Neurocomputing 207 (2016) 150–164

Contents lists available at ScienceDirect

Neurocomputing

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

Improvement on the linear and nonlinear auto-regressive model for predicting the NOx emission of diesel engine

Jiaxin Ma [a], Feiyun Xu [a,*], Kai Huang [b], Ren Huang [a]

[a] School of Mechanical Engineering, Southeast University, Nanjing 211189, China
[b] Special Equipment Safety Supervision Inspection Institute of Jiangsu Province, Nanjing 210036, China

A R T I C L E I N F O

Article history:
Received 9 December 2015
Received in revised form
3 March 2016
Accepted 18 March 2016
Communicated by Wei Chiang Hong
Available online 20 May 2016

Keywords:
Linear and nonlinear autoregressive model
NOx emission prediction
Diesel engine
Parameter estimation
Structure identification

A B S T R A C T

Given the increasingly stringent emission regulations, an accurate model of emission prediction is required for the aftertreatment systems of diesel engines. For example, the selective catalytic reduction system can realize higher accuracy emission control if the mass of nitrogen oxides (NOx) is known. Given its simplicity, convenience, and effectiveness, the method of data-driven modeling has been widely researched and considered a primary method to estimate the NOx mass before it reaches the after-treatment device of a diesel engine. To fully use the known engine operating data and therefore improve the prediction accuracy, this study proposes and develops a general linear and nonlinear auto-regressive model with exogenous inputs (GNARX) for NOx prediction. A recursive least square algorithm with forgetting factor is given to estimate the model parameters, and a new simulated annealing based pruning algorithm is developed to identify the model structure. The proposed methods are first used to model the simulation and engineering data to validate their effectiveness and superiority in comparison to the conventional methods. Based on gray relational analysis, the main factors that influence NOx formation, such as the net engine torque, turbo speed, and accelerator pedal position, are then determined as the inputs for modeling the NOx emission of the diesel engine. The results show that the modeling and prediction accuracy of the GNARX model are higher than those of other models, which indicates that the GNARX model is feasible to predict NOx emission.

1. Introduction

Given the growing environmental concerns and increasingly stringent emission regulations, the diesel engines can no longer meet the requirements only by relying on in-engine purification. Studies have shown that it is challenging for diesel engines alone to achieve nitrogen oxide (NOx) emission levels to satisfy the increasingly strict emission standard [1]. Thus, exhaust gas after-treatment techniques are widely used for most diesel engine-powered vehicles [2–4]. The selective catalytic reduction (SCR), one of the most promising aftertreatment techniques, has been shown to be capable of reducing more than 90% of the NOx emission and is chosen as a more feasible solution to the NOx reduction of diesel engines in China [5]. Although the formation of NOx and chemical reaction of SCR have been elaborated in previous studies [6–8], the control of urea injection for SCR remains a great challenge in practice because of its complicated dynamics and limited feedback information. Several approaches to urea injection control have been proposed [9,10], and the on-time adjustment of urea injection based on real-time NOx emission is particularly important in practice.

Several methods can be used to estimate the amount of NOx that reaches the aftertreatment device [11]. (a) The NOx emitted by a reference engine can be directly mapped as a function of rotation speed and torque implemented as a series of look-up tables [12]. (b) A physical-based model, which has been proposed by engine experts, can be used [13]. (c) The NOx emission in the exhaust gases can be directly measured [14]. However, given the complexity of engine behavior, the method of direct engine mapping, which extends the engine-map calibrated under stationary condition to the transients, is usually not accurate enough to estimate the NOx mass. Incorporating experts' deeper knowledge of the physics of the engine and the chemism of emission behavior, physical-based models compensate for the above-mentioned weakness of direct engine maps. As such, physical-based models are widely used [15,16]. Nevertheless, physical-based models usually require huge computational power and significant development time and are very specific (i.e., one model

* Corresponding author.
E-mail address: fyxu@seu.edu.cn (F. Xu).

http://dx.doi.org/10.1016/j.neucom.2016.03.075
0925-2312/

only for one type of engine). For direct NOx measurement, the technology to produce low-cost, precise, and drift-free NOx sensors is still under development [17]. Moreover, literatures have shown that most onboard NOx sensors are significantly cross-sensitive to ammonia [18–20],which poses another difficulty for NOx emission measurement.

Data-driven modeling is another method to predict NOx emission and presents several advantages. First, unlike physical-based models, which depend on expert knowledge of the system, data-driven models require low human intervention because they can be automatically generated from process data [21]. Therefore, compared with physical-based models, data-driven models can save much time and money for model development [22]. More importantly, data-driven models can handle both steady and dynamic data [23]. Given these benefits, data-driven models are widely used in many fields [24–27] and are also successfully applied in the field of diesel engines. Burke et al. [28] used the parametric Volterra series calculated from dynamic measurements to model various gaseous emission species from a multiple-cylinder diesel engine and assessed the predictive performance over the New European Driving Cycle. Antory [29] applied a data-driven monitoring technique to diagnose air leaks in an automotive diesel engine and developed the data-driven diagnostic model using measurement signals taken from sensors in a modern automotive vehicle for condition monitoring purposes. Formentin et al. [30] proposed a data-driven technique to deal with a multivariable fixed-order controller design and validated the effectiveness of the method by numerical comparison with other techniques on a benchmark simulation example and practical test on the airpath control of a real diesel engine. Svard et al. [31] combined a set of general methods for model-based sequential residual generation and data-driven statistical residual evaluation into an automated design methodology and utilized it to create a complete fault detection and isolation system for an automotive diesel engine.

In the present study, a data-driven time series model is considered for NOx emission modeling. To account for the two factors of the NOx emission of diesel engines (i.e., one is the dynamic and nonlinear property of the system, and the other is the knowledge of a part of the system inputs), a general expression for linear and nonlinear auto-regressive model with exogenous inputs (GNARX) is proposed and applied to the prediction of the NOx emission of diesel engine, which has good performance in modeling and predicting nonlinear system data and can fully exploit known information [32,33]. Moreover, the improvement of parameter estimation and structure identification makes the GNARX model superior to other models. Finally, with the properly chosen model inputs, the GNARX model obtains high modeling and prediction accuracy of NOx emission and can therefore provide dependable data feedback for the SCR system to achieve a closed-loop control of urea injection.

The rest of the paper is organized as follows: In Section 2, the model expression, parameter estimation, and structure identification methods of the GNARX model are presented. In Section 3, the simulation and engineering data are applied to verify the superiority of the proposed parameter estimation and structure identification algorithms. In Section 4, the application of the model in predicting practical NOx emission is discussed and compared with several data-driven models to validate its effectiveness. Finally, in Section 5, a summary and some conclusive remarks are provided.

2. Description of the GNARX model

2.1. Model expression

According to the modeling strategy of time series analysis, the general linear and nonlinear auto-regressive model (GNAR) takes a zero mean white noise $\{a_t\}$ as input to the system [33,34]. When one of the exogenous inputs $\{u_t\}$ is known, the GNAR model is convert into the GNARX model with a single exogenous input.

Suppose the system has two exogenous inputs, u_t and v_t, the GNANX model with double inputs is used as shorthand for GNARX (p; τ_u, τ_v; $n_{w,1}$, $n_{w,2}$,…, $n_{w,p}$; $n_{u,1}$, $n_{u,2}$,…, $n_{u,p}$; $n_{v,1}$, $n_{v,2}$,…, $n_{v,p}$), which is expressed as follows:

$$\boldsymbol{x}_{t,i,1} = \{w_{t-1}, \cdots, w_{t-n_{w,i}}, u_{t-\tau_u}, \cdots, u_{t-\tau_u-n_{u,i}+1}, v_{t-\tau_v}, \cdots, v_{t-\tau_v-n_{v,i}+1}\} \tag{1}$$

$$\begin{aligned} w_t = &\sum_{i_1=1}^{n_{w,1}+n_{u,1}+n_{v,1}} \theta(i_1)\boldsymbol{x}_{t,1,1}(i_1) \\ &+ \sum_{i_1=1}^{n_{w,1}+n_{u,1}+n_{v,1}} \sum_{i_2=1}^{n_{w,2}+n_{u,2}+n_{v,2}} \theta(i_1,i_2)\boldsymbol{x}_{t,2,1}(i_1)\boldsymbol{x}_{t,i,1}(i_2) \\ &+\cdots+ \sum_{i_1=1}^{n_{w,1}+n_{u,1}+n_{v,1}} \cdots \sum_{i_p=1}^{n_{w,p}+n_{u,p}+n_{v,p}} \theta(i_1,\cdots,i_p) \prod_{k=1}^{p} \boldsymbol{x}_{t,p,1}(i_k) \\ = &\sum_{j=1}^{p} \sum_{i_1=1}^{n_{w,1}+n_{u,1}+n_{v,1}} \cdots \sum_{i_j=1}^{n_{w,j}+n_{u,j}+n_{v,j}} \theta(i_1,\cdots,i_j) \prod_{k=1}^{j} \boldsymbol{x}_{t,j,1}(i_k) + a_t \end{aligned} \tag{2}$$

where $\boldsymbol{x}_{t,i}$ (i=1, 2,…, p) is the ith-order term; $\boldsymbol{x}_{t,i,j}$ (j=1 ,2,…, i) is the jth-order transitional term in the derivation process of $\boldsymbol{x}_{t,i}$; $\boldsymbol{x}_{t,i,1}(j)$ is the jth element of vector $\boldsymbol{x}_{t,i,1}$; w_{t-i} is the observation at time $t-i$; $u_{t-\tau u-i}$ is the exogenous input u_t at time $t_-\tau_u-i$; $v_{t-\tau v-i}$ is the exogenous input v_t at time $t-\tau_v-i$; a_{t-i} is the white noise at time $t-i$, i=1, 2,…, n; τ_u and τ_v are the input delay of u_t and v_t, respectively; $\theta(i_1)$, $\theta(i_1,i_2)$,… are the model parameters; p is the model order; $n_{w,j}$ (j=1, 2,…, p) is the memory step of the jth-order term of output $\{w_t\}$; $n_{u,j}$ and $n_{v,j}$ (j=1, 2,…, p) are the memory step of the jth-order term of input $\{u_t\}$ and input $\{v_t\}$, respectively.

Similarly, Eq. (2) can also be generalized into multi-input systems, which need not be repeated here.

2.2. Parameter estimation

The forgetting factor recursive least square (FFRLS) is applied to the parameter estimation for the GNARX model [35], which is appropriate for time varying system identification.

Using the GNARX model with double inputs indicated in Eq. (2) as example, the FFRLS algorithm for the parameter estimation of GNARX is deduced as follows:

$$\begin{aligned} &\boldsymbol{x}_{t,i,1} = \{w_{t-1}, \cdots, w_{t-n_{w,i}}, u_{t-\tau_u}, \cdots, u_{t-\tau_u-n_{u,i}+1}, v_{t-\tau_v}, \cdots, v_{t-\tau_v-n_{v,i}+1}\} \\ &\boldsymbol{x}_{t,i,2} = \left\{ \begin{array}{l} \boldsymbol{x}_{t,i,1}(1)\{\boldsymbol{x}_{t,i,1}(1)\}, \boldsymbol{x}_{t,i,1}(2)\{\boldsymbol{x}_{t,i,1}(1), \boldsymbol{x}_{t,i,1}(2)\}, \cdots, \\ \boldsymbol{x}_{t,i,1}(m_{i,1})\boldsymbol{x}_{t,i,1} \end{array} \right\} \\ &\vdots \\ &\boldsymbol{x}_{t,i,i} = \left\{ \begin{array}{l} \boldsymbol{x}_{t,i,i-1}(1)\{\boldsymbol{x}_{t,i,i-1}(1)\}, \boldsymbol{x}_{t,i,i-1}(2)\{\boldsymbol{x}_{t,i,i-1}(1), \boldsymbol{x}_{t,i,i-1}(2)\}, \cdots, \\ \boldsymbol{x}_{t,i,i-1}(m_{i,i-1})\boldsymbol{x}_{t,i,i-1} \end{array} \right\} \end{aligned} \tag{3}$$

where $m_{i,j} = C^j_{n_{w,i}+n_{u,i}+n_{v,i}+j-1}$ (j=1, 2,…, i).

$$\boldsymbol{x}_{t,p} = \boldsymbol{x}_{t,p,p} \tag{4}$$

$$\begin{aligned} &\boldsymbol{x}_t = \{\boldsymbol{x}_{t,1}, \boldsymbol{x}_{t,2}, \cdots, \boldsymbol{x}_{t,p}\} \\ &\boldsymbol{\theta} = \{\theta(i_1), \theta(i_1,i_2), \cdots\}^T \end{aligned} \tag{5}$$