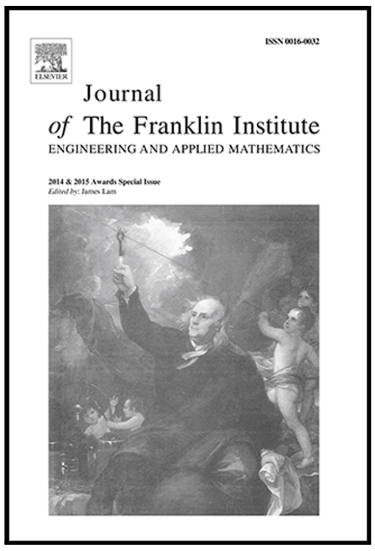

Soft sensor modeling with a selective updating strategy for Gaussian process regression based on probabilistic principle component analysis

Weili Xiong , Xudong Shi

PII: S0016-0032(18)30329-6
DOI: 10.1016/j.jfranklin.2018.05.017
Reference: FI 3451

To appear in: *Journal of the Franklin Institute*

Received date: 6 September 2017
Revised date: 25 April 2018
Accepted date: 13 May 2018

Please cite this article as: Weili Xiong , Xudong Shi , Soft sensor modeling with a selective updating strategy for Gaussian process regression based on probabilistic principle component analysis, *Journal of the Franklin Institute* (2018), doi: 10.1016/j.jfranklin.2018.05.017

Soft sensor modeling with a selective updating strategy for Gaussian process regression based on probabilistic principle component analysis

Weili Xiong [a,b*] ,Xudong Shi [a]

[a] School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China

[b] Key Laboratory of Advanced Process Control for Light Industry of Ministry of Education, Jiangnan University, Wuxi 214122, China

*Corresponding author at: Key Laboratory of Advanced Process Control for Light Industry of Ministry of Education, Jiangnan University, Wuxi 214122, China. E-mail address: greenpre@163.com

Abstract

Considering the deviation of the working condition and the high updating frequency of the traditional moving window methods, this paper proposes a selective strategy of moving window for the Gaussian process regression in the latent probabilistic component space. First, the probabilistic principle component analysis (PPCA) is employed to deal with the multi-dimensional issue and extract essential information of the process data. Because the latent probabilistic components are more sensitive to the deviation of the working condition in the industrial process than the original data, the regression performance is improved under the PPCA framework. Under the proposed strategy, the soft sensor is able to detect the change of the working condition, and the updating is activated only when the predicted error exceeds the preset threshold, otherwise the model is kept unchanged. Furthermore, the promotion of both predicted accuracy and efficiency can be obtained by regulating the threshold. To test the effectiveness of the proposed method, a wastewater case study is provided, and the result shows that the proposed strategy works better under the probabilistic than other conventional methods.

Keywords : Probabilistic Principle Component Analysis; Selective Updating; Gaussian Process Regression; Soft Sensor

1. Introduction

Nowadays, with the demand of high industrial products, soft sensors are more and more important to obtain reliable and real-time estimations of the key quality variables. Because of hostile environment and expensive cost, it is difficult to establish hardware sensors for online measurement of the quality variables [1-2]. Constructing soft sensors is a potential solution to address this problem, which is based on the functional relationship between the easy-to-measure variables (usually the input or process variables) and the hard-to- measure variables (usually the output or quality variables). Generally, the soft sensing methods can be categorized into three types: model-based methods, data-based methods, and the mixed methods. Compared with the model-based methods relying on chemical and biological mechanism of process greatly, the data-based methods are much more flexible and less complex, and thus these methods have been widely used. Up to now, plenty of data-based soft sensor models have been proposed, such as principle component regression (PCR) [3], partial least square(PLS) [4], artificial neural network (ANN) [5], and support vector machine (SVM) [6] for chemical and biological engineering. Recently, a novel probabilistic nonlinear modeling method, namely, the Gaussian process regression (GPR) has caught much attention in the field of soft sensor [7-10], due to its excellent regression performance and the ability to provide the uncertainty of the predicted result.

Usually historical data of offline models will cover as much as possible working conditions to deal with the deviation of the working conditions, such as the drift of instrument parameters, the varied properties of raw materials, and the changes of the environment. However, the characteristics of the industrial object will inevitably drift out of after the model was constructed, because of the complex