Chemometrics and Intelligent Laboratory Systems 151 (2016) 228–244

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Dual learning-based online ensemble regression approach for adaptive soft sensor modeling of nonlinear time-varying processes

Huaiping Jin [a], Xiangguang Chen [a,*], Li Wang [a], Kai Yang [a,b], Lei Wu [a]

[a] Department of Chemical Engineering, Beijing Institute of Technology, Beijing 100081, People's Republic of China
[b] Beijing Research & Design Institute of Rubber Industry, Beijing 100081, People's Republic of China

ARTICLE INFO

Article history:
Received 25 September 2015
Received in revised form 3 December 2015
Accepted 10 January 2016
Available online 18 January 2016

Keywords:
Adaptive soft sensor
Just-in-time learning
Statistical hypothesis testing
Ensemble learning
Incremental local learning
Partial least squares

ABSTRACT

Soft sensors have been widely used to estimate difficult-to-measure variables in the process industry. However, the nonlinear nature and time-varying behavior of many processes pose significant challenges for accurate quality prediction. Thus a novel adaptive soft sensor, referred to as dual learning-based online ensemble regression (DLOER), is proposed for nonlinear time-varying processes. To deal with process nonlinearity, just-in-time (JIT) learning is used to build local domains and local models simultaneously while statistical hypothesis testing is employed to remove redundant local models. As a result, multiple diverse local models are constructed for characterizing various process states. Then the posterior probabilities of each test sample with respect to different local models are estimated through Bayesian inference and further set as adaptive weights to combine local predictions into a final output. Moreover, DLOER is equipped with incremental local learning and JIT learning for model adaptation, which enables recursive adaptation and online inclusion of local models, respectively. Therefore, process nonlinearity can be well handled under the local learning framework while both gradual and abrupt changes of processes can be efficiently addressed using the dual learning-based adaptation mechanism. The effectiveness of the DLOER approach is demonstrated through a fed-batch penicillin fermentation process.

1. Introduction

The reliable and realtime measurements of quality variables play an important role in process monitoring, control, stability and improving product quality. However, in the process industry, there remain some difficult-to-measure variables, which can be determined either at low sampling rates or through offline analysis only. Even if online analyzers are available, they still encounter limitations such as unacceptable cost and heavy maintenance load. Over the last decade, soft sensor technology, as a promising solution towards online prediction, has attracted fast-growing interests in both academia and process industries [1–5]. By using a soft sensor model between these easy-to-measure variables and those difficult-to-measure ones, the key process variables can be estimated online.

Traditionally, soft sensors are based on first principles models. Nevertheless, this class of methods requires in-depth process knowledge and tremendous effort for model development. Alternatively, the data-driven soft sensors gained growing popularity in the process industry, which is due to the increasing availability of operating data, as well as the prosperity of computational learning techniques to process the data [6–9]. The most common data-driven soft sensors are based on multivariate statistical techniques, such as principal component regression (PCR) [10–12], partial least squares (PLS) [13–15] and independent component regression (ICR) [16–17]. Meanwhile, the machine learning methods, such as artificial neural networks (ANN) [18–20], support vector regression (SVR) [21–23], and Gaussian process regression (GPR) [24–26], have also been accepted as useful tools for soft sensor development. Despite the availability of a variety of soft sensor methods, it remains challenging to develop high-performance soft sensors, because industrial processes often exhibit nonlinear and time-varying behaviors.

To address process nonlinearity, a straightforward strategy is to use nonlinear modeling techniques such as ANN, SVR, and GPR. However, such approaches are usually based on global models given the underlying assumption of a constant operating phase/mode throughout the process. In practice, industrial processes are often characterized by multiple operating phases/modes [27–30]. In such a case, global models fail to function well. Thus considerable efforts have been paid to local learning, which employs the "divide and rule" philosophy. By exploiting the local learning framework, the complex nonlinear relationship between input and output data can be well described by multiple local valid models, each of which is responsible for one specific operating region of the process.

In general, one can distinguish three classes of local learning methods, namely JIT learning [31,32], multi-model modeling [33,34],

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

http://dx.doi.org/10.1016/j.chemolab.2016.01.009
0169-7439/

and ensemble learning [35,36]. JIT learning can cope with process nonlinear by building local models repeatedly during online operation. However, it encounters heavy online computational load and difficulties in defining appropriate similarity measures. An alternative local learning method is the multi-model approach, where a model library is defined as a collection of local models. Then the output variable can be predicted using the corresponding local model relevant to the query state. Though multiple localized models can handle multiphase/multimode operations, they may not efficiently characterize the transient dynamics between phases or modes. This problem can be solved using ensemble learning methods, in which local models are first trained and then combined to provide an overall output estimate.

Apart from process nonlinearity, another crucial problem remaining to be solved for soft sensor development is how to deal with the time-varying behavior of processes and maintain high model performance for a long period of time. This problem is also known as model maintenance or adaptation [37,38], and has been even recognized as the most important problem of current soft sensors, as revealed in a recent questionnaire survey [39]. Even if an accurate soft sensor model can be built from the training data, its prediction accuracy will deteriorate over time. The factors contributing to such behavior include changes of raw materials, catalyzing performance loss, sensor drift, abrasion of mechanical components and external environmental changes.

To reduce degradation, various adaptation mechanisms have been proposed to update soft sensors, such as moving window (MW) [40,41], recursive adaptation (RA) [42–44], and JIT learning [45,46]. Generally, these adaptation strategies can be categorized into two groups [47]: temporal adaptation and spatial adaptation methods. MW and RA belong to the former type since they update the soft sensor model using the newest data. In contrast, JIT learning is classified into the spatial adaptation because it builds a new local model from scratch using the most relevant data in the data space.

The temporal adaptation mechanisms assume that the current process state is highly similar to the newest state, which enables to efficiently handle gradual changes of processes. Nevertheless, they cannot function well when abrupt changes occur, because in such situation the process characteristics change rapidly from one state to another. In comparison, the spatial adaptation methods allow dealing with abrupt changes in process characteristics because they utilize the most relevant data from a database covering a wide range of operating conditions. However, the temporal relationship of samples is ignored in spatial adaptation models. In practical applications, it is often the case that both gradual and abrupt changes exist simultaneously. It is therefore appealing to combine the temporal and spatial adaptation mechanisms to update soft sensor models, thereby addressing both types of time-varying behavior.

To allow handling nonlinearity and time-varying issues simultaneously, in recent years, there has been an increasing interest in the integration of local learning and adaptation capability [25,26,48–53]. By using such hybrid strategies, process nonlinearity can be well handled under local learning framework while time-varying behavior can be captured through online model adaptation. Therefore, this paper aims to develop a local learning based adaptive soft sensor for nonlinear time-varying processes. Among those aforementioned local learning methods, only the ensemble learning is involved in this work for two reasons. On one hand, once an initial ensemble model is built from the training data, the online prediction can be conducted very fast. On the other hand, in terms of the prediction robustness, the ensemble methods outperform other local learning methods that only use one single local model for each run of prediction. However, developing a high-performance adaptive ensemble soft sensor needs to well solve three critical issues: How are the local domains constructed? How are the local models combined? How is the ensemble model updated?

The first key task in ensemble learning is to construct local domains so that local models can be built from the corresponding data subsets. The most frequently used approaches for this purpose are the clustering methods, such as fuzzy c-means (FCM) [34,54] and Gaussian mixture models (GMM) [25,33]. However, there are several issues associated with such partition methods. First, the number of clusters needs to be preset, whereas in practice the precise quantitative information of process divisions is often unavailable. Second, it is difficult to include new local domains online without retraining from scratch, thus limiting its ability to capture new process states. Especially, when significant variations of processes occur, the division results are no longer valid. Thus the capability of launching new local model online is highly desirable for the ensemble models. Third, the best clustering results in terms of the data distribution do not mean the best partitions for ensemble learning from the point of view of the prediction performance.

Another method for process divisions is the moving window (MW), which builds local domains by collecting successive samples included in a certain period of time [10,55]. The rationale of this strategy lies in the assumption of highly similar correlation among variables within a time window. By repeatedly shifting the window forward, a large number of windows can be constructed. However, severe redundancy may exist between windows. To address this issue, statistical hypothesis testing such as Student's t-test can be employed to evaluate whether two windows belong to the same process state or not [48,50,51]. The redundant windows can be included to the same local domain or discarded directly. However, a potential limitation of the MW approach is that local domain data are selected only based on time relevance and thus samples within a certain window may contain multiple distinct process states, especially when abrupt changes occur and the window size is not set appropriately.

The next crucial operation for ensemble modeling is the combination of local models. Given the input samples, each of local models can make predictions of the target variable. Then these local predictions have to be integrated into an overall estimation of the ensemble model. Traditionally, local models are combined by building a linear or nonlinear relationship between the local outputs and the final output based on the training data. Examples of such ensemble methods include simple mean, trimmed mean, weighted mean based on the training accuracy, PLS coefficients, and ANN model [2,16,56,57]. Nevertheless, these methods are nonadaptive, and the resulting combination weights or weighting models remain unchanged once deployed into real-life operation. One significant limitation of such nonadaptive combination approaches is that they are prone to assign larger weights to the models that give better prediction on the training data. In fact, if some of the models are severely overfitted on the training data, the impact effect of overfitting problem could be further amplified by these models assigned large weights. Consequently, the generalization capability of the ensemble models for new test data may deteriorate.

As opposed to the nonadaptive ensemble methods, adaptive weighting strategies are promising since the importance of local models is dynamically determined based on the relevance between the query data and local models. So far, several frequently used indices for adaptive weighting are summarized as follows: (i) distances between the query data and centers of local domains [58]; (ii) posterior probabilities of the query data with respect to different local models [23,25,33]; (iii) fuzzy memberships [34]; prediction uncertainty [24,54]; (iv) prediction accuracy on the recently measured samples [49,52,53] or on the similar samples to the query data [50,51]; and (v) monitoring statistics [17].

Another major concern with the ensemble soft sensors is to incorporate adaptation mechanisms into the offline built ensemble models. The ensemble learning framework allows performing flexible adaptations, which can combine a subset of the following strategies [36,37]: (i) adaptation of the models' parameters; (ii) adaptation of the combination weights; and (iii) launching new models. Nevertheless, many of the current adaptive ensemble soft sensors only pay attention to the former two adaptation operations, where the temporal adaptation mechanisms such as recursive strategy and MW approach are usually used for model updating. As for the combination weights, the aforementioned adaptive