Keywords: SBR process, MPCA, situation assessment, fuzzy tool, classification.

Abstract

In this work, a combination between Multivariate Statistical Process Control (MSPC) and an automatic classification algorithm is developed to application in Waste Water Treatment Plant. Multiway Principal Component Analysis is used as MSPC method. The goal is to create a model that describes the batch direction and helps to fix the limits used to determine abnormal situations. Then, an automatic classification algorithm is used to situation assessment of the process.

1 Introduction

The treatment of wastewater has become one of the most important environmental topics. Wastewater treatment is an important part of maintaining the highest possible quality of natural water resources. With the new regulations for quality monitoring of Wastewater Treatment Plants (WWTP), directive 91/271/CEE [8], it is necessary to introduce new technology for control and supervision. In recent years, major advances in wastewater treatment using activated sludge have been developed. This work has been developed using a Sequencing Batch Reactor (SBR) pilot plant shown in Figure 1. The SBR has a cyclic nature, each cycle consist of several stages depending on the objective of operations. This process is highly nonlinear, time-varying and subjected to significant disturbances such equipment defects, atmospheric changes (rain), variation in the composition, among others.

In this work, a Statistical Process Control (SPC) can be used to determine abnormal situations, because the statistical model considers the correlation structure between variables [4], it is characterized by the covariance matrix, due to the existing relation between variables and process. Based on this knowledge, it is natural to consider the use of different models in different phases to achieve better results. However, when SPC is used to batch process, often false alarms are generated [5]. Fortunately, this problem can be solved using Multivariate Statistical Process Control (MSPC). MSPC compresses the multidimensional information in few latent variables that explain the variability of the measured variables, including their relationships. In this paper, a model is developed using an extension of the Principal Component Analysis method to detect abnormal batch behaviours. This extension is based on Multiway Principal Component Analysis (MPCA) for batch data [9,14]. An automatic fuzzy classification algorithm is used after MSPC, allowing a good situation assessment.

This paper is organized as follows: Operation of the SBR process is presented in section 2. The section 3 describes those MPCA methods for process monitoring. In section 4, the fuzzy classification method is presented. In section 5 a numerical example is presented by using data recorded from pilot plant SBR during a period of 4 months. Finally, future work and conclusions are summarised in section 6 and 7.

2 SBR pilot plant: influent composition and operational conditions

SBR Pilot plant is composed of a metal square reactor of 1m3. The minimum and maximum volumes of the reactor are 483 litres and 683 litres, respectively and it defines 200 litres of water to process. Wastewater is taken directly from the Cassà-WWTP (Spain) by means of a peristaltic pump and it is stored in a storage tank. Next, the wastewater is pumped to the reactor by means of another peristaltic pump (Figure 1). The SBR pilot plant carries out advanced treatment, in which the nitrogen is removed. Nitrogen removal in the SBR pilot plant is a two steps process: Nitrification: Ammonia is converted to nitrate by aerobic micro organisms. Denitrification: Nitrate is converted to N2O or nitrogen gas under anoxic (without oxygen) conditions by anoxic micro organisms. Until now, the duration of operation stages are fixed. Each batch take 8 hours and has 392 samples (obtained every 60 seconds) per variable: pH, Oxidation Reduction Potential (ORP), Dissolved Oxygen (DO) and Temperature. Each cycle is based on alternating anoxic and aerobic reaction. The anoxic period is longer than the aerobic because of increasing denitrification

This work has been developed on the basis of the research project “Development of a system of control and supervision applied to a Sequencing Batch Reactor (SBR) for the elimination of organic matter, nitrogen and phosphorus DPI2002-04579-C02-01”.

MULTIVARIATE STATISTICAL PROCESS CONTROL TO SITUATION ASSESSMENT OF A SEQUENCING BATCH REACTOR

M. Ruiz*, J. Colomer*, J. Colprim† and J. Meléndez*

*Control Engineering and Intelligent Systems (eXiT), University of Girona, Spain, e-mail: (mlruizo,colomer,quimmel)@eia.udg.es
† Laboratory of Chemical and Environmental Engineering (LEQUIA), University of Girona, Spain, e-mail: jesus@lequia.udg.es
MSPC in batch processes

Consider a typical batch run in which \( j = 1, 2, \ldots, J \) variables are measured at \( k = 1, 2, \ldots, K \) time intervals throughout the batch. Similar data will exist on a number of such batch runs \( i = 1, 2, \ldots, I \) [9]. All the data can be summarized in the \( X \) \((I \times J \times K)\) see figure 2 and 3.

MPCA is equivalent to performing ordinary PCA on a large two-dimensional \((2 - D)\) matrix constructed by unfolding the three-way. The three-way array \( X \) can be unfolded between the rearranged data matrix in the batch direction \((I \times KJ)\) (figure 2) and in direction of variable \((IK \times J)\) (figure 3)[5].

The objective of MPCA is to decompose the three-way \( X \), into a large two-dimensional matrix \( X \). It accomplishes this decomposition in accordance with the principle of PCA and separates the data in an optimal way into two parts. The noise or residual part \((E)\), which is as small as possible in a least squares sense, and the systematic part \((P)\), which expresses it as one fraction \((t)\) related only to batches and a second fraction \((P)\) related to variables and their time variation [9]. The MPCA algorithm derives directly from the NIPALS algorithm and as a result the matrix \( X \). It is the product of score vector \( t \) and loading matrices \( P\), plus a residual matrix \( E\), that is minimized in a least-squares sense as

\[
X = \sum_{r=1}^{R} t_r \otimes P_r + E = TP + E = \hat{X} + E
\]  

(2)

Abnormal behaviour of batch can be identified by projecting the batch onto the model. Control charts that are used in monitoring batch processes are generally based on the the \( Q\)-statistic and \( D\)-statistic, in which control limits are used to determine whether the process is in control or not. The assumption behind these approximate confidence limits is that underlying process exhibits a multivariate normal distribution with a population mean of zero. This is to be expected, because any linear combination of random variables, according to the central limit theorem, should toward a normal distribution.

The MPCA decomposes the three-way array where \( \otimes \) denotes the Kronecker product \((X = t \otimes P\) is \(X(i, j, k) = t(i)P(j, k)\)) and \( R\) denotes the number of principal components retained. The equation (1) is the 3-D decomposition while the equation (2) displays the more common 2-D decomposition [13].

The loading vectors \((P)\) define the reduced dimension space \((R)\) and are the directions of maximum variability. Each element of the score vectors \((t)\) corresponds to a single batch and depicts the overall variability of this batch with respect to the other batches in the database throughout the entire batch duration [9]. A few principal component can express most of the variability in the data when there is a high degree of correlation among the data \((R < < \min(I, JK))\). \( R\) is chosen such that most of the systematic variability of the process data is described by these principal components and that the residual matrix, \( E\), is a small as possible in a least-squares sense. When process measurements array \( X \) is unfolded to preserving the batch direction and variable direction there are combination of slices of matrices of different sizes, these are:

<table>
<thead>
<tr>
<th>In batch direction</th>
<th>In variable direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X ) ((I \times J \times K))</td>
<td>( X ) ((I \times J \times K))</td>
</tr>
<tr>
<td>( X ) ((I \times KJ))</td>
<td>( X ) ((IK \times J))</td>
</tr>
<tr>
<td>( T ) ((I \times R))</td>
<td>( T ) ((IK \times R))</td>
</tr>
<tr>
<td>( P ) ((KJ \times R))</td>
<td>( P ) ((J \times R))</td>
</tr>
<tr>
<td>( E ) ((I \times KJ))</td>
<td>( E ) ((IK \times J))</td>
</tr>
</tbody>
</table>

Abnormal behaviour of batch can be identified by projecting the batch onto the model. Control charts that are used in monitoring batch processes are generally based on the the \( Q\)-statistic and \( D\)-statistic, in which control limits are used to determine whether the process is in control or not. The assumption behind these approximate confidence limits is that underlying process exhibits a multivariate normal distribution with a population mean of zero. This is to be expected, because any linear combination of random variables, according to the central limit theorem, should toward a normal distribution.

The \( Q\)-statistic is a measure of the lack of fit with the established model. For batch number \( i\), \( Q_i\) is calculated as:
\[
Q_j = \sum_{j=1}^{J} \sum_{k=1}^{K} (e_{jk})^2 \approx g_{x(j)}^2
\] (3)

where \(e_{jk}\) are the elements of \(E\). \(Q\) indicates the distance between the actual values of the batch and the projected values onto the reduced space.

The D-statistic or Hotelling \(T^2\) statistic, measures the degree to which data fit the calibration model:
\[
T_i^2 = I/S - I/R \approx F_{(I,J-R)}
\] (4)

where \(S\) is the estimated covariance matrix of the scores. The \(T^2\) gives a measure of the Mahalanobis distance in the reduced space between of batch and the origin that designates the point with average batch process behaviour.

4 Fuzzy classification method

In classification purpose the Learning Algorithm for Multivariate Data Analysis (LAMDA) has been used. This method combines both, numeric and symbolic classification algorithms, taking profit of fuzzy logic and hybrid connectives. LAMDA is proposed as a classification technique to determine the current situation of the process [2,6]. The following paragraphs resumes the classification principle used in LAMDA.

One \(\hat{X}\) row (object) is a batch. This object has a number of principal components called "descriptors". These descriptors are obtained by MPCA and which are used to describe the batch. Every object is assigned to a "class" in the classification process [10]. Class \((k)\) is defined as the universe of descriptors, which characterize one set objects (batch) as pictured in figure 4.

\[\begin{array}{c}
\hat{X} \\
\hline
\end{array}\]

Figure 4: Basic LAMDA recognition methodology

MAD (Marginal Adequacy Degree) concept is a term related to how similar is one object descriptor to the same descriptor of a given class, and GAD (Global Adequacy Degree) is defined as the pertinence degree of one object to a given class as in fuzzy membership functions \(m(x)\) [1].

Classification, in LAMDA, is performed according to similarity criteria computed in two stages. FirstMADto each existing class is computed for each descriptor of an object. Second, these partial results are aggregated to get a GAD of an individual to a class [2]. The former implementation of LAMDA included a possibility function to estimate the descriptors distribution based on a “fuzzification” of the binomial probability function computed as equation 5. This approach was used in this work.

\[
MAD \left( \frac{d_i x_j}{\rho_{jk}} \right) = \rho_{jk}^{d_i x_j} (1 - \rho_{jk})^{1-d_i x_j}
\] (5)

where \(d_i x_j = \) Descriptor \(i\) of the object \(j\), \(\rho_{jk} = \) \(\rho\) of descriptor \(i\) and class \(k\).

On the other hand, GAD computation was performed as an interpolation between a \(\gamma\)-norm and a \(\gamma\)-conorm by means of the \(\beta\) parameter such that \(\beta = 1\) represents the intersection and \(\beta = 0\) means the union [7].

\[
GAD = \beta T(MAD) + (1 - \beta)S(MAD)
\] (6)

5 Results and discussion

The MPCA algorithms were applied to the three-way data array, \(X\), with dimensions 179 x 4 x 392, where \(k = 392\) is time intervals throughout the batch (samples), \(J = 4\) are measured variables and \(I = 179\) are the historical data batches. The three-way array \(X\) has been unfolded in the batch direction (I x KJ) so, the new dimensionality are (179 x 1568), with dimensionality reduction technique are obtained only (179 x 8). The model is created with eight components which explained 92.79% of the total variability see table 1.

<table>
<thead>
<tr>
<th>Principal component number</th>
<th>Eigenvalue of Cvs (%)</th>
<th>% Variance Captured This PC</th>
<th>% Variance Captured Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0000E+00</td>
<td>25.44</td>
<td>25.44</td>
</tr>
<tr>
<td>2</td>
<td>9.03E-01</td>
<td>2.03</td>
<td>27.47</td>
</tr>
<tr>
<td>3</td>
<td>8.39E-01</td>
<td>1.58</td>
<td>29.05</td>
</tr>
<tr>
<td>4</td>
<td>7.89E-01</td>
<td>1.13</td>
<td>30.18</td>
</tr>
<tr>
<td>5</td>
<td>7.49E-01</td>
<td>0.94</td>
<td>31.12</td>
</tr>
<tr>
<td>6</td>
<td>6.99E-01</td>
<td>0.85</td>
<td>31.97</td>
</tr>
<tr>
<td>7</td>
<td>6.60E-01</td>
<td>0.58</td>
<td>32.55</td>
</tr>
<tr>
<td>8</td>
<td>6.40E-01</td>
<td>0.40</td>
<td>32.95</td>
</tr>
</tbody>
</table>

Table 1: This is an example of a table caption.

To examine the process data in the reduced projection spaces, is plotted the first two principal component, dashed line is the model (see figure 5). In the score plot a group the batch are out the model. This batch corresponds of disturbance.

The SBR process is very complex so chemistry engineers and control engineers provided the types batch classification depicted in table 2. It has been used to locate and establish the relationship between classes and types batch. In this analysis were used two methods. First, the analytical methods proposed in [11] where the use of organic matter for denitrification is purposed and Second, the results preliminary of MPCA. This study creates five types of batch process. These are: Electrical fault, Variation in the composition, Equipment defects, Atmospheric changes and Normal behaviour. The normal behaviour was the most common type with higher nitrogen efficiency than legally-required effluent standards and there are 60 batches with abnormal behaviour (table 3).
Abnormal behaviour of batch is identified by projecting the batch onto the model. Figure 6 shows the $Q$ and $T^2$-statistic charts. In the $Q$ chart, only a third of the total abnormal behaviour is detected, furthermore there are 8 false alarms. The $T^2$ chart has 20 batches with abnormal behaviour (without false alarm), besides two batches are outside. These batches had electrical fault (EF). Table 4 summarise all information. In conclusion, 31 about 60 of the abnormal behaviour can be detected, 9 batches are in both charts.

SBR process is highly nonlinearly and MPCA is a dimensionality reduction lineal technique for this reason there are batches with abnormal behaviour that are not detected. LAMDA is used under a no supervised strategy. The descriptors matrix is based on eight principal components of MPCA. LAMDA classify automatically the data in eleven. The table 5 compares the classes and type batch presented in process, there are some batches are wrong classified, in this table is showing the number and percentage of data in each class. According to these results, it is possible to identify classes as equipment defects, electrical faults, atmospheric changes and variation in the composition. The classes 1,9 and 10 correspond to normal behaviour. The group 6 is atmospheric changes. The classes 3 and 11 are variation in the composition. The classes 7 and 8 are electrical fault. Finally, the classes 2, 4 and 5 are composed by different types of batch process. The class 5 is abnormal behaviour because there are atmospheric changes and equipment defects. The relation between the class and principal components is studied. The 8th component is less predominant because it does not change. It is the same into the whole classes. If only seven components are selected, the total variability will be 90.54%. It indicates that the control model can be created with seven components.
The three-way array $X$ has been unfolded in the variable direction too (IK x J). The model was developed with dimensions (70168 x 4), where MPCA squeezes in 3 principal components explaining the 95.18% of the total variability. In figure 7 a projection on the first and second component plane of the statistical model. The batches are sequentially ordered and there are 3 sections into the model. Each section corresponds to one period determinate of SBR process. Test of SBR process match the first month with monitoring; Spring is the batches developed in spring season finally summer correspond to cycles in summer season. Temperature is less contribution than others variables (-0.25 by first component) in consequence, Temperature was omitted and it was developed a new model only with 3 variables, figure 8 shown that process is highly related with environment.

Figure 7: Score plot in variable direction

Figure 8: Score plot in variable direction without temperature variable

6 Conclusions

The MSPC has been used to detect abnormal behavior in SBR process, it was projecting the data into a lower dimensional space that accurately characterizes the state of the process. In general, MSPC is a good tool for situation assessment, but it was necessary to complement it with the classification tool. This classification tool can not be used directly because the process dynamics are not taken into account. However MSPC summarises one process (variables and sampling) in 7 variables in the batch direction and 3 in the variable direction with high percentage of total information.

References