONERA, BP74025, 2 Avenue Édouard Belin, 31055 Toulouse Cedex 4, France
[c] Institut de Mathématiques de Toulouse – INSA Toulouse, 135 Avenue de Rangueil, 31077 Toulouse Cedex 4, France

ARTICLE INFO

Article history:
Received 20 December 2012
Received in revised form
28 June 2013
Accepted 1 July 2013
Available online 2 August 2013

Keywords:
Surrogate model
Recursive multilayer perceptron
Spatio-temporal data
Model selection
Penalized criterion
Thermal engineering

ABSTRACT

Thermal engineering deals with the estimation of the temperature at different spatial points and different instants for a given set of boundary and initial conditions. For this purpose, the reference model is a numerical simulation model but it is time-consuming. Consequently we build a surrogate model in order to replace it. This surrogate model is a recursive multilayer perceptron, independent of the boundary conditions and parametrized by the statistical learning of multidimensional temporal trajectories computed with the reference model. It emulates the outputs of the reference model over time from the only knowledge of initial conditions and exogenous variables. Moreover this model is able to predict these outputs in steady state, even if its formulation is time-dependent.

A new methodology is proposed so as to overcome the learning problem associated to the very weak number of trajectories available for the surrogate model construction. The first step attempts to build a more robust surrogate model by considering it as the average of local models resulting from the *V*-folds cross-validation technique. This new kind of multilayer perceptron is much more robust and accurate, in particular when the learning dataset is very small. The second step consists in the creation of a new learning dataset which is made up of each time observation coming from each trajectory. In this way, we artificially obtain a sizeable sample allowing all the classic neural networks constructions.

Furthermore, many approaches exist in order to select the best hidden neurons number but most of them are costly or require a lot of observations. We consider here a non-asymptotic approach based on the minimization of a penalized criterion providing accurate results in an economical computational way. In order to calibrate precisely the penalty term, we use the slope heuristic or the dimension jump, recently introduced in a regression framework. The validation of the method is performed on a toy function.

The prediction ability of the surrogate model built with the new methodology is successfully compared to usual constructions on a simplified problem and then applied to thermal engineering.

1. Introduction

Thermal engineering deals with the estimation of the temperature at some points of interest of a physical system in steady or transient states according to the values of the initial and boundary conditions. The reference model for this problem is a numerical simulation model based upon the spatio-temporal discretization of non-linear physics equations using finite elements or boundary elements methods for example. This model is very time-expensive; consequently it is useful to replace it by a faster one. To this purpose, compact thermal models based upon a thermal–electrical analogy are physical reduced models widely used in thermal analysis (Lasance, 2008) which significantly reduce the computational time. Their parametrization is realized by fitting some observations coming from the reference model which highly depend on specified boundary conditions: these models are "boundary-conditions dependent". In this way, these reduced models quickly estimate the outputs of the reference model but with an accuracy decreasing with the distance to the fitted observations, thus leading to very poor results moving far from them. So, the important CPU time reduction brought by these compact thermal models unfortunately goes along with a sizeable decrease of the global estimation accuracy and new approaches have to be thought.

Therefore, this paper deals with the construction of an accurate surrogate model quickly emulating the outputs of the reference model in transient and steady states for a wide range of boundary

* Corresponding author at: ONERA, BP74025, 2 Avenue Édouard Belin, 31055, Toulouse Cedex 4, France. Tel.: +33 688015388.
E-mail address: matthias.de_lozzo@onera.fr (M. De Lozzo).

0952-1976/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.engappai.2013.07.001

conditions. The surrogate model is usually obtained by the statistical learning of a dataset made up of few observations coming from the reference model and illustrating various boundary conditions. These observations are time-discretized trajectories of the inputs–outputs pair provided by the reference model. The size of the learning sample is very small because of the prohibitive cost of the numerical simulations.

The autoregressive models (AR) are very classical and widely used (Ljung, 1999; Long et al., 2013). These models take in input of their own delayed outputs and have the advantage to be linear. Nevertheless in thermal engineering, temperatures commonly obey to nonlinear laws what can be an obstacle for such a linear model. In a nonlinear context, support vector machine (SVM) (Vapnik, 1998) for regression is a model developed in recent years; this kernel method has been applied with dynamical data (Wang and Meng, 2011). Another kernel method is the Gaussian process model; this is a probabilistic model where the dependences between the observations are specified by the covariance structure of the process (Rasmussen and Williams, 2006). This model has recently been used in dynamical systems identification (Ažman and Kocijan, 2011). It has the advantage of the kernel methods but it cannot be used with high numbers of inputs and observations.

Furthermore many surrogate models based upon a recursive formulation and a neural architecture are commonly met in the literature, e.g. nonlinear time series (Ljung, 1999) and recursive neural networks (Narendra, 1990; Wang and Chin-Teng, 1998). Usually the subjacent function in these models is a multilayer perceptron (MLP) with one hidden layer (Dreyfus, 2005). MLPs have great properties such as parsimoniousness and universal approximation (Bishop, 1995). This kind of neural network has also a global formulation, contrary to a radial basis function (RBF) neural network which has a local one characterized by center-spread pairs (e.g. Li and Zhao, 2006 for an application in identification of dynamical systems). The latter thing is important in our situation where the input space dimension can be high (some dozens of input parameters) and the output to predict has very different behaviors. Moreover, wavelet networks have been developed in a similar manner for nonlinear dynamic input–output modeling (Oussar et al., 1998).

Among these neural network constructions, some of them use uniformly time spaced observations while others require not only the values of exogenous variables and measured outputs at the previous time, but also older versions. Moreover many of these surrogate models are one step-ahead predictors because they need measured outputs; so their use is limited to the next value prediction of an outputs temporal evolution, with applications e.g. in financial forecasting or control system. In our situation, measured outputs are not available and observations are not time equispaced. So, all of these considerations are restrictive for the wished construction and consequently, a new multidimensional temporal multilayer perceptron is proposed in Section 2. The formulation of this estimator is recursive: the predictions it has done at the time just before are considered as inputs with the exogenous variables at the current instant and the time step between the past and the current ones.

The surrogate model should also take into account the small, even very small number of observations coming from the reference model due to its high CPU time. In such a situation, the surrogate model can be very dependent of the learning sample and robustification methods have to be looked for. This is the purpose of Section 3 where two new complementary approaches of construction are detailed. The first one (see Section 3.1) consists in averaging V local recursive multilayer perceptrons resulting from the well-known V-folds cross-validation (Hastie et al., 2001, chapter 7). The second one (see Section 3.2) proposes an innovative point of view for the observations in transient state: from one trajectory measured at K times, K sub-trajectories are extracted, thus artificially increasing the size of the learning dataset. Thence, classical tools from the neural network community can be now considered: indeed if the number of observations is important, building a test sample is possible and consequently, methods such as early-stopping and weight-decay (Bartlett, 1998) can be used. More particularly we use such a test sample in order to select the initial parameters of multilayer perceptrons in multi-start optimization.

In addition, the hidden neurons number in the multilayer perceptron has to be selected, leading to a model selection purpose. This value characterizes the architecture of the multilayer perceptron and so, its complexity. When this complexity is too weak, the ability of the surrogate model to predict accurately the value of the outputs can be very poor. When it is too important, there is a problem of parsimoniousness and the data learning might meet some difficulties. In order to select this number, penalized criteria exist in the statistical community such as AIC (Akaike, 1973) and BIC (Schwarz, 1987) when the number of observations is important. Furthermore, cross-validation and bootstrap error are approximations of the generalization error of the surrogate model and these tools are interesting but costly for model selection (Hastie et al., 2001, chapter 7). The cross-validation error is used in Section 3.1 and Section 4 presents a penalized criterion whose penalty coefficient is data-driven calibrated by means of two different methods based upon a slope heuristic. This last technique has been developed for 10 years now by the statistical community (Barron et al., 1999; Arlot and Massart, 2009; Baudry et al., 2012) and for the best of our knowledge, this is its first application with multilayer perceptrons.

Finally the surrogate model has to be able to deal with stationary situations, that is to say to predict the values of temperatures of interest when the exogenous variables are constant and when the time tends to infinity. Because the steady state corresponds to the phase where the transient phenomena become unchanging with time, we propose in Section 5 to predict the temperatures in steady state via the use of the surrogate models developed in the previous parts for temporal situations. This is possible thanks to the sub-trajectories approach considered in Section 3.2.

1.1. Layout

In Section 2, the physical problem is put into a mathematical framework and a multidimensional temporal multilayer perceptron is presented. In Section 3, new approaches for the construction of the surrogate model and its robustification are described, principally in order to deal with small learning samples. Then a convincing method of model selection based on a penalized criterion is described in Section 4 with an illustration on a toy example. An algorithm is presented in Section 5 so as to consider steady state situations with the models built in Sections 2 and 3. Finally in Section 6, the surrogate model as well as the different methods of construction is validated and compared for a simplified thermal problem; the best approach is applied on an industrial case in Section 7.

2. A multidimensional temporal multilayer perceptron

2.1. The physical problem in a mathematical framework.

Let $\mathbf{y}\in\mathbb{R}^{n_\mathbf{y}}$ be a multidimensional quantity of interest and $\mathbf{y}^k:=\mathbf{y}(t^k)$ its value at time t^k. In the thermal context of this paper, $\mathbf{y}$ merges the values of the temperature at $n_\mathbf{y}$ points of the aeronautical equipment. We assume that $\mathbf{y}^k$ is function of the sets