Asian Journal of Information Technology 6 (2): 230-235, 2007

Bioprocess Monitoring Using Local Model Networks

Azzeddine Amri, Messaoud Ramdani and Mouldi Bedda
Department of Electronics, Faculty of Engineering, University of Annaba,
BP. 12, 23000, Annaba, Algeria

Abstract: Monitoring of fermentation processes is of great importance to ensure their safe operation and consistent high quality products. Unfortunately, some of the difficulties such as the lack of on-line sensors for indication of fermentation performance, the presence of significant nonlinear behaviour and difficulties in designing accurate mechanistic models limit our ability to provide adequate monitoring. The amount of time and cost involved in developing detailed fundamental models combined with the commercial pressure to reduce the time-to-market requires different modelling, monitoring and control techniques. The local modelling methodology can be used in the design of soft-sensors. In this study, we propose a Local Model Network (LMN) with improved learning scheme for the bioprocess monitoring. The validity of the approach is illustrated on a gluconic acid fermentation process for the design of a soft-sensor to provide an estimation of the product concentration.

Key words: Multi-model, nonlinear dynamics modelling, fermentation process, LMN, soft-sensor, gluconic acid

INTRODUCTION

Bioprocess technology is currently employed for the production of several fermentation derived products such as pharmaceuticals and food products. Because of the complex nature of micro organism's growth and product formation in batch and fed-batch cultures, the supervision of bioprocesses continues to present a challenge to chemical engineers. Current researchers to monitoring fermentation processes have focused on the use of either fundamental mathematical models with state estimation or knowledge-based models.

At the heart of bioprocess supervision is the ability to monitor the bioprocess progress in terms of physicochemical environment (pH, temperature, pressure, dissolved oxygen) and process dynamics (feed rates, carbon dioxide evolution, oxygen uptake rate). However, the inability to measure some key variables represents an obstacle for the direct implementation of supervision strategies. To overcome this obstacle, the so-called "Soft-sensors" have been proposed and used for a variety of purposes. "Soft sensors" are mathematical algorithms that can provide estimations of some unmeasured variables from measured ones (Isermann, 1997; 1998; Shimizu, 1996). For example, in the case of fermentation systems, variables such as temperature, dissolved oxygen, pH, agitator power input and flow rates which are recorded frequently can be used to infer unmeasured of difficult to measure quantities such biomass, substrate or product concentrations.

Different techniques have been proposed until now for on-line estimation in bioreactors. Extensive reviews on this subject have been presented. These techniques are always based on either an empirical or first-principles models of the process. The bioprocesses are highly nonlinear and operate within a wide range of operating regimes. Typically, during fermentation, the microbial species continuously undergo physiological changes contributing to highly nonlinear dynamics of the process. Many researchers define these physiological changes as physiological states, phases, or operating regimes of the microbial population. In every phase, the cell population expresses stable characteristics and behaves in a fairly linear manner. Attempting to model or control the process using global approach would be a challenging task and only lead to suboptimal results.

A better approach would be to decompose the process into its phases and model as well as control each phase locally. In this study, the nonlinear dynamics modelling methodology is adopted for the purposes of bioprocess supervision. This objective can be achieved by using either a data-driven models or local models derived from first principles. From the different data-driven empirical models, the multi modelling strategy

Corresponding Author: Azzeddine Amri, Department of Electronics, Faculty of Engineering, University of Annaba, BP. 12, 23000, Annaba, Algeria

represents an effective way to model the bioprocess. As a data-driven empirical models, we introduce a Local Model Network (LMN) based on decomposition of the process into its operating phases. The overall model consists of multiple linear local models of the outputs from the local models. Phase identification for model development is performed using fuzzy clustering.

The rest of this study is organized as follows. Firstly, we give an overview of the local modelling strategy. Secondly, the whole process of structure identification is described. Then, the hybrid learning algorithm for tuning the parameters efficiently is described. Experimental results concerning a simulated process illustrate the validity of the proposed approach. Finally, some concluding remarks are given in the last section.

LOCAL MODEL NETWORKS

The local model networks are hybrid models which allow the easy integration of a priori knowledge, as well as the ability to learn from available data to model the underlying complex nonlinear relationships. The basic philosophy behind this modelling strategy is to partition the input domain into multiple subsets. Such local representations include the modular networks and fuzzy systems (Johansen and Foss, 1997; Takagi and Sugeno, 1985). The locality property can be used to make models more interpretable and computationally efficient. Assume we have a complex nonlinear Multi-Input and Single-Output (MISO) system where $x = [x_1, x_2, ..., x_n]^T \in X \subset \Re^n$ is the vector of input variables and $y \in Y \subset \Re^m$ is the vector of output variables.

In the multi-input and multi-output LMN network given in Fig. 1 the overall output is defined as:

$$\hat{y}(x) = \sum_{i=1}^{H} f_i \phi_i \Big/ \sum_{i=1}^{H} \phi_i \tag{1}$$

where $I = 1,2,...,c$ and $l = 1,2,..., n$;

$$\phi_i = \prod_{l=1}^{n} \exp\left\{-(x_l - c_{il})^2 / (\sigma_{il})^2\right\} \tag{2}$$

Here, we assume that $c_{il} \in X_l$, $\sigma_{il} > 0$ and $f_l \in Y$ where X_l and Y are the variation domains of the input x_l and output y, respectively. The local linear models are defined as

$$f_i \equiv f_i(x, \theta_i) = \left[1 x^T\right] \cdot \theta_i \tag{3}$$

If the centres and standard deviations of the validity functions are known, the estimation of the local linear model parameters is θ_i a linear optimisation problem. The parameters can therefore be evaluated by employing linear least squares optimisation algorithms. In the following subsection, we consider the learning of parameters of the local models for a given model structure, i.e., we are estimating θ for an a priori given set of c, σ.

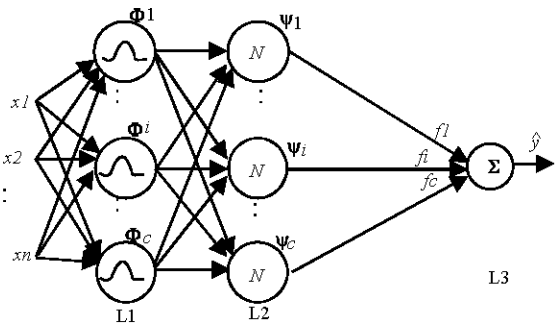

Fig. 1: Architecture of the Local Model Network (LMN)

HYBRID LEARNING SCHEME

The learning scheme is a two stage algorithm. Firstly, a coarse model that roughly approximates the underlying input-output relationship is determined by a clustering process. Secondly, parameter optimisation procedure is performed for a better tuning of the initial structure. In principle, once an appropriate structure is identified, the learning task can be accomplished by any suitable training algorithm such as the standard Back-Propagation Algorithm (BPA). However, because of slow convergence speed of pure BPA, in the following a more efficient training method, namely the combination of gradient descent with least squares optimisation procedure will be used.

Structure identification by clustering: Assume that we have N input-output samples, a regression matrix and an output matrix are constructed

$$X = \left[x_1, \cdots, x_N\right]^T, \quad Y = \left[y_1, \cdots, y_N\right]^T \tag{4}$$

Let these vectors be partitioned into clusters, each represented by the centre vector. c_i Denote by $U = W \in R^{c \times N}$ the partition matrix of the elements u_{ik} representing the membership degrees of the data vectors x_k ($k = 1, ..., N$). The fuzzy clustering algorithms search for a partition matrix and cluster centers such that the objective function J_m is minimized

$$J_m = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^m d^2(x_k, c_i) \tag{5}$$

subject to

$$\sum_{i=1}^{c} u_{ik} = 1 \tag{6}$$

For $(k = 1, \cdots, N)$. The parameter *m* controls the fuzziness of the clusters (typically *m*=2). The function $d(x_k, c_i)$ measures the distance between the data vector and the centre of the *i*th cluster. One of the most efficient clustering algorithms, able to take account different shapes of the clusters, is the Gustafson-Kessel (G-K) algorithm, extending the c-means algorithm (Bezdek, 1981) by using scaled metric norm for the distance (Gustafson and Kessel, 1979)

$$d^2(x_k, c_i) = (x_k - c_i)^T M_i (x_k - c_i) \tag{7}$$

Where is a positive definite matrix adjusted to the actual shape of the *i*th cluster. The matrix is defined as follows:

$$M_i = \sqrt[n]{\det(F_i)} F_i^{-1} \tag{8}$$

Where F_i is the cluster covariance matrix

$$F_i = \frac{\sum_{k=1}^{N} u_{ik}^m (x_k - c_i)(x_k - c_i)^T}{\sum_{k=1}^{N} u_{ik}^m} \tag{9}$$

For each input , using the Lagrangian multiplier method, for *m* > *1*, local minimum of Eq. 8 was demonstrated if and only if

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(m-1)}} \tag{10}$$

$$c_i = \frac{\sum_{k=1}^{N} u_{ik}^m x_k}{\sum_{k=1}^{N} u_{ik}^m} \quad \forall i \tag{11}$$

The algorithm iterates until $\|U^t - U^{t-1}\| \le \varepsilon$ for two succeeding iterations. The special case of G-K algorithm is developed by Bezdek (1981). In c-means, the scattering matrix is fixed to unity (M =1) and is not adapted during the learning. It means that the distances are calculated in a simple way as $d^2(x_k, c_i) = (x_k - c_i)^T (x_k - c_i)$.

Parameter optimisation procedure: The parameters obtained by the identification procedure can be optimized or fine tuned by a variant of gradient descent optimisation techniques. This is achieved by an iterative two stage forward-backward optimisation algorithm. In the forward stage, with the validity functions being constant, the local models f_i; *i = 1, ..., c* are identified by solving a least squares problem. Then, in the backward stage, the functional models are fixed and the parameters of the validity functions $c_{il}, \sigma_{il}, i = 1, \cdots, c; l = 1, \cdots, n$ are updated by an effective nonlinear Gradient-Descent (GD) optimisation technique, which requires the computation of the derivatives of the objective function to be minimized with respect to the parameters c_{il} and σ_{il}.

The optimisation algorithm uses a variable step learning rates. Given a set, $D = \{(x^p, d^p)\}_{p=1}^{N}$ such that $x^p \in X \subset \Re^n, d^p \in Y \subset \Re$ the objective is to find the system $\hat{y}(x^p)$ in the form of (1), such that the Sum of Squared Error (SSE) function

$$E = \frac{1}{2} \sum_{x^p \in D} (\hat{y} - d^p)^2 \tag{12}$$

is minimized. The problem is reduced to the adjustment of the f_i and the mean c_{il} and variance σ_{il} of the model's validity functions, so that the SSE is minimized.

Now it can be seen that the network output $\hat{y}$ and hence E, depends on c_{il} and σ_{il} only through ϕ_i where $\hat{y}, f_i, b$ and Ψ_i are represented by the following equations:

$$\hat{y} = \sum_{i=1}^{H} f_i \psi_i \tag{13}$$

$$\psi_i = (\phi_i / b) \text{ and } b = \sum_{i=1}^{H} \phi_i \tag{14}$$

Derivatives of E w.r.t c_{il} and σ_{il}:

$$\frac{\partial E}{\partial c_{il}} = \frac{\partial E}{\partial \phi_i} \frac{\partial \phi_i}{\partial c_{il}} = \frac{\partial E}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \phi_i} \frac{\partial \phi_i}{\partial c_{il}} \tag{15}$$

$$\frac{\partial E}{\partial \sigma_{il}} = \frac{\partial E}{\partial \phi_i} \frac{\partial \phi_i}{\partial \sigma_{il}} = \left(\frac{\partial E}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \phi_i} \right) \frac{\partial \phi_i}{\partial \sigma_{il}} \tag{16}$$

Finally, the results of the chain rules are written as follows:

$$\frac{\partial E}{\partial c_{il}} = A.\left\{ 2.\phi_1.(x_1 - c_{il}) / (\sigma_{il})^2 \right\} \tag{17}$$

$$\frac{\partial E}{\partial \sigma_{il}} = A.\left\{ 2.\phi_1.(x_1 - c_{il})^2 / (\sigma_{il})^3 \right\} \tag{18}$$

with. $A = (\hat{y} - d) \cdot (f_i - \hat{y}) / b$

Estimation of the local models: As stated previously, the coefficients of the local models can be estimated using the least squares method on the prediction targets *Y*. Once an initial structure is defined (i.e. and) and local models as in (3), so the learning problem is a straightforward application of linear regression techniques to find parameters θ_which best fit the data. Staking the data into matrices, we get the following regression model:

$$Y = \Gamma\theta^* + \varepsilon \tag{19}$$

where Γ is the design matrix, the rows of which are defined by

$$\varphi_k = \left[\phi_1(x_k)\left[1x_k^T\right], \cdots, \phi_c(x_k)\left[1x_k^T\right]\right] \tag{20}$$

So that the design matrix Γ, vector of output measurement *Y* and errors ϵ_are

$$\Gamma = \left(\varphi_1^T, \cdots, \varphi_N^T\right)^T,\ Y = \left(y_1, \cdots, y_N\right)^T$$
$$\varepsilon = \left(\varepsilon_1, \cdots, \varepsilon_N\right)^T$$

The standard least squares criterion for this estimation problem is

$$J(\theta) = \frac{1}{N}\left(Y - \Gamma\theta\right)^T\left(Y - \Gamma\theta\right) \tag{21}$$

and the Moore-Penrose pseudo inverse of Γ, Γ+is used to estimate the weights:

$$\hat{\theta}_{LS} = \Gamma^+ Y = \left(\Gamma^T\Gamma\right)^{-1}\Gamma^T Y \tag{22}$$

The computation of the pseudo inverse uses the Singular Value Decomposition (SVD) to decompose any $N \times p$ matrix , such that $\Gamma = USV^T$ and the pseudo inverse of Γ is:

$$\Gamma^+ = VS^+U^T$$

Then, the solution of the regression problem (6) can be calculated:

$$\hat{\theta}_{LS} = VS^{-1}U^T Y \tag{23}$$

In this way, the learning is global because it is based on assumption that all the parameters _would be learned in a single regression operation and the local models cooperate to solve the regression task. An alternative to global learning which is less prone to these disadvantages is to locally estimate the parameters of each of the local models as defined in Eq. 3 independently. This is achieved using a set of local estimation criteria for the *i*th local model

$$J_i(\theta_i) = \frac{1}{N}\left(Y - \Gamma_i\theta_i\right)^T W_i\left(Y - \Gamma_i\theta_i\right) \tag{24}$$

where $i = 1, \ldots, c$. W_i is an $N \times N$ diagonal weighting matrix defined as:

$$W_i = \mathrm{diag}(\phi_i(x_1), \cdots, \phi_i(x_N)) \tag{25}$$

Now, the criteria J_i is minimized by the locally Weighted Least Squares (WLS) estimate of the local model parameters vector. In matrix terms, now we have

$$\hat{\theta}_{WLS} = \left(\hat{\theta}_{WLS,1}, \cdots, \hat{\theta}_{WLS,c}\right)^T$$
$$\hat{\theta}_{WLS,i} = \left(\Gamma_i^T W_i \Gamma_i\right)^{-1} W_i^T W_i Y,\quad i = 1, \cdots, c \tag{26}$$

where Γ_i is a $N \times (n+1)$ submatrix of Γ corresponding to the *i*th local model.

SIMULATION STUDIES

In this subsection, the effectiveness of the proposed approach is demonstrated through the modelling of a gluconic acid batch fermentation bioprocess. This process was chosen since a nonlinear state-space model which could be used to simulate it is available in the literature (Shimizu, 1996; Takagi and Sugeno, 1985). The modelling is based on decomposition of the process into its different operating regimes. The model consists of multiple linear models, one for each regime and its output is the interpolation of the outputs from the local models. Regime identification is performed automatically by fuzzy clustering. The model is described as:

$$\begin{aligned}
\frac{dx_1}{dt} &= \mu_m \frac{x_1 x_4 x_5}{K_s x_5 + K_0 x_4 + x_4 x_5} \\
\frac{dx_2}{dt} &= V_L \frac{x_1 x_4}{K_L + x_4} - 0.9082 K_p x_2 \\
\frac{dx_3}{dt} &= K_p x_2 \\
\frac{dx_4}{dt} &= -\frac{1}{Y_s}\mu_m \frac{x_1 x_4 x_5}{K_s x_5 + K_0 x_4 + x_4 x_5} - 1.011 V_L \frac{x_1 x_4}{K_L + x_4} \\
\frac{dx_5}{dt} &= K_{La}\left(x_5^* - x_5\right) - 0.09\, V_L \frac{x_1 x_4}{K_L + x_4} \\
&\quad \frac{1}{Y_0}\mu_m \frac{x_1 x_4 x_5}{K_s x_5 + K_0 x_4 + x_4 x_5}
\end{aligned} \tag{27}$$

Table 1: Parameters for process model

K_{lg}	K_L	V_L
180 h^{-1}	12.80 litre	8.3 mg UOD^{-1} h^{-1}
K_p	K_o	K_S
0.64 h^{-1}	0.0055 g $litre^{-1}$	2.5 g $litre^{-1}$
Y_S	K_o	μ_m
0.375 UOD mg^{-1}	0.890 UOD mg^{-1}	0.39 h^{-1}

Table 2: Initial conditions and noise level used in simulations

Process number	Cell concentration	Glucose concentration	Noise level (%)
1	0.4	40	2
2	0.4	50	6
3	0.5	40	3
4	0.5	50	5
5	0.43	47	3

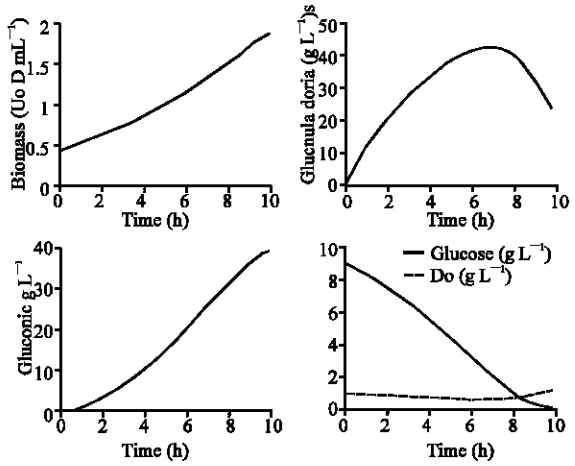

Fig. 2: A process simulation characterized by three phases. Phase 1: High DO and glucose concentration. Phase 2: Low DO and glucose concentration is decreasing. Phase 3: High DO concentration and low glucose concentration

where

x_1 is the biomass concentration, UOD mL^{-1};
x_2 is the gluconolactine concentration, g L^{-1};
x_3 is the gluconic acid concentration, g L^{-1};
x_4 is the glucose concentration, g L^{-1};
x_5 is the dissolved oxygen concentration, g L^{-1}
x_5 is the concentration of oxygen in liquid in equilibrium with gas phase, g L^{-1} (the value 0.00685 g L^{-1} is used).

The parameters values for this nominal model are shown in Table 1. To create simulated data that reflects different operating conditions, initial states of cell and glucose concentrations were chosen from the intervals (0.4 0.5) and (40 50) Table 2 shows initial conditions as well as noise levels under which these simulations were performed. Note that the model is used here only to simulate the process and provide access to variables, which would normally be monitored online. For most bioprocesses, such models do not exist, which is the reason the type of work describe here is necessary.

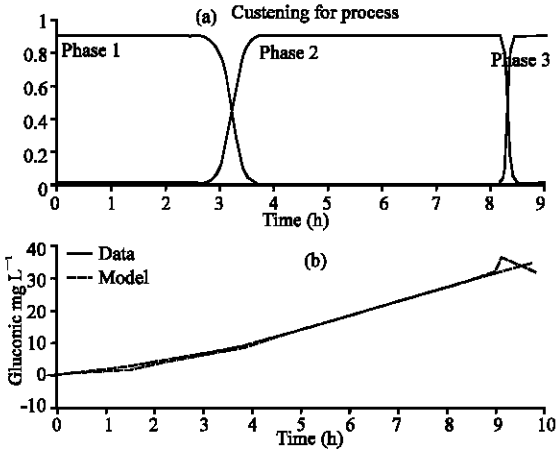

Fig. 3: (a) Identification of process phases (b) Estimation of the product concentration

For the gluconic acid batch fermentation process and according to Foss *et al* (1993), the fermentation consists in three phases. At the beginning of the batch, there is a relatively high concentration of both DO and glucose. In the intermediate phase, there is a relatively low concentration of DO and concentration of glucose is decreasing. During the final phase, there is a low glucose concentration and high DO concentration. Thus, the fuzzy clustering is used to cluster the data into $c = 3$ clusters. The two variables, glucose and DO are used as inputs.

Figure 2 shows a simulation of the gluconic acid fermentation process. Figure 3 gives fuzzy clustering results for the first three processes in Table II (used as training data). As expected, the smooth transitions identified by clustering are process dependent due to the different initial conditions and noise levels. It is worth to notice that the relationships between glucose and DO concentration are in accordance with those identified by expert knowledge (Takagi and Sugeno, 1985).

Figure 3(a) gives phase prediction of the Local Model Network (LMN) for process 1. With each new time sample of glucose and DO concentrations, the LMN predicts accurately to what degree the fermentation is in phase 1, 2 or 3.

In order to demonstrate how the LMN can be used as a "soft sensor" enabling the product estimation, the product concentration is considered as the output. Figure 3(b) shows the estimated product for the process 1 when the local models are estimated by a global least squares.

CONCLUSION

A new hybrid optimisation approach for Local Model Networks (LMN) was applied to the identification of the

global dynamics of bioprocesses. The technique is based on the decomposition of the process into its operating regimes by a well sounded clustering algorithm and multi-linear models. The method was effectively applied to the estimation of product for multiple simulations on a gluconic acid batch fermentation performed at various initial conditions and noise level.

REFERENCES

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function, Plenum Press, New York.

Foss, B.A., T.A. Johansen, A.V. Sorensen, 1993. Nonlinear predictive control using local models applied to a batch process, Control Eng. Practice, 3: 389-396.

Gustafson, D.E. and W.C. Kessel, 1979. Fuzzy clustering with a fuzzy covariance matrix, in Proc. IEEE. CDC. San Diago, CA, pp: 761-766.

Isermann, R., 1997. Supervision, fault-detection and fault-diagnosis methods. An introduction, Control Eng. Practice, 5: 639-652.

Johansen, T.A., B.A. Foss, 1997. Operating regime based process modelling and identification, Computers chem., Eng., 21: 159-176.

Rolf Isermann, 1998. On Fuzzy Logic Applications for Automatic Control, Supervision and Fault Diagnosis, IEEE . Trans. Syst., Man, Cybern., 28: 221-235.

Shimizu, K., 1996. A tutorial review on bioprocess systems engineering, Comput. Chem. Eng., 20: 915-941.

Takagi, T. and M. Sugeno, 1985. Fuzzy identification of systems and its applications to modeling and control, IEEE. Trans. Sys., Man, Cybern., 15: 116-132.