Multi-Phase Support Vector Regression Soft Sensor for Online Product Quality Prediction in Glutamate Fermentation Process

1,2 Rongjian Zheng and 1 Feng Pan

1 Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi, China
2 Faculty of Automation, Huaiyin Institute of Technology, Huaian, China

Abstract: Glutamate fermentation is inherently nonlinear, multi-phase and an aerobic fermentation process. As long measurement delays and expensive apparatus cost, on-line measurement of the product concentration is not necessarily available. The present fermentation process monitoring and quality prediction involve manual interpretation of highly informative, however, the concentrations of substrates, biomass and products are only low frequency off-line measurements. In this paper, we propose a novel Multi-Phase Support Vector Regression (MPSVR) based soft sensor model for online quality prediction of glutamate concentration. The glutamate fermentation process can be divided into a sequence of five phases by detecting the trend variation events (also termed as singular points or inflection point) of online measured O2 in the exhaust gas, the Inflection Point (IP) are easily identified through combining Moving Window (MW) with Pearson Correlation Coefficient (PCC). For each estimation phase, SVR soft sensor model are constructed and their performance is evaluated against fermentation data in a 5 L fermenter. The efficiency of the proposed soft sensor model for online product quality prediction has been demonstrated to be superior compared to that of reported techniques in a 5 L glutamate fermentation process.

Keywords: Multi-phase Quality Prediction, Glutamate Fermentation, Support Vector Regression, Soft Sensor

Introduction

Glutamate is commercially one of the most important amino acids produced mainly by fermentation process, its fermentative production amount exceeds 2.2 million tons annually (Xiao et al., 2006; Khan et al., 2005). Glutamate are widely used for human and animal alimentation, as ingredients of pharmaceutical products, agrochemicals and some other industrial derivatives (Pal et al., 2016).

Like other fed-batch processes, glutamate fermentation process requires a sophisticated operator involvement. In addition to run-to-run modifiability, this could result in abnormal situation, in which any deviation from desired operating regimes could lead to product quality change. This provides a strong incentive for automating operation supervision. During manual operation, a human operator is commonly responsible for setting points to regulatory controllers, performing control actions, process supervising and taking remedial measures when an abnormal condition is detected (Muthuswamy and Srinivasan, 2003). The batch-to-batch change may root in the variation in the raw material quality or the seed culture variations. Typically, during manipulation, the product quality and batch performance are desired to be controlled and monitored by way of offline laboratory assays of concentrations of the product, substrates and biomass, which might take up to 2 h. These laboratory assays are high investment costs, manpower inputs and time consuming, are obtained at low frequencies and hence, may not necessarily acquire timely information about the fermentation status of the batch. Online measurements that are easily acquirable include temperature, pH, dissolved oxygen, agitation speed, exhaust CO2 and O2, whereas these measurements do not show direct state of the process (Doan et al.,...
provide analysis of product concentration, nutrient compositions and other metabolites (Alford, 2006). However, these methods suffer from the aforementioned disadvantages. Another class of approach has focused on using the routinely available online data to qualitatively identify fermentation phases, this class most common methods include process knowledge, phase analysis and the process data. A formal framework for inferring process trends from the online variables was exploited (Cheung and Stephanopoulos, 1990) and applied to fermentation process data (Stephanopoulos et al., 1997; Doan et al., 2007). Another method for detecting phase change uses singular points detection based only on online measurements (Maiti et al., 2009; Régis et al., 2008). The reader can find more multi-phase analysis methods in (Yao and Gao, 2009; Camacho et al., 2008; Doan et al., 2007; Luo et al., 2016), which give different kinds of phase identification methods. Knowledge based phase identification fails when the process prior knowledge is not enough to divide processes into phases legitimately and difficult to customise for diverse fermentation processes. Process analysis based phase identification works well when certain required process features are known. Finally, process data based methods carry out phase partition by detecting variation in process data. Compared with the aforementioned two methods, data-driven methods are easier to perform because of their data-driven property. However, their phase partition results obtained by data-driven methods may or may not always consistent with actual operation phases (Sun et al., 2011; Luo et al., 2016).

In this paper, a novel phase partition method and a Multi-Phase SVR (MPSVR) modeling strategy are presented for online estimation and prediction of glutamate concentration. Glutamate fermentation process goes through a number of phases based on serial cell growth, substrate uptake and product formation. Besides, the production of glutamate is an aerobic process, the glutamate fermentation performance and the metabolic flux distribution are affected drastically by the concentrations of dissolved oxygen in the liquid phase in fermentor or oxygen concentrations in exhaust gas (Golobič and Gjerkeš, 1999; Xiao et al., 2006), the fed-batch process can be divided into 5 phases based on the detection of Inflection Point (IP) by online measured O2 in the exhaust gas, the IP are easily identified through combining Moving Window (MW) with Pearson Correlation Coefficient (PCC). The phase division result agrees well with actual fermentation process, it depends only on on-line measurements and fermentation processes can be easily automated to work. Then, for each estimation phase, SVR soft sensor models are designed for online prediction of glutamate concentration and their performance is evaluated against glutamate fermentation data in a 5 L fermenter. Also, a
Materials and Methods

Experimental Methods

C. glutamicum S9114 was used in the present study, it was kept by the laboratory of industrial biotechnology, Jiangnan University. The fermentation conditions and the compositions of medium were the same as those ahead reported (Zhang et al., 2005; Xiao et al., 2006; Ding et al., 2012; Cao et al., 2013; Zheng and Pan, 2016).

C. glutamicum S9114 was cultured for glutamate production in a 5 L bioreactor. PH was controlled in 7.0-7.2 by feeding 25% (v/v) ammonia water. The O2 and CO2 concentrations in the exhaust gas were measured on-line by a gas analyzer (LKM2000, Lokas Co., Korea). Dissolved Oxygen (DO) was controlled at various levels by automatically or manually controlling the agitation speed, O2 Uptake Rate (OUR) and CO2 Evolution Rate (CER) were computed accordingly. Temperature was controlled at about 32°C. Electronic balances, which was assumed to be identically distributed and independent. When working with the online measured O2 data, this postulation may not be valid. Whereas, this assumption is more likely to be satisfied if time average is used as the response variable (Li et al., 2013; Hess et al., 2001).

Support Vector Regression

The SVR aims to provide a nonlinear mapping function of the feature space. The empirical risk can be defined as following:

\[ R_{emp}(f) = \frac{1}{n} \sum_{i=1}^{n} \Theta_{\varepsilon}(y_i, f(x_i)) \]  

where, \( \Theta_{\varepsilon} \) denote the \( \varepsilon \)-insensitive loss function and is described as follows:

\[ \Theta_{\varepsilon}(y_i, f(x)) = \begin{cases} |f(x) - y| - \varepsilon, & \text{if } |f(x) - y| \geq \varepsilon \\ 0, & \text{otherwise} \end{cases} \]

Then, an optimum hyper plane can be acquired by minimizing the empirical risk subject to the following constraints:

\[ \min \ w^T w + C \sum_{i=1}^{n} (\zeta_i^+ + \zeta_i^-) \]

where, \( C \) is the regularization parameter. The constraint conditions of this optimization problem are as follows:

\[ \begin{cases} y_i - w^T \phi(x_i) - b \leq \zeta_i^+, & i = 1,...,n \\ -y_i + w^T \phi(x_i) + b \leq \zeta_i^-, & i = 1,...,n \\ \zeta_i^+, \zeta_i^- \geq 0, & i = 1,...,n \end{cases} \]

By solving the above describing optimization problem, the coefficients of Equation (1) can be got as below:

\[ w = \sum_{i=1}^{n} (\beta_i^+ - \beta_i^-) \phi(x_i) \]

where, \( \beta_i \) is the Lagrangian coefficients. The SVR regression function can be described as below:

\[ f(x) = \sum_{i=1}^{n} (\beta_i^+ - \beta_i^-) K(x, x_i) + b \]

where, \( K(x, x_i) \) denotes the Kernel function, it can be described as follows in the feature space:

\[ K(x, x_i) = \phi(x) \cdot \phi(x_i) \]

In this study, the RBF Kernel function is utilized, it can be expressed as:

\[ K(x, x_i) = e^{-\frac{|x - x_i|^2}{\sigma^2}} \]

Then, an optimum hyper plane can be acquired by utilizing the function. With the help of hyper plane, the training data were divided into two linear separable subsets with maximum separation distance. As it is, SVR is an optimizing problem with objective function is:

\[ \min_{w,\epsilon} R_{\epsilon}(w, \epsilon) = 0.5w^T w + C \sum_{i=1}^{n} (\zeta_i^+ + \zeta_i^-) \]

where, \( C \) is the regularization parameter. The constraint conditions of this optimization problem are as follows:

\[ \begin{cases} y_i - w^T \phi(x_i) - b \leq \zeta_i^+, & i = 1,...,n \\ -y_i + w^T \phi(x_i) + b \leq \zeta_i^-, & i = 1,...,n \\ \zeta_i^+, \zeta_i^- \geq 0, & i = 1,...,n \end{cases} \]
After taking time averages, the regression coefficients are estimated by using the least squares method. Thus the estimates of the intercept and slope are given by:

\[ b = \frac{\sum (t - \bar{t})(Y - \bar{Y})}{\sum (t - \bar{t})^2} \]  

(11)

\[ a = \bar{Y} - b\bar{t} \]  

(12)

where, \( \bar{t} \) and \( \bar{Y} \) are the arithmetic means for \( t_i \) and \( Y_i \).

Since glutamate fermentation is a continuous process, \( b \) changes continuously and smoothly relative to \( Y \) (\( O_2 \)), we then employ Pearson Correlation Coefficient (PCC) to identify the Inflection Point (IP) of variation trend \( b \). The PCC was developed by Karl Pearson from a related idea introduced by Galton in the late 19th century, it is a well-established measure of correlation and has range of -1 (perfect but negative correlation) to +1 (perfect correlation) with 0 denoting the short of a relationship (Adler and Parmryd, 2010).

PCC \( r \) is given as follows:

\[ r = \frac{\sum (b_i - \bar{b})(T_i - \bar{T})}{\sqrt{\sum (b_i - \bar{b})^2 \sum (T_i - \bar{T})^2}} \]  

(13)

where, \( \bar{b} \) and \( \bar{T} \) is the means for \( b_i \) and \( T_i \). Fig. 1 gives the phase partition procedure.

As the glutamate fermentation continues, the fermentation characteristics will change over time, the Moving Window (MW) technology can be used for tracking \( O_2 \) changes. It is essential to discard the old data and add the newest data to the model. In fact, the most challenging lies in the selection of the window length, enough information can be included to detect the real change trend of parameter via setting suitable window size. If the window size is too small, the variation trend will be disturbed by the process noises and phases recognition may be lagged when the window size is too long (Yuan et al., 2016). The window length of \( t \) is set to be 180 by trial-and-error, while the window length of \( b \) is set to be 60. Fig. 2 shows the recognition process of inflection point, the profile of the online measured \( O_2 \) is shown in Fig. 2A. As shown in Fig. 2B, the \( r \) changes sharply at transition point of two different phases.

Glutamate fermentation experiences about five phases: Growth phase, transition phase, initial and middle production phases, late production stage, end of fermentation phase. Nevertheless, according to offline measurements and analysis, the growth of the cells passes through 5 phases: Lag phase, accelerate phase, decelerate phase, stationary phase and decline phase.

A key challenge for exploiting such a multi-phase partition is finding a appropriate \( b \) in normal fermentation process, analysis and experiments of the fermentation data for all the normal batches revealed that \( b_1 \) is set to \( b_1<0 \), while \( (b_1/10)<b_2<0 \) and \( b_2>b_4 \). \( b_1 \), \( b_2 \), \( b_3 \), \( b_4 \) and \( b_5 \) denotes phase 1, phase 2, phase 3, phase 4 and phase 5, respectively.

**Quality Prediction Based on Multi-phase SVR**

The prediction model based on Multi-Phase SVR (MPSVR) for glutamate concentration is shown in Fig. 3.
Fig. 2. Recognition process of inflection point

Fig. 3. Prediction model structure of multi-phase SVR

Table 1. Input and output variables of the prediction SVR

| Input variables | Output variable       |
|-----------------|-----------------------|
| t               | Glutamate concentration |
| T               |                       |
| pH              |                       |
| DO              |                       |
| AG              |                       |
| OUR             |                       |
| CER             |                       |
| AR              |                       |

In soft sensors, the secondary variables are used to act as the inputs of the soft sensor model and the primary variables such as glutamate concentration is employed to act as the output of soft sensor model.

The online measurement variables, such as fermentation time (t), fermentation Temperature (T), pH, Dissolved Oxygen Concentration (DO), Agitation Rate (AG), O₂ Uptake Rate (OUR), CO₂ Evolution Rate (CER) and ammonia water consumption Rate (AR) were chose as input variables. Table 1 shows the inputs and outputs configuration of the prediction model. There are 10 batches data for modeling and testing, 9 batch used for training and the other one for testing (Zheng and Pan, 2016).

In order to evaluate the prediction performance, Root Man Square Error (RMSE) and coefficient of determination ($R^2$) are computed as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$  \hspace{1cm} (14)

$$R^2 = \frac{\left( n \sum_{i=1}^{n} \hat{y}_i y_i - \left( \sum_{i=1}^{n} \hat{y}_i \right) \left( \sum_{i=1}^{n} y_i \right) \right)^2}{\left( n \sum_{i=1}^{n} \hat{y}_i^2 - \left( \sum_{i=1}^{n} \hat{y}_i \right)^2 \right) \left( n \sum_{i=1}^{n} y_i^2 - \left( \sum_{i=1}^{n} y_i \right)^2 \right)}$$  \hspace{1cm} (15)

where, $y_i$ is the offline measured value; $\hat{y}_i$ is the prediction value and $n$ is the number of samples.

Results and Discussion

The phase partition results of glutamate fermentation process are shown in Fig. 4, comparisons of multi-phase partition and offline measurements method are shown in Fig. 5. It can be seen that the phase partition results are well consistent with the physiological states of the real process by offline measurements and analysis method. The phase partition approach can be applied to other fermentation processes where minimal process information concerning phase shifts is available. Moreover, it can be used as a guide online in making a decision about the timing of off-line sampling. Note that in Fig. 4, range of variables has been stretched processing for clear show of each curve.

Once the phase recognition of fermentation process has been determined, the proposed MPSVR prediction models are therefore built for online product quality prediction of glutamate fermentation base on the phase partition results. For comparison, the SVR, NN and Multi-Phase NN (MPNN) models are equally constructed to predict production concentration with the same fermentation batches.
For the NN, Functions included to automatically train and test standard 1-layer neural networks using the MATLAB functions "train" and "sim". The number of hidden neurons is cross validated. For the SVR, standard support vector implementation for regression and function approximation using the Libsvm toolbox. Prediction values comparisons of different models are shown in Fig. 6.

| Models   | RMSE   | R²     |
|----------|--------|--------|
| NN       | 7.2582 | 0.9694 |
| SVR      | 3.7979 | 0.9942 |
| MPNN     | 7.8957 | 0.9145 |
| MPSVR    | 2.8653 | 0.9908 |

Fig. 4. Phase partition results of glutamate fermentation process

Fig. 5. Comparisons of multi-phase partition and offline measurements method

Fig. 6. Prediction values comparisons of different models. (A) Overall simple models, (B) Multi-phase models

Fig. 7. Prediction residuals comparisons of different models. (A) Overall simple models, (B) Multi-phase models
Prediction residuals and residuals boxplot comparisons of different models are shown in Fig. 7 and 8. It can be seen that the best prediction result has been obtained by the multi-phase SVR (MPSVR) model based soft sensor. The maximum residuals reaches 7.32 at fermentation time of 12 h. The reason is that it is decelerate phase, the bacteria activity becomes stronger at fermentation time of 12 h. Meanwhile, From Fig. 6-8, we can see that the MPSVR based prediction model exhibits better tracking performance than that by NN, SVR and MPNN models. Quality prediction results of different models are shown in Table 2, as can be seen that MPSVR has got the best prediction performance, since it has the minimum Root Mean Square Error (RMSE) 2.8653 among all prediction models, however, RMSE of SVR model is 3.7979. Furthermore, R^2 of the MPSVR is 0.9908.

**Conclusion**

In this paper, a soft sensor prediction model based on Multi-Phase SVR (MPSVR) was proposed for online quality prediction in glutamate fermentation process. The states detection and partition is based on the detection of inflection points of the online measured O2 variations using Moving Window (MW) and Pearson Correlation Coefficient (PCC). The phase identification result agrees well with the physiologic phase changing of the glutamate fermentation process. Hence this method can help to automatically control and optimize the glutamate fermentation process. As the phase identification merely lies on the process data, it can be applied to other fermentation processes where minimal process knowledge regarding phase shifts is available. Besides, the phase partition approach can be used as a guide online in making decisions about the timing of off-line sampling.

The performance comparisons among SVR, NN, Multi-phase NN and Multi-phase SVR are implemented. Through glutamate fermentation process case study, the feasibility of the proposed MPSVR soft sensor has been conformed, the MPSVR model exhibits excellent prediction performance, it can provide effective information for monitoring and operation of glutamate fermentation process.

**Acknowledgement**

The Authors would like to appreciate the National Natural Science Foundation of China under Grant No.61273131 for financial support.

**Author’s Contributions**

Rongjian Zheng: Designed the research plan, participated in all experiments and writing the manuscript. Feng Pan: Provided guidance, acquisition and analysis of data.

**Ethics**

The authors declare that they have no conflict of interest.

**References**

Acuña, G., C. Ramirez and M. Curilem, 2014. Software sensors for biomass concentration in a SSC process using artificial neural networks and support vector machine. Bioprocess Biosyst. Eng., 37: 27-36. DOI: 10.1007/s00449-013-0925-3

Adler, J. and I. Parmryd, 2010. Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient. Cytometry Part A, 77: 733-742. DOI: 10.1002/cyto.a.20896

Alford, J.S., 2006. Bioprocess control: Advances and challenges. Comput. Chem. Eng., 30: 1464-75. DOI: 10.1016/j.compchemeng.2006.05.039

Camacho, J., J. Picó and A. Ferrer, 2008. Multi-phase analysis framework for handling batch process data. J. Chemometrics, 22: 632-643. DOI: 10.1002/cem.1151
Were, K., D.T. Bui, Ø.B. Dick and B.R. Singh, 2015. A comparative assessment of support vector regression, artificial neural networks and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecological Indicators, 52: 394-403. DOI: 10.1016/j.ecolind.2014.12.028

Xiao, J., Z.P. Shi, P. Gao, H.J. Feng and Z.Y. Duan et al., 2006. On-line optimization of glutamate production based on balanced metabolic control by RQ. Bioprocess Biosyst. Eng., 29: 109-117. DOI: 10.1007/s00449-006-0059-y

Yao, Y. and F.R. Gao, 2009. A survey on multistage/multiphase statistical modeling methods for batch processes. Annu. Rev. Control, 33: 172-183. DOI: 10.1016/j.arcontrol.2009.08.001

Yuan, X.F., Z.Q. Ge and Z.H. Song, 2016. Spatio-temporal adaptive soft sensor for nonlinear time varying and variable drifting processes based on moving window LWPLS and time difference model. Asia-Pac. J. Chem. Eng., 11: 209-219. DOI: 10.1002/apj.1957

Zhang, C.Y., Z.P. Shi, P. Gao, Z.Y. Duan and Z.G. Mao, 2005. On-line prediction of products concentrations in glutamate fermentation using metabolic network model and linear programming. Biochem. Eng. J., 25: 99-108. DOI: 10.1016/j.bej.2005.03.012

Zheng, R.J. and F. Pan, 2016. Soft sensor modeling of product concentration in glutamate fermentation using gaussian process regression. Am. J. Biochemistry Biotechnol., 12: 179-187. DOI: 10.3844/ajbbsp.2016.179.187