Chemometrics and Intelligent Laboratory Systems 145 (2015) 103–113

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

An adaptive least squares support vector machine model with a novel update for NOx emission prediction

You Lv *, Tingting Yang, Jizhen Liu

The State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Changping District, 102206 Beijing, China

ARTICLE INFO

Article history:
Received 1 January 2015
Received in revised form 16 March 2015
Accepted 10 April 2015
Available online 1 May 2015

Keywords:
Data-driven model
Model update
Least squares support vector machine
NOx emissions
Coal-fired boiler

ABSTRACT

This paper presents an adaptive least squares support vector machine (LSSVM) model with a novel update to tackle process variations. The key idea of the update is to divide the process variations into two main categories, namely, irreversible and reversible variations. Correspondingly, sample addition and sample replacement are proposed to update the model. The incremental LSSVM algorithm and detailed update procedure are also provided. A benchmark simulation with a time-varying nonlinear function is conducted to evaluate the effectiveness of the update algorithm. Finally, the proposed method is applied to predict the nitrogen oxide (NOx) emissions of a coal-fired boiler using real operation data from a power plant. Results reveal that the LSSVM model with the novel update maintains high prediction accuracy despite different process characteristics. Meanwhile, the time consumed in the update process is decreased because of the incremental form compared with the model reconstruction.

1. Introduction

In industrial processes, certain types of primary variables in industrial processes, such as product qualities and flue gas concentration, should be measured accurately and reliably at all times. These variables may be required to be maintained within specified limitations in accordance with government regulation and manufacturing criteria, besides, they are also very important in guiding optimal operation [1,2]. However, the accurate measurement of primary variables is hindered by high costs and technical limitations. Although online analyzers are available in a number of plants, these hardware-based instruments are not only highly vulnerable to failure because of being operated under harsh environments but are also expensive and difficult to maintain. Therefore, a new way of realizing a redundant measurement has crucial significance to ensure the safe, economical and efficient operation of plants [3,4].

Soft sensor techniques are widely accepted for estimating primary variable with the use of other relevant variables that are easy to measure online. Soft sensors can operate in parallel with hardware-based sensors, thereby providing a back-up and redundant measurement. Moreover, if the soft sensor model that describes the relationship between the primary and other operating variables is developed, the process operation can also be optimized by regulating the operating parameters [5,6].

Soft sensor models are generally established on the basis of first-principle and data-driven methods. Although often desirable, first-principle models are impractical in most cases because they involve immense complexity and often require several differential equations to be solved mathematically [7]. Meanwhile, data-driven models have wide applications because they can be directly developed based on operation data and detailed fundamental knowledge about the process is unnecessary. Many data-driven modeling methods have been proposed recently, and an extensive review can be found in the work of Kadlec, Gabrys and Strandt [8]. In particular, artificial intelligence techniques such as artificial neural networks (ANN) and support vector machine (SVM) have been widely applied in designing data-driven soft sensors because of their capability to describe highly nonlinear processes in chemical and energy industries. SVM is established based on the structural risk minimization (SRM) principle [9] and exhibits better generalization performance than ANN. An important variation of SVM is the least squares support vector machine (LSSVM), which uses squared errors and equality constraints to substitute inequality ones [10]. Thus, results can be obtained by solving a group of linear equations instead of a quadratic programming problem, thereby significantly reducing the training time of the model.

However, the data source is an important factor to consider when establishing a data-driven model. Several methods have been studied on the basis of experimental data acquired from field experiments [6]. The field experiment can be effectively designed to cover the majority of possible operating conditions, which makes it easier to implement a global approximation [11]. However, all of the concerned parameters

* Corresponding author. Tel.: +86 010 61772965; fax: +86 010 61772849.
E-mail addresses: you.lv@hotmail.com, you.lv@ncepu.edu.cn (Y. Lv).

http://dx.doi.org/10.1016/j.chemolab.2015.04.006
0169-7439/

are required to be set to specified values in this approach, and the normal operation and production are disturbed. Therefore, conducting a field experiment is costly and time-consuming. On the contrary, operation data can be easily stored and accessed in real time because of the wide availability of distributed control systems (DCS) and supervisory information systems (SIS). The accumulated historical data represent the real operation status of the industry process, thus, they can provide useful information for developing data-driven soft sensors [12].

Although excellent prediction results have been achieved with the application of soft-sensing techniques, some difficulties are still involved in practical applications. Prediction accuracy often decreases over time because of process drift, plant equipment abrasion, as well as the variations in the external environment and input materials. Thus, the most important problem is maintaining high prediction accuracy at all times after a soft-sensor model is initially established [13–15]. The factors that influence the accuracy of models based on historical operation data can be summarized into two points.

The first point is when the initial model is developed, operation segments with extensive representation should be selected as the training samples to ensure that a large operating region can be covered. Reliable predictions are more likely to be obtained if the current operation condition is within the coverage of the training dataset [16]. However, guaranteeing that all probable working regimes have been incorporated in the initial data set is almost impossible because a number of new working operation conditions could appear at a later time. In other words, collecting adequate representative operation samples is a successive and accumulating process.

The other important point is related to internal changes. For example, irreversible variations such as the abrasion of plant equipment and the alteration of input materials can change the characteristics of the industrial process, thus rendering the initial model incapable to describe the new relationship [17]. 'Irreversible' means that the process characteristics would not be recovered once this type of variation occurs.

Therefore, a maintenance and update strategy should be performed on the model. The most straightforward way to reduce the degradation is to reconstruct a new model periodically. The reconstruction can be implemented based on time difference of the variables, or using the data that are recently measured or have high similarities with the prediction [18]. However, in such an approach, previous training results would be abandoned, thus resulting in a heavily repeated computation. Moreover, reconstructed models are inclined to specialize in a narrow operation range, and they cannot predict the process variations accurately [19]. The model has to be trained again to be adequate for the former process if the previous operation conditions appear.

To tackle this issue, several update approaches have been presented, including the moving window method [20,21], recursive method [22], and their variants. Numerous studies have utilized the abovementioned frameworks to update data-driven soft sensor models. For example, Qin [23] proposed a recursive-PLS algorithm to tackle model degradation. Kaneko, Arakawa and Funatsu [24] presented a moving window method that uses recently acquired data to update the model. Online SVM and LSSVM models with time variables were also proposed to improve the prediction accuracy when the process changed [25–27]. In addition, a number of studies focused on the aspects of maintenance-free learning [28] and local learning, such as just-in-time learning, instance-based learning, and lazy learning [29,30], in which local models were established adaptively using relevant data samples to describe the current process characteristics. Further, Kaneko and Funatsu [31] classified the degradation of the models into three types and discussed different update effects based on the classification results. Besides that, they proposed an index to manage the database considering the variation, based on which the prediction accuracy of adaptive soft sensor models increased [19]. In addition, Kadlec, Grbić and Gabrys [32] and Slišković, Grbić and Hocenski [33] presented comprehensive reviews on different types of update techniques.

The conventional update framework is implemented by prioritizing samples that are newly measured or similar to the prediction data to solve the model degradation when the process characteristics vary [27]. However, a detailed analysis of the fundamental causes of the process variations can be helpful in the implementation of the update measures, which is seldom discussed in previous studies. This paper presents an adaptive LSSVM with a novel update based on the analysis of process variations, which are categorized into irreversible and reversible variations according to the similarity criterion. Addition and replacement strategies are used on the samples to update the LSSVM model incrementally according to the variation types. This method is then assessed with a benchmark nonlinear function and applied to predict the nitrogen oxide (NOx) emissions of a coal-fired boiler. Finally, the performance of the proposed model is compared with that of the original LSSVM without updates.

The paper is organized as follows. The process variations are analyzed and the update scheme is provided in Section 2. Section 3 introduces the incremental LSSVM. Section 4 presents the detailed update procedure based on the LSSVM model. A benchmark simulation is conducted in Section 5, and a real industrial application is described in Section 6. Section 7 presents the conclusions.

2. Process variations and model update

2.1. Process variation characteristics

Based on the changes in the explanatory and primary variables, and the rapidity of the changes, Kaneko and Funatsu [31] made a systematic study on the degradation of the models. Different from that, in this paper, the process variation characteristics are classified into two categories according to actual process characteristics and the variation causes. The first is irreversible variation, which is caused by internal factors such as changes in fuel property, abrasion and maintenance of plant equipment. This type of variation often occurs irreversibly because the characteristics of fuel and equipment cannot be reverted back to the previous condition. The other category is reversible variation, which is caused by external factors such as production load changes and valve opening alteration. Unlike the case with irreversible variations, the operation characteristics with reversible variations could return to the previous condition as the process proceeds.

A specific example of a single-in-single-out system with process variations is illustrated in Fig. 1. The initial operation condition is denoted as I in Fig. 1(a). The soft-sensing model $y = f(x)$ ($x \in [x_1, x_2]$) is established based on the representative training samples selected from the historical operation database. New operation instructions are received with the changes in the production loads. Accordingly, the operating variable x is set to a new value, and the process condition shifts to II ($x \in [x_3, x_4]$) and III ($x \in [x_5, x_6]$). Poor accuracy is obtained if the initial model $y = f(x)$ ($x \in [x_1, x_2]$) is still applied to predict the process characteristics. Notably, the abovementioned operation variation is reversible because the condition could possibly return to the initial area I with certain changes in the operation instructions. Therefore, the new samples collected in conditions II and III should be incorporated with condition I to update the initial model. The operation space is then expanded from $x \in [x_1, x_2]$ to $x \in [x_1, x_6]$, which is shown in Fig. 1(b).

Fig. 1(c) illustrates an irreversible variation from condition I to IV, which can be primarily attributed to the abrasion and maintenance of plant equipment. Different from the previous variation, the process characteristics could no longer change back. Therefore, the old samples representing former characteristics are already of little use and should be replaced by the newly-collected ones. On the basis of the above analysis, the final model should be adequate to describe the new process characteristics shown in Fig. 1(d) after the model update.