Chinese Journal of Chemical Engineering 25 (2017) 116–122

Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

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

Process Systems Engineering and Process Safety

Dynamic soft sensor development based on Gaussian mixture regression for fermentation processes☆

Congli Mei *, Yong Su, Guohai Liu, Yuhan Ding, Zhiling Liao

School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China

ARTICLE INFO

Article history:
Received 1 April 2016
Received in revised form 9 June 2016
Accepted 6 July 2016
Available online 14 July 2016

Keywords:
Dynamic modeling
Process systems
Instrumentation
Gaussian mixture regression
Fermentation processes

ABSTRACT

The dynamic soft sensor based on a single Gaussian process regression (GPR) model has been developed in fermentation processes. However, limitations of single regression models, for multiphase/multimode fermentation processes, may result in large prediction errors and complexity of the soft sensor. Therefore, a dynamic soft sensor based on Gaussian mixture regression (GMR) was proposed to overcome the problems. Two structure parameters, the number of Gaussian components and the order of the model, are crucial to the soft sensor model. To achieve a simple and effective soft sensor, an iterative strategy was proposed to optimize the two structure parameters synchronously. For the aim of comparisons, the proposed dynamic GMR soft sensor and the existing dynamic GPR soft sensor were both investigated to estimate biomass concentration in a Penicillin simulation process and an industrial Erythromycin fermentation process. Results show that the proposed dynamic GMR soft sensor has higher prediction accuracy and is more suitable for dynamic multiphase/multimode fermentation processes.

1. Introduction

In fermentation processes, some important quality variables, *e.g.*, biomass concentration, are difficult to measure online due to measurement limitations such as cost, reliability, and long dead time. From the viewpoint of control and optimization, these measurement limitations may cause important problems such as product loss, energy loss, and undesired byproduct generation. Over the past decades, soft sensors have been widely used to tackle these problems, which provide frequent estimations of key process variables through those that are easy to be measured online [1–4].

The most popular soft sensor methods are partial least squares (PLS) [5,6], artificial neural networks (ANN) [7,8], and support vector machines (SVM) [2,9]. Recent reviews of soft sensor methods can be found in [4,1,10]. Usually, soft sensors are constructed based on process measurements easy to measure online. From the viewpoint of measuring, one of the main disadvantages of those traditional soft sensors is lacking information of precision. That restricts above-mentioned soft sensors in practical cases. Another important problem which cannot be ignored is that dynamic multiphase/multimode processes are wide in fermentation processes and cannot be modeled effectively by single data driven regression models [11], *e.g.*, PLS, ANN and SVM. Generally speaking, these processes may result in complexity and poor performance of single models.

Recently, a relative new machine learning method, *i.e.*, Gaussian process regression (GPR), has been developed, and began to be applied in soft sensor modeling [12,13]. GPR is usually trained by optimizing the hyperparameters using the expectation maximization (EM) algorithm with the squared exponential covariance function which is commonly employed [14]. This regression method has many useful features that distinguish it from other machine learning techniques, particularly in the field of nonlinear modeling, such as ability to measure prediction confidence, few training hyper-parameters and possibility to include prior knowledge into the model. It should be noticed that GPR soft sensors are constructed based on the assumption that process data are generated from a single operating region and follow a unimodal Gaussian distribution. However, for complex multimode/multiphase processes, the basic assumption of multivariate Gaussian distribution may not be met because of the mean shifts or covariance changes. Then Gaussian mixture regression (GMR) was introduced to construct soft sensors for those complex processes [11]. Besides the process characteristics of multimode/multiphase, it cannot be neglected that fermentation processes are dynamic systems. Conventional static soft sensors commonly rely on the assumption that processes are operating at steady states. It was pointed out that based on static regression models applied in a dynamic process may result in complexity of modeling and large errors of estimates [15,16]. Recently, considering the merits of GPR and dynamics in processes, a dynamic GPR soft sensor was proposed to estimate biomass concentration in a fermentation process [13]. In the model, besides input measurements, delayed outputs are also fed back and used as regressors. However,

☆ Supported by the Natural Science Foundation of Jiangsu Province of China (BK20130531), the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD [2011]6) and Jiangsu Government Scholarship.
* Corresponding author.
E-mail address: clmei@ujs.edu.cn (C.L. Mei).

http://dx.doi.org/10.1016/j.cjche.2016.07.005
1004-9541/

regressors of the GPR soft sensor were selected heuristically from numerous alternatives [13]. A systematic method of selecting regressors needs to be studied further.

For dynamic multiphase/multimode fermentation processes, this paper presented a systematic method of constructing a dynamic GMR soft sensor to overcome the abovementioned problems. How to optimize the number of Gaussian components and select regressors are crucial to the dynamic GMR soft sensor. The former is related to the number of process phases/modes [11], and the latter can be represented using the order of the model. In the work, for simplicity and effectiveness of modeling the soft sensor, an iterative strategy was presented to optimize the two parameters simultaneously. In the strategy, the Bayesian information criterion (BIC) commonly used in the field of modeling is employed as an evaluation criterion to determine the structure of the soft sensor model.

The remainder of this paper is organized as follows. The GMR and the expectation maximization (EM) estimation are introduced in Section 2. Then, in Section 3, we introduce the proposed dynamic GMR soft sensor and the iterative strategy. A numerical example and an application example are investigated together to verify the effectiveness of the proposed GMR soft sensor in Section 4. Finally, in Section 5, conclusions are made.

2. Introduction to GMR and Expectation Maximization Estimation

2.1. GMR

Assume $\boldsymbol{X}$ represents the space of the explanatory variables and $\boldsymbol{Y}$ is the space of the response variables. x is the input of training data ($x \in \boldsymbol{X}$) and y is the ideal output data ($y \in \boldsymbol{Y}$). For the given x and y, the joint probability density is given as [17]

$$f_{XY}(x,y) = \sum_{j=1}^{K} \pi_j \phi\left(x, y; \mu_j, \sum_j\right) \tag{1}$$

where, subsequently, the mean μ_j and covariance $\sum_j$ can be divided into the input and output parts like the following

$$\mu_j = \begin{bmatrix} \mu_{jx} \\ \mu_{jy} \end{bmatrix}, \sum_j = \begin{bmatrix} \sum_{jXX} & \sum_{jXY} \\ \sum_{jYX} & \sum_{jYY} \end{bmatrix}$$

Eq. (1) shows that the relationship between the explanatory variables and the prediction value can be described by Gaussian mixture models (GMMs) where $\phi(x,y;\mu_j,\sum_j)$ denotes the probability density function of the multivariate GMM. The parameters of this model include the number of the mixture components K, the prior π_j, the mean value μ_j, and the variance of each Gaussian component $\sum_j$, which are represented as $\theta = (\theta_1, \theta_2, \cdots, \theta_k)$ with $\theta_j = (\pi_j, \mu_j, \sum_j)$ and the constraint $\sum_{j=1}^{k} \pi_j = 1 (0 \le \pi_j \le 1)$.

Similarly, the marginal probability density can be given as [17]

$$f_X(x) = \int f_{X,Y}(x,y)\mathrm{d}y = \sum_{j=1}^{K} \pi_j \phi\left(x; \mu_{jX}, \sum_{jX}\right) \tag{2}$$

The global GMR function can be deduced by combining Eqs. (1) and (2)

$$f_{Y/X}(y/x) = \frac{f_{XY}(x,y)}{f_X(x)} = \sum_{j=1}^{K} w_j(x) \phi\left(y; m_j(x), \sigma_j^2\right) \tag{3}$$

with the mixing weight

$$w_j(x) = \frac{\pi_j \phi\left(x; \mu_{jX}, \sum_{jX}\right)}{\sum_{j=1}^{K} \pi_j \phi\left(x; \mu_{jX}, \sum_{jX}\right)} \tag{4}$$

The mean and the variance of the conditional distribution can be acquired in closed form by

$$m_j(x) = \mu_{jX} + \sum_{jYX} \sum_{jX}^{-1} \left(x - \mu_{jX}\right) \tag{5}$$

$$\sigma_j^2 = \sum_{jYY} - \sum_{jYX} \sum_{jX}^{-1} \sum_{jXY} \tag{6}$$

The prediction given a new input can be obtained by computing the expectation over the conditional distribution $f_{Y/X}(y/x)$ [17]

$$E\left[f_{Y/X}(y/x)\right] = \sum_{j=1}^{K} w_j(x) m_j(x) \tag{7}$$

It can be seen that the weight function $w_j(x)$ is not determined by the local structure of the data but the components of a global GMM. Therefore, the GMR model is a global parametric model with nonparametric flexibility.

2.2. EM algorithm for GMMs

To use a GMM, the unknown parameter set θ of probabilistic weights and model parameters of each Gaussian component should be estimated first. Common methods for this problem include the maximum likelihood estimation (MLE) and EM algorithm. With a set of given data ($\boldsymbol{X}, \boldsymbol{Y}$) which is realized by estimating model parameters θ in Eq. (1). This process can be realized by maximizing the log-likelihood function $L(\theta_k)$ which can be expressed as [18]

$$L(\theta_k) = \ln \prod_{i=1}^{N} p(x_i, y_i) = \sum_{i=1}^{N} \ln \sum_{j=1}^{K} \pi_j \phi\left(x, y; \mu_j, \sum_j\right) \tag{8}$$

For the given training data, θ is calculated by maximizing this function with the EM algorithm in the iterative means. It includes two steps [19].

(1) E-step (expectation step)

$$p^{(s)}(l_k/x_i) = \frac{\pi_j^{(s)} \phi\left(X, \mu_j^{(s)}, \sum_j^{(s)}\right)}{p^{(s)}(x_i, \theta)} (i = 1, 2, \cdots N; j = 1, 2, \cdots k) \tag{9}$$

where $p^{(s)}(l_k/x_i)$ denotes the posterior probability of the ith training sample within the kth Gaussian component at the sth iteration.

(2) M-step (maximum step)

$$\mu_j^{(s+1)} = \frac{1}{\pi_j N} \sum_{i=1}^{N} p^{(s)}(l_k/x_i) x_i \tag{10}$$

$$\sum_j^{(s+1)} = \frac{1}{\pi_j N} \sum_{i=1}^{N} p^{(s)}(l_k/x_i) \left[(x_i - m_j)(x_i - m_j)^T\right] \tag{11}$$

$$\pi_j^{(s+1)} = \frac{1}{N} \sum_{i=1}^{N} p^{(s)}(l_k/x_i) \tag{12}$$

where $\mu_j^{(s+1)}$, $\sum_j^{(s+1)}$, and $\pi_j^{(s+1)}$ are the mean, covariance, and prior probability of the kth Gaussian component at the $(s+1)$th iteration, respectively.