Journal of Zhejiang University SCIENCE
ISSN 1009-3095
http://www.zju.edu.cn/jzus
E-mail: jzus@zju.edu.cn

On-line estimation of concentration parameters in fermentation processes*

XIONG Zhi-hua (熊志化)†, HUANG Guo-hong (黄国宏), SHAO Hui-he (邵惠鹤)

(*Institute of Automation, Shanghai Jiaotong University, Shanghai 200030, China*)

†E-mail: zhxiong@sjtu.edu.cn

Received June 7, 2004; revision accepted Sept. 6, 2004

Abstract: It has long been thought that bioprocess, with their inherent measurement difficulties and complex dynamics, posed almost insurmountable problems to engineers. A novel software sensor is proposed to make more effective use of those measurements that are already available, which enable improvement in fermentation process control. The proposed method is based on mixtures of Gaussian processes (GP) with expectation maximization (EM) algorithm employed for parameter estimation of mixture of models. The mixture model can alleviate computational complexity of GP and also accord with changes of operating condition in fermentation processes, i.e., it would certainly be able to examine what types of process-knowledge would be most relevant for local models' specific operating points of the process and then combine them into a global one. Demonstrated by on-line estimate of yeast concentration in fermentation industry as an example, it is shown that soft sensor based state estimation is a powerful technique for both enhancing automatic control performance of biological systems and implementing on-line monitoring and optimization.

Key words: Gaussian processes (GP), Expectation maximization (EM), Multiple models, Soft sensor, Yeast concentration, Fermentation processes

doi:10.1631/jzus.2005.B0530 **Document code:** A **CLC number:** Q815

INTRODUCTION

Saccharomycete, closely linked with human life, is the only microbe that has been utilized to the extent of over a million tons in the world, and is one of the modelled biological bacteria mostly used in modern genetic engineering. Since there are no effective online instruments for measuring the production or resultant concentration, which is a very important quality index in fermentation process, soft sensors (Martin, 1997; Tham *et al*., 1991; Assis and Filho, 2000; Salgado *et al*., 2004; Andres-Toro *et al*., 2004) provide a convenient solution to eliminate the problem. And the key problem is modelling.

Most conventional industrial process models are global, where the industrial process is assumed to be fully characterized by a single model. However, because of multiple variables, serious nonlinear and multiple work modes, many real industrial process are too complex to be described by a single model. Furthermore, single model method not only reduces the estimate precision, but also increases the computational complexity (Li *et al*., 2001).

Multiple model (MM) approaches (Murray-Smith and Johansen, 1997; Babuska, 1998; Johansen and Babuska, 2002) to the empirical modelling of nonlinear systems have been of interest for many years, and have been widely used in the last few years. Mixtures of Gaussian processes (GP) (Tresp, 2001; Shi *et al*., 2002; 2003; Rasmussen and Ghahramani, 2002) have appeared in various forms.

The rest of this article is organized as follows. In Section 1 we discuss how to set the mixture of GP based soft sensor. Section 2 presents the parameter estimation of the above mixture model, which in-

* Project (No. 2002AA412010) supported by the National High-Tech Research and Development Program (863) of China

cludes Gath-Geva (GG) clustering algorithm (Gath and Geva, 1989), expectation maximization (EM) (Bilmes, 1998) and weighted least-squares (LS) estimation. Following that, we present the real-world simulation of yeast concentration derived from a fermentation process as an example (Section 3); followed by some discussions in Section 4.

MIXTURE OF GP BASED SOFT SENSOR

Briefly, multiple models involve several local models, with the global model output being a combination of local models using an interpolation technique.

The construction of a soft sensor is described below. Given the sample $\{\boldsymbol{x}_k, y_k\}_{k=1}^N$, where N is the size of data samples, $\boldsymbol{x}_k \in R^n$ are inputs of soft sensor model, and $y_k \in R$ are the corresponding desired targets. Suppose that the input vector for a test case is $\boldsymbol{x}$ and the targets are scalar. From this training set, we wish to establish a learning model of the dependency of the targets on the inputs with the objective of making accurate predictions of $\hat{y}$ for previously unseen values of $\boldsymbol{x}$.

The operating regime based model (Murray-Smith and Johansen, 1997) of the system is formulated as:

$$\hat{y} = \sum_{i=1}^{c} \phi_i(\boldsymbol{x})(\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} + b_i) \tag{1}$$

where $\phi_i(\boldsymbol{x})$ is the validity function for the ith operating regime and $\theta_i=[\boldsymbol{a}_i^{\mathrm{T}}\ b_i]^{\mathrm{T}}$ is the parameter vector of the corresponding local linear model.

The available data samples are collected in matrix $\boldsymbol{Z}$ formed by concatenating the input data matrix $\boldsymbol{X}$ and the output vector $\boldsymbol{y}$

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \boldsymbol{x}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_N^{\mathrm{T}} \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad \boldsymbol{Z}^{\mathrm{T}}=[\boldsymbol{X}\ \boldsymbol{y}]$$

Each observation thus is an n+1 dimension column vector $\boldsymbol{z}_k=[x_{1,k},\ \ldots,\ x_{n,k},\ y_k]^{\mathrm{T}}=[\boldsymbol{x}_k^{\mathrm{T}},\ y_k]^{\mathrm{T}}$. Through clustering, the data set $\boldsymbol{Z}$ is partitioned into c cluster. The result is a fuzzy partition matrix $\boldsymbol{U}=[\mu_{i,k}]_{c\times N}$, whose element $\mu_{i,k}$ represents the degree of membership of the observation $\boldsymbol{z}_k$ in cluster i.

Denote $p(\eta_i)$ the unconditional cluster probability normalized such that $\sum_{i=1}^{c} p(\eta_i) = 1$, given by the fraction of the data that it explains. $p(\boldsymbol{z}|\eta_i)$ is the domain of influence of the cluster, and will be taken to be multivariate Gaussian $N(\boldsymbol{v}_i, \boldsymbol{F}_i)$ in terms of a mean $\boldsymbol{v}_i$ and covariance matrix $\boldsymbol{F}_i$.

$$p(\boldsymbol{z} \mid \eta) = \sum_{i=1}^{c} p(\boldsymbol{z}, \eta_i) = \sum_{i=1}^{c} p(\boldsymbol{z} \mid \eta_i) p(\eta_i) \tag{2}$$

where the $p(\boldsymbol{z}|\eta_i)$ distribution generated by the ith cluster is represented by the Gaussian function

$$p(\boldsymbol{z} \mid \eta_i) = \frac{1}{(2\pi)^{\frac{n+1}{2}}\sqrt{|\boldsymbol{F}_i|}} \exp\left(-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{v}_i)^{\mathrm{T}}(\boldsymbol{F}_i)^{-1}(\boldsymbol{z}-\boldsymbol{v}_i)\right) \tag{3}$$

In this paper, we propose to use the Gath-Geva (GG) clustering method (Gath and Geva, 1989; Hoppner *et al*., 1999; Abonyi *et al*., 2002) algorithm instead of the widely used Gustafson-Kessel (GK) method, because with the GG method, the parameters of the univariate membership functions can be derived directly from the parameters of the clusters. Through GG clustering, the $p(\boldsymbol{z})=p(\boldsymbol{x}, y)$ joint density of the response variable y and the regressor $\boldsymbol{x}$ is modelled as a mixture of c multivariate n+1 dimensions Gaussian functions. And the conditional density $p(y|\boldsymbol{x})$ is also a mixture of Gaussian models. Therefore, the prediction of $\hat{y}$ based on unseen value $\boldsymbol{x}$ can be formulated as:

$$\hat{y} = E(y \mid \boldsymbol{x}) = \int y p(y \mid \boldsymbol{x}) \mathrm{d}y = \frac{\int y p(y, \boldsymbol{x}) \mathrm{d}y}{p(\boldsymbol{x})}$$

$$= \sum_{i=1}^{c} \frac{\left[[\boldsymbol{x}^{\mathrm{T}}\ 1]\boldsymbol{\theta}_i\right] p(\boldsymbol{x} \mid \eta_i) p(\eta_i)}{p(\boldsymbol{x})} = \sum_{i=1}^{c} p(\eta_i \mid \boldsymbol{x})\left[[\boldsymbol{x}^{\mathrm{T}}\ 1]\boldsymbol{\theta}_i\right] \tag{4}$$

Here, $\boldsymbol{\theta}_i$ is the parameter vector of the local models and $p(\eta_i|\boldsymbol{x})$ is the probability that the ith Gaussian component is generated by the regression vector $\boldsymbol{x}$:

$$p(\eta_i \mid \boldsymbol{x}) = \frac{\dfrac{p(\eta_i)}{(2\pi)^{n/2}\sqrt{\left|\boldsymbol{F}_i^{xx}\right|}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{v}_i^x)^{\mathrm{T}}(\boldsymbol{F}_i^{xx})^{-1}(\boldsymbol{x}-\boldsymbol{v}_i^x)\right)}{\sum_{i=1}^{c}\dfrac{p(\eta_i)}{(2\pi)^{n/2}\sqrt{\left|\boldsymbol{F}_i^{xx}\right|}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{v}_i^x)^{\mathrm{T}}(\boldsymbol{F}_i^{xx})^{-1}(\boldsymbol{x}-\boldsymbol{v}_i^x)\right)} \quad (5)$$

where $\boldsymbol{F}_i^{xx}$ is obtained by partitioning the covariance matrix $\boldsymbol{F}_i$ as follows

$$\boldsymbol{F}_i = \begin{bmatrix} \boldsymbol{F}_i^{xx} & \boldsymbol{F}_i^{xy} \\ \boldsymbol{F}_i^{yx} & \boldsymbol{F}_i^{yy} \end{bmatrix} \quad (6)$$

where $\boldsymbol{F}_i^{xx}$ is the $n\times n$ sub-matrix containing the first n rows and columns of $\boldsymbol{F}_i$, $\boldsymbol{F}_i^{xy}$ is an $n\times 1$ column vector containing the first n elements of the last column of $\boldsymbol{F}_i$, $\boldsymbol{F}_i^{yx}$ is an $1\times n$ row vector containing the first n elements of the last row of $\boldsymbol{F}_i$, and $\boldsymbol{F}_i^{yy}$ is the last element in the last row of $\boldsymbol{F}_i$.

The mixture of Gaussian processes defined by Eq.(4) and Eq.(5) is in fact a kind of operating regime based model (1) where the validity function is chosen as $\phi_i(\boldsymbol{x})=p(\eta_i|\boldsymbol{x})$.

PARAMETER ESTIMATION

Through a linear transformation of the input variables (Kim *et al*., 1998), the antecedent partition can be accurately captured and no decomposition error occurs. Unfortunately, the resulting model is not transparent as it is hard to interpret the linguistic terms defined on the linear combination of the input variables. To form an easily interpretable model that does not rely on transformed input variables, a new clustering algorithm is proposed based on the expectation maximization (EM) (Bilmes, 1998). The clusters obtained by GG clustering are multivariate Gaussian functions. The alternating optimization of these clusters is identical to the EM identification of the mixture of these Gaussian models when the fuzzy weight exponent m=2.

The EM algorithm is widely used for parameter estimation of mixture of models, in particular mixture of Gaussian models. The basics of EM are as follows. Suppose we know the observed values of a random variable $\boldsymbol{z}$ and wish to model the density of $\boldsymbol{z}$ using a model parameterized by η. EM obtains parameter estimate $\hat{\eta}$ which maximize the likelihood $L(\eta)=p(\boldsymbol{z}|\eta)$ of the data. The EM assumes that this estimation is intractable and that the values of a missing or hidden random variable $\boldsymbol{h}$ would make the problem more tractable. Let $p(\boldsymbol{z}, \boldsymbol{h}|\eta)$ denote the joint probability of $\boldsymbol{z}$ and $\boldsymbol{h}$ parameterized by η. It is assumed that $\boldsymbol{z}$ and $\boldsymbol{h}$ are such that maximizing the complete data likelihood $L_c(\eta)=p(\boldsymbol{z}, \boldsymbol{h}|\eta)$ is more tractable than maximizing $L(\eta)$. However, the values of $\boldsymbol{h}$ are not known. The EM algorithm tackles this problem by iteratively generating a probability over the values $\boldsymbol{h}$ and estimating the parameters which maximize the expected value of $L_c(\eta)$ with respect to $\boldsymbol{h}$.

The optimal $\boldsymbol{\theta}_i$ parameter vector of the local model can be obtained as:

$$\min_{\boldsymbol{\theta}_i}\frac{1}{N}(\boldsymbol{y}-\boldsymbol{X}_e\boldsymbol{\theta}_i)^{\mathrm{T}}\boldsymbol{\Phi}_i(\boldsymbol{y}-\boldsymbol{X}_e\boldsymbol{\theta}_i) \quad (7)$$

where $\boldsymbol{X}_e=[\boldsymbol{X}\ 1]$ denotes the extended regression matrix obtained by adding a unitary column to $\boldsymbol{X}$, and $\boldsymbol{\Phi}_i$ is a matrix having the membership degrees on its main diagonal:

$$\boldsymbol{\Phi}_i = \begin{bmatrix} \mu_{i,1} & 0 & \cdots & 0 \\ 0 & \mu_{i,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_{i,N} \end{bmatrix} \quad (8)$$

The weighted least-squares estimate of the consequent parameters is given by

$$\boldsymbol{\theta}_i = (\boldsymbol{X}_e^{\mathrm{T}}\boldsymbol{\Phi}_i\boldsymbol{X}_e)^{-1}\boldsymbol{X}_e^{\mathrm{T}}\boldsymbol{\Phi}_i\boldsymbol{y} \quad (9)$$

EXPERIMENTAL RESULTS

Important parameters are required to be computed or monitored for real-time use in fermentation processes. However, we tend to meet problems for lack of online instruments. The simulation experiment presented in the article is derived from the corresponding off-line assay value and the historical data from SIEMENS PCS in Hubei Angel Yeast Co., Ltd. It takes 16 h for each batch, and the assay value of

yeast concentration will be obtained per hour, which means that we can get 17 pairs data samples during each batch. We can easily find that the current method of determining the concentration is discrete sampling and characteristic of time-delay. According to techniques' analysis, the secondary variables that are already available are as follows: pH value, liquid level, quality of O_2 and volume ratio of alcohol. The four variables are the input of soft sensor model for estimating the yeast concentration. From the soft sensor, we can get the estimation of yeast concentration online and continuously.

When we specify the cluster c=3 and 204 (17×12) samples for the test, the generalized mean square error (MSE) is 0.6173. The results for the estimated output of the soft sensor and the actual assay value are shown in Fig.1. The actual output is the dotted line and estimate output is the solid one.

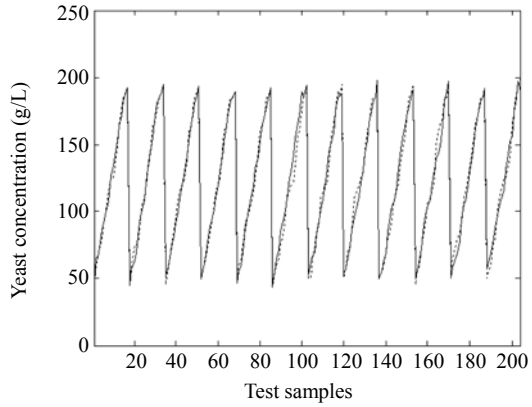

Fig.1 Results for the predictive output and the actual output

Through a series of experimental simulations, we found that if we increase the value of c, we will get less generalized error at the cost of more computational time and vice versa. The results for both generalized MSE and computational time with the change of c are shown in Table 1. The computer used for all these simulations was a PIV 1.5 GHz PC with 256 MB RAM and MATLAB version 6.5 (R13) running under a Windows 2000 operating system.

Table 1 Series of results for the change of c

c	MSE	Time (s)
2	0.7071	0.16
3	0.6173	0.23
4	0.5237	0.32

From the above result, we can conclude that the proposed soft sensor can potentially replace the analyzer or be used as a backup for automatic running since the analyzer is expensive and for doing the same job offline. And we can also realize online optimization and economic running of the fermentation process via the estimated output of the above soft sensor.

DISCUSSION

Software sensors which provide on-line estimation of unmeasurable process variables from available sensors, are powerful tools in bioprocess monitoring and control. Their performance depends on both the measurement quality delivered by the sensor and the associated estimation algorithm.

Gaussian processes are probabilistic kernel machines and moderately simple to implement and use without loss of performance compared with other kernel methods. Now GP is becoming popular in the community of kernel machines (Seeger, 2004; Kocijan *et al*., 2003; Brahim-Belhouari and Bermak, 2004). The main drawback of GP is heavy computational burden, which is greatly alleviated by the introduction of multiple models method as shown in this article. And the multiple models method accords with the real-world situation of fermentation processes.

The proposed soft sensor can be implemented via online computation which can be employed in fermentation processes for optimal running and on-line monitoring. Furthermore, we believe the proposed method opens new possibilities for applying kernel methods to potential fields.

References

Abonyi, J., Babuska, R., Szeifert, F., 2002. Modified Gath-Geva fuzzy clustering for identification of Takagi-Sugeno fuzzy models. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, **32**(5):612-621.

Andres-Toro, B., Giron-Sierra, J.M., Fernandez-Blanco, P., Lopez-Orozco, J.A., Besada-Portas, E., 2004. Multiobjective optimization and multivariable control of the beer fermentation process with the use of evolutionary algorithms. *Journal of Zhejiang University SCIENCE*, **5**(4):378-389.

Assis, A.J., Filho, R.M., 2000. Soft sensors development for on-line bioreactor state estimation. *Computers & Chemical Engineering*, **24**(2-7):1099-1103.

Babuska, R., 1998. Fuzzy Modeling for Control. Kluwer Academic Publishers, Boston.

Bilmes, J.A., 1998. A Gentle Tutorial of the EM Algorithm and Its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. ICSI TR-97-021, U.C. Berkeley, USA.

Brahim-Belhouari, S., Bermak, A., 2004. Gaussian process for nonstationary time series prediction. *Computational Statistics & Data Analysis*, **47**(4):705-712.

Gath, I., Geva, A.B., 1989. Unsupervised optimal fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **7**:773-781.

Hoppner, F., Klawonn, F., Kruse, R., Runkler, T., 1999. Fuzzy Cluster Analysis–Methods for Classification, Data Analysis and Image Recognition. John Wiley and Sons.

Johansen, T.A., Babuska, R., 2002. On Multi-Objective Identification of Takagi-Sugeno Fuzzy Model Parameters. Proceedings of IFAC World Congress, Barcelona, Spain, p.847-860.

Kim, E., Park, M., Kim, S., Park, M., 1998. A transformed input-domain approach to fuzzy modeling. *IEEE Transactions on Fuzzy Systems*, **6**(4):596-604.

Kocijan, J., Girard, A., Banko, B., Murray-Smith, R., 2003. Dynamic Systems Identification with Gaussian Processes. Proceedings of 4th Mathmod Conference, Vienna, Int. Association for Mathematics and Computers in Simulation, p.776-784.

Li, N., Li, S., Xi, Y., 2001. A multiple model approach to modeling based on LPF algorithm. *Journal of Systems Engineering and Electronics*, **12**(3):64-70.

Martin, G., 1997. Consider soft sensors. *Chemical Engineering Progress*, **7**:66-70

Murray-Smith, R., Johansen, T.A., 1997. Multiple Model Approaches to Nonlinear Modeling and Control. Taylor & Francis, London, UK.

Rasmussen, C.E., Ghahramani, Z., 2002. Infinite Mixtures of Gaussian Process Experts. Advances in Neural Information Processing Systems 14, MIT Press, p.881-888.

Salgado, A.M., Folly, R.O.M, Valdman, B., Valero, F., 2004. Model based soft-sensor for on-line determination of substrate. *Applied Biochemistry and Biotechnology*, **113**(1-3):137-144.

Seeger, M., 2004. Gaussian Processes for Machine Learning. Technical Report, Department of EECS, University of California at Berkeley, USA.

Shi, J.Q., Murray-Smith, R., Titterington, D.M., 2002. Hierarchical Gaussian Process Mixtures for Regression. Technical Report TR-2002-107, Department of Computing Science, University of Glasgow, Scotland, UK.

Shi, J.Q., Murray-Smith, R., Titterington, D.M., 2003. Bayesian regression and classification using mixtures of Gaussian processes. *International Journal of Adaptive Control and Signal Processing*, **17**:1-16.

Tham, M.T., Montague, G.A., Morris, A.J., Lant, P.A., 1991. Soft-sensors for process estimation and inferential control. *Journal of Process Control*, **1**:3-14.

Tresp, V., 2001. Mixtures of Gaussian Processes. Advances in Neural Information Processing Systems 13, MIT Press, p.654-660.