Biochemical Engineering Journal 112 (2016) 178–185

Contents lists available at ScienceDirect

Biochemical Engineering Journal

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

Process state classification of fed-batch fermentation based on process variables analysis

Jan Mareš*, Jaromír Kukal, Pavel Hrnčiřík, Jan Náhlík

Department of Computing and Control Engineering, University of Chemistry and Technology, Prague, Technická 5, 166 28 Prague 6, Czech Republic

ARTICLE INFO

Article history:
Received 9 July 2015
Received in revised form 29 March 2016
Accepted 20 April 2016
Available online 26 April 2016

Keywords:
Bioprocess monitoring
Process state classification
Dissolved oxygen
Expert system
Fed-batch fermentation

ABSTRACT

The success of fermentation processes operated in the fed-batch mode depends, among other factors, on appropriate substrate feeding. Overfeeding should be avoided because even slightly higher concentrations may result in an inhibition or even poisoning of the microbial culture when toxic substrates are used. Therefore, bioprocess monitoring and control play a key role. This paper introduces a new bioprocess state classification methodology combining expert knowledge and automatic signal analysis, suitable for on-line application. A fed-batch cultivation of the strain *Pseudomonas putida* KT2442 grown on octanoic acid was used as a model process. The classification was performed in two steps – a manual classification done by an expert, and a subsequent automatic classification using the results obtained by the manual classification. The manual classification strategy was based on the analysis of time profiles of selected process variables related to substrate feeding, such as dissolved oxygen tension and substrate feeding rate. Three process states were recognized – normal feeding, overfeeding and underfeeding. Ridge regression was then applied to the data results of the manual classification in order to design an automatic classification strategy for easier on-line use. This strategy can distinguish between the normal and other feeding states, using a limited number of on-line measurable output variables and their time fluctuations.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Several approaches are available for control of fed-batch bioprocesses. Control methods include simple exponential feeding strategies [1], closed loop control strategies [2] and alternative control strategies relying on fuzzy logic [3,4] and neural networks [5,6]. However, bioprocess performance indicators like productivity and yield are significantly influenced by the type of metabolism used by the cultivated microorganism for the processing of substrates. Therefore, the design of bioprocess control strategies should not be limited just to the issue of cell environment control but should ideally also aim at the control of the cell physiology itself. As a result, strategies applying this type of control operate in closed loop in respect to the cell state. Obviously, for this type of control it is essential to be able to identify the physiological state of the cultivated microorganism in on-line regime. This issue has been addressed by Konstantinov and Yoshida, who introduced a control concept referred to as physiological state control [7], in which the online process state classification plays a key role.

Since the introduction of the concept of physiological state control there have been several successful on-line applications, especially for the control and monitoring of the most common model bioprocess—a baker's yeast fed batch cultivation process. Siimes et al. developed an object-oriented real-time fuzzy-knowledge-based system for fault diagnosis and control of the baker's yeast production process. One of the systemís functionalities was the on-line detection of the current process phase (metabolic state) from five process variables – four measured online and one off-line. In all, five process phases (metabolic states) were recognized – two within the mixed oxidative-fermentative and three within the purely oxidative metabolism [8]. Shimizu et al. applied a fuzzy system to recognize on-line the physiological states of the yeast culture with respect to cell growth (three states) and ethanol production (two states). Fuzzy membership functions were constructed from the error vectors as defined by a macroscopic elemental balance equation. As input variables, five conversion rates (oxygen, carbon dioxide, ammonia, ethanol, glucose) were taken [9].

A more recent alternative approach to the task of on-line physiological state recognition of fermentation processes is represented

* Corresponding author.
E-mail addresses: jan.mares@vscht.cz (J. Mareš), jaromir.kukal@fjfi.cvut.cz (J. Kukal), pavel.hrncirik@vscht.cz (P. Hrnčiřík), jan.nahlik@vscht.cz (J. Náhlík).

http://dx.doi.org/10.1016/j.bej.2016.04.019
1369-703X/© 2016 Elsevier B.V. All rights reserved.

Nomenclature

List of Symbols

pH	pH of the cultivation broth in the bioreactor
DO	Dissolved oxygen tension in the cultivation broth in the bioreactor (% sat.)
Fm	Carbon source (100% octanoic acid) feeding rate to the bioreactor (mL/min)
S1	Dissolved oxygen fluctuations—case 1
S2	Dissolved oxygen fluctuations—case 2
Δt	Sampling time (min)
DDO	1st derivative of the filtered DO signal (% sat./min)
D(m,n)	Matrix of historical process data (m—measured samples, n—process variables)
$\mathbf{D}_T$	Matrix of historical process data—training set
$\mathbf{D}_V$	Matrix of historical process data—verification set
s	State vector (results from process state classification done by a human expert)
$\mathbf{s}_T$	State vector—training set
$\mathbf{s}_V$	State vector—verification set
H	Time window length (number of samples)
x	Statistical sample (vector of H subsequent values) of a given process variable
s	Standard deviation
IQR	Interquartile range
E	Sample mean
Q_1	First quartile
Q_2	Sample median
Q_3	Third quartile
MAD_1	Median of absolute differences from the sample median
MAD_2	Mean of absolute differences from the sample median
MAD_3	Median of absolute differences from the sample mean
MAD_4	Mean of absolute differences from the sample mean
ADIFF	First quartile of absolute pair differences
X	Data matrix
y	Vector of states
$\mathbf{y}_{calc}$	Vector of calculated states
λ	Weighting coefficient
λ_{opt}	Optimal weighting coefficient
err	Error between calculated and observed classifications
err_V	Error between calculated and observed classifications—verification set
err_T	Error between calculated and observed classifications—training set
$err^0{}_T$	Error between calculated and observed classifications—baseline
$err^*{}_T$	Error between calculated and observed classifications—corresponding to λ_{opt}

by the use of metabolic flux analysis (MFA) in various forms, combining the information from extracellular measurements with those obtained from intracellular data. The potential of MFA for on-line monitoring of physiological states has been demonstrated by the studies of Takiguchi et al. and Henry. In the first study [10] an on-line state recognition scheme was proposed for a fed-batch bacterial fermentation process (lysine production). Based on this scheme combining an intracellular metabolic model of 11 metabolic reactions with an on-line measurement and estimation by an Extended Kalman Filter of 4 extracellular conversion rates this system was capable of on-line recognition of 3 different physiological states. A similar approach, yet using a more extensive intracellular metabolic model (40 metabolic fluxes and 46 metabolic components), has been used by Henry [11] for the on-line monitoring of the physiological state of a mammalian cell perfusion process (human embryo kidney cells). An adaptive variant of MFA has been even applied for on-line monitoring of a mixed microbial culture in a polyhydroxybutyrate production process in a sequencing batch reactor. The proposed monitoring system was capable of on-line identification of shifts between 3 different physiological states [12]. An important limitation of MFA for on-line monitoring of physiological states is, however, that the number of stoichiometric and uptake/production rate measurements is often insufficient to observe all relevant intracellular metabolic pathways. A possible solution to this problem is the use of ^{13}C-labeled tracers (isotope-based MFA) in combination with advanced sampling and analytical methods [13,14]. Additional complications arise when this approach is to be applied for the monitoring of bioprocesses operated at a non-steady state, *i.e.* batch or fed-batch. The number of reported solutions for this problem is still relatively low, *e.g.* Antoniewicz [15] have proposed an alternative approach combining isotopomer spectral analysis and MFA for the on-line monitoring of metabolic shifts during a bacterial fed-batch fermentation (*E. coli*). However, despite the aforementioned studies, the use of MFA for on-line physiological state monitoring in industrial bioprocesses is still rather limited due to the inherent complexity of underlying metabolic models and the need for multiple on-line and off-line measurements.

For bioprocesses where historical process data, including information on physiological or process states in general, are available, data-driven classification-based approaches provide a suitable alternative for the design of on-line physiological state monitoring systems. This approach is applicable for the general task of on-line process state monitoring including the area of Fault Detection and Isolation (FDI). For FDI of continuous chemical and biochemical processes in steady state, several different classification-based techniques (*e.g.* discriminant partial least squares, Fisher discriminant analysis and support vector machines) were successfully applied to a number of case studies [16–19]. On the other hand, in the case of inherently unsteady-state batch and fed-batch processes the number of reported applications of the classification-based FDI techniques is relatively low, and these are predominantly off-line approaches, because the proposed process state classification tools need to be trained on entire faulty batches. As a consequence, during fault state classification the monitored batch needs to be either completed or the missing future values of corresponding process variables must be estimated for the remainder of the batch duration. In a recent study by Gins et al. [20], a more appropriate approach for on-line classification-based FDI of batch and fed-batch processes has been proposed, combining a PCA-based fault state detection (12 input process variables from a simulated industrial-scale bioprocess) and two classification methods – k-Nearest-Neighbors (k-NN) or Least Squares Support Vector Machine (LS-SVM). Using simulated historical fed-batch process data a set of on-line fault state classifiers were trained and validated. The resulting classification-based fault state diagnosis was then performed within a sliding time window of the length of up to 10 process data samples.

All of the aforementioned approaches to on-line bioprocess state monitoring typically use the whole set of available on-line variables or even a combination of multiple on-line and off-line measurements (*e.g.* MFA-based techniques), which causes a number of data-related complications (*e.g.* high dimensionality, complex intercorrelations), often requiring the application of advanced data pretreatment methods like PCA or ICA. However, for industrial applications it would be sufficient in many cases to have an on-line bioprocess state monitoring system focused on a specific, narrowly

Download English Version:

https://daneshyari.com/en/article/2721

Download Persian Version:

https://daneshyari.com/article/2721

Daneshyari.com