Chemometrics and Intelligent Laboratory Systems 143 (2015) 58–78

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Online local learning based adaptive soft sensor and its application to an industrial fed-batch chlortetracycline fermentation process

Huaiping Jin, Xiangguang Chen *, Jianwen Yang, Li Wang, Lei Wu

Department of Chemical Engineering, Beijing Institute of Technology, Beijing 100081, People's Republic of China

ARTICLE INFO

Article history:
Received 1 September 2014
Received in revised form 18 February 2015
Accepted 20 February 2015
Available online 28 February 2015

Keywords:
Adaptive soft sensor
Local learning
Just-in-time learning
Adaptive sample selection
Dual updating
Fed-batch chlortetracycline fermentation processes

ABSTRACT

This work presents a new method for adaptive soft sensor development by further exploiting just-in-time modeling framework. In the presented method, referred to as online local learning based adaptive soft sensor (OLLASS), the samples used for local modeling are selected based on the mutual information (MI) weighted or neighbor sample based similarity measure. Then, two adaptive methods, namely self-validation and neighbor-validation, are developed to adaptively select the optimal local modeling size for scenarios without and with the neighbor output information, respectively. Further, a real-time performance improvement strategy is used to enhance the online modeling efficiency. Moreover, an online dual updating strategy is proposed to activate infrequent local model updating and model output offset updating in turn, which allows significantly reducing the online computational load by avoiding unnecessary local model reconstruction while at the same time maintaining high estimation accuracy by performing offset compensation. A maximal similarity replacement rule using MI weighted similarity measure is used for database updating. The superiority of the proposed OLLASS method over traditional soft sensors in terms of the estimation accuracy, adaptive capability and real-time performance is demonstrated through an industrial fed-batch chlortetracycline fermentation process.

1. Introduction

The lack of reliable online sensors, which can measure the critical process variables that are usually obtained by offline laboratory analysis, poses great challenges for implementing advanced control, efficient monitoring and optimization of industrial processes. In recent years, soft sensor techniques have gained increasing popularity in process industry and aim to provide online estimates of difficult-to-measure variables in a real-time fashion [1–5]. Instead of using hardware instruments, soft sensing method relies upon an inferential model to predict the target variable by using other highly correlated but easy-to-measure variables as model inputs.

In general, one can distinguish two types of soft sensors, namely model-driven and data-driven [2]. Though modeling soft sensors from first principles are desirable, in most cases it is impossible to develop accurate mechanistic models due to the unavailability of in-depth process knowledge as well as the heavy workload involved, especially for complex processes. Furthermore, the mechanistic models are usually built under the ideal operating conditions, whereas the actual process characteristics may be significantly different. Alternatively, data-driven soft sensors have become increasingly popular in industrial applications as minimal process knowledge is needed while the plant historians provide abundant operation data for empirical model development [2,4,6,7]. Modern measurement techniques enable large amounts of plant data to be collected, stored and analyzed, thereby rendering data-driven modeling more attractive than model-driven methods for soft sensor development.

The most common data-driven modeling techniques for soft sensor design are multivariate statistical methods such as principle component regression (PCR) [8–10], partial least squares (PLS) [11,12], independent component analysis (ICA) [13,14], and their nonlinear variants such as kernel PLS (KPLS) [15–18] and kernel PCR (KPCR) [18,19]. Methods of this type usually identify models within the lower-dimensional subspace projected from the original input data. They gained popularity due to their statistical background, ease of interpretability as well as their strong capability of dealing with data colinearity. Meanwhile, the machine leaning methods have been widely accepted as useful techniques for soft sensor design such as artificial neural networks (ANN) [20,21], support vector regression (SVR) [22,23], neuro-fuzzy systems [24], Gaussian process regression (GPR) [25,26] and Gaussian mixture regression [27]. Various data-driven soft sensor modeling methods and their applications were reviewed by Kadlec et al. [2].

Traditionally, a soft sensor usually relies upon a global regression model and aims to achieve universal generalization performance. However, global methods may lead to inaccurate estimations in some local regions where process characteristics change. Meanwhile, when dealing with large datasets, global approaches become less competitive because of the difficulties in determining model structure and the complexity

* Corresponding author. Tel.: +86 13601333018; fax: +86 10 68914662.
E-mail addresses: jinhuaiping@gmail.com (H. Jin), xgc1@bit.edu.cn (X. Chen), yangjianwen@bit.edu.cn (J. Yang), wangli@bit.edu.cn (L. Wang), wulei@bit.edu.cn (L. Wu).

http://dx.doi.org/10.1016/j.chemolab.2015.02.018
0169-7439/

associated with the optimization problem. Another fundamental limitation of global methods lies in the difficulty of updating model online when the process dynamics have moved away from the nominal operating space. To overcome these shortcomings, various local learning methods have been developed, mainly including multi-model methods and ensemble learning. In multi-model modeling, process data are divided into different sub-domains and local models are constructed over every domain [16,28,29]. Then, the target variable is predicted by using a deterministic local model representing the identical operation phase or mode. The commonly used techniques for achieving phase or mode identification are clustering methods such as *k*-means algorithm [30] and Gaussian mixture model [31]. In contrast to multi-model strategy, ensemble learning requires building a set of local models and then provides the final prediction by combing all available local model outputs [32]. The way in which the global data is split into the local partitions depends upon the algorithm. Typical approaches for this purpose include clustering algorithms [31,33], bagging [13,26,34], boosting [35], etc. To obtain the ensemble prediction, the predictions of local models have to be combined by using simple averaging, Bayesian inference strategy [9,26,32,33,36], or regression coefficient weighting [13,34], etc. Though multi-model and ensemble strategies can provide more robust and reliable predictions than global models, such methods often suffer from the drawback of requiring a priori knowledge to decide the partition of operation data. In practice, the quantitative and precise information of data divisions is often unavailable. Even worse, such offline local learning methods are essentially nonadaptive since the historical data partitions remain unchanged once deployed into real-life operation, which may lead to performance deterioration when new process states take place.

Apart from the model structure, another crucial issue concerning soft sensor application is the model maintenance. Even if a highly accurate soft sensor is developed, its estimation performance will deteriorate due to changes in the state of plants and process characteristics, such as set-point changes, catalyst deactivation, seasonal effects, variances of raw materials, and equipment aging [37,38]. This is because the built model usually represents only the process states observed during the training phase, whereas the current process dynamics may be significantly different. In fact, a recent questionnaire survey revealed that the most important problem of current soft sensors was how to cope with changes in process characteristics and maintain high estimation accuracy for a long period of time, i.e. model maintenance [39]. Thus, from the practical viewpoint, soft sensor should be appropriately updated using the newest data to avoid performance degradation.

To cope with changes in process characteristics and automatically update soft sensors, various adaptation mechanisms have been developed, such as recursive adaptation methods [37,40], moving window (MW) techniques [37,41–48], time difference (TD) modeling [49], and offset compensation [41–45]. Recursive methods can adapt models to new operating conditions recursively. However, when a process is operated within a narrow range, the model may adapt excessively and thus result in blind updating. Another issue is that they cannot cope with abrupt changes. Additional challenging issue for recursive methods is related to the selection of an appropriate forgetting factor. Moving window methods have also been developed to update predictive models using a set of data points that are measured most recently. Similar to the selection of parameters like forgetting factor, it is also difficult to select the optimal size of adaptation window (window size) and the adaptation intervals between the updates (step size). Another adaptation method is the time difference modeling, which was proposed by constructing models based on the time difference of a target variable and that of input variables [49]. Though TD models can effectively handle the effects of deterioration with age, such as drift and gradual changes in the state of plant, it cannot account for complex process changes but those changes progressing at a constant rate. Moreover, the estimation accuracy of soft sensors can be improved by performing offset compensation [41–45]. In addition, an adaptive soft sensor with online Bayesian model updating strategy was proposed [50]. The basic idea of this approach is that the nominal nonlinear state space model is first identified through expectation–maximization algorithm and online prediction of difficult-to-measure variable is achieved by adaptive synthesis of data from various hardware sensors and model predictions through sequential Bayesian filtering. The adaptive capability of the resulting soft sensor is attributed to the Bayesian model calibration scheme and particle filter, which are applied to simultaneously update the process state and the calibration parameters using multirate measurements from online-analyzer and lab data. However, such time-series models using dynamic model structure may be not well suited for online real-time quality and state prediction in practice due to their limited capability of handling missing values and inadequate sampling intervals. To further improve the performance of adaptive soft sensors, adaptation methods are often combined with local learning. Examples of combination of recursive methods and local learning can be found in [43,51,52] while the examples integrating MW methods with local leaning can be found in [53,54]. An ensemble of TD soft sensor has also been reported in [55].

Meanwhile, just-in-time (JIT) modeling was proposed to cope with changes in process characteristics as well as nonlinearity, and it has been widely used for nonlinear process monitoring and soft sensing [5,15,56–63]. The general idea of JIT modeling is to build a local model from past data around the query point only when the estimation for the query data is required, and then the local model is discarded after estimation. The superiority of JIT modeling over conventional methods is due to its particular online local learning framework. Since only samples similar to current state are selected for local model construction, JIT modeling can cope with abrupt changes as well gradual ones in process characteristics. It can also efficiently deal with process nonlinearity by repeatedly building local models. Traditionally, nonlinear methods such as ANN and SVM are employed to handle process nonlinearity. The methods such as PLS and PCA, which are essentially linear modeling techniques, cannot address process nonlinearity unless certain nonlinear variations such as kernel and spine functions are integrated [64–66]. In contrast, JIT learning or locally weighted method such as locally weighted PLS allows capturing nonlinear characteristics and achieving high model performance based on local learning even though simple linear techniques such as PCR and PLS are used. Another advantage of JIT method is that it can avoid the problems of region division and interpolation between local models, which are often encountered in offline local learning. Additional advantage of JIT modeling is its inherently adaptive nature, which is obtained by simply adding the new data into the database. Therefore, JIT method has become a promising technique for adaptive soft sensor development and has been successfully applied to industrial processes [3,5,61].

Over the last decade, research works concerning JIT modeling have mainly focused on the definition of similarity measure and the selection of regression function to enhance the estimation accuracy. To construct highly accurate JIT soft sensors, it is crucial to define the similarity measure. So far, various similarity measures have been proposed, such as Euclidean distance [5,67], the angle [58,59], the correlation among variables [63,68–70], the weighted Euclidean distance (WED) [15,67,72], and the adaptive WED [73] measures. On the other hand, the estimation accuracy of JIT models can be further improved by selecting suitable regression function. Apart from the most frequently used linear modeling techniques such as PCR and PLS methods, some attempts have also been paid to nonlinear techniques such as KPLS [15], KPCR [19], SVR [74], least squares SVR (LSSVR) [67], and weighted LSSVR [75]. While some success have been reported based upon the use of JIT modeling, there remain some problems that have yet to be overcome for the introduction of JIT soft sensors into practice.

First of all, the definition of similarity measure should be further explored. For example, WED similarity measure performs better than the commonly unweighted one since different importance is assigned to input variables according to their relevance to the output variable.