Adaptive modeling of an urban wastewater treatment plant for Monitoring Purpose

Satyam Agnihotri

Course project report for CHE-765

Chemical Engineering

McMaster University

18 Dec 2018

0.1 Abstract

Conventional PCA modeling is widespread for monitoring purposes of many industrial processes. However, the conventional method of building a PCA model using a representative training set for the process is not the best way to monitor all kinds of processes. Suitable adaptation of model is necessary for accurate monitoring of the process in cases such as slow drifting of process to a different state which can be due to various reasons such as seasonal changes, equipment lifetime, startup of a process after shutdown etc.

Application of recursive PCA techniques to adaptively model the process and raise alarms before a process upset occurs is shown in this report. Daily measurements of input and output parameters from an urban wastewater treatment plant in Barcelona were recorded and process upsets were also recognized and noted down along with the data. The data was segregated into training and monitoring sets. A comparison of fixed model PCA and adaptive PCA is shown in this report.

0.2 Introduction

The concept of adaptive modeling and related algorithms had emerged in the 1990s with significant contribution from S. Joe Qin [**cite1**, **cite2**] and others. While fixed Principle Component Analysis (PCA)and Partial Least Squares (PLS) models work well for processes where variables and their correlations are time invariant, adaptive modeling is required where variables can slowly drift to different values due to known or unknown reasons. The correlation structure between some variables may also change with time which can be taken care of only when the model is updated suitably.

In the following subsection 2.1, a brief account of monitoring using fixed PCA and monitoring using adaptive PCA is given. Subsection 2.2 has a brief account about a conventional municipal wastewater treatment plant.

0.2.1 Fixed and Adaptive PCA

Fixed PCA model refers to a PCA model built for a process on a predefined training dataset. Confidence limits for two statistical values namely the SPE (Square prediction error) and T^2 (Hotelling's T^2) are calculated and used to monitor incoming observations. SPE and T^2 are calculated for the new observation. If the observation crosses any of those limits an alarm is raised and the operator can use the model to look into variables contributing to the phenomena using a contributions plot and take preventive measure as necessary.

A few assumptions while building a fixed PCA model using a training dataset are:

(i) Most of the 'normal' operating conditions are captured within the training data. While this is true for most steady state processes which are very sensitive to abnormal operations when variables change their values by a significant amount, this is not true for a municipal wastewater treatment plant where normal operation can be achieved over a wide range of wastewater parameters (discussed in later section).

(ii) The process does not drift while in the monitoring phase. Drifting is a phenomena when the mean value of a variable in the monitoring phase has changed by a significant amount when compared to its mean in the training set. This means that the process is slowly moving towards a different state while still operating normally. False alarms are inevitable in such cases and hence these new normal operating conditions need to be incorporated in the model somehow to improve model performance.

A simple schematic of an adaptive PCA model is shown in figure 1. An initial PCA model is made using a sufficiently large training set. Subsequent observations are monitored using the SPE and T^2 statistic and the new observation is included in the training set if process operation went normally (decided using SPE, T^2 values or using the actual knowledge of product quality). A new model is then built using the new training set available now using recursive techniques. Recursive techniques are used to reduce computation load. Confidence limits for SPE and T^2 are calculated again and now the model is ready to monitor the next observation.

0.2.2 Wastewater Treatment Plant (WWTP)

Municipal WWTPs treat millions of litres of wastewater everyday coming in from many small point sources through sewer lines. The main objective of such a treatment plant is to treat this water below a certain regulation limit (determined by environmental laws) for discharge into local water bodies. Figure 2 shows the schematic of the WWTP in this study.

This is a typical configuration of any WWTP and it consists of three parts, pretreatment, primary treatment and secondary treatment. Pretreatment helps in screening out big solid particles from water. Primary treatment consists of a number of sedimentation tanks known as primary settlers which remove sedimentable particles. Secondary treatment consist of aerobic digestion of the wastewater using bacteria to degrade the organic matter present in the water. This process is called as the well known activated sludge process. The primary and secondary treatment are batch processes and happen in big tanks with volume of the order of 2000 m^3.

0.3 Database

Daily measurements of 8 input variables to primary settler and 7 input variables to secondary treatment unit were recorded for 527 days from 01/01/1990 to 30/10/1991 in an urban wastewater treatment plant in Barcelona. Corresponding output variables were also recorded but not used in this study. However, the knowledge of inconsistent output quality observations (due to abnormal operation) was available and used in this study. This dataset was obtained from UCI machine learning repository (**Dua:2017**).

Input variables to primary settler include: (i)Volume of wastewater (WW), (ii) Concentration of zinc in WW, (iii) pH of wastewater, (iv) Biological Oxygen Demand (BOD) of WW, (v) Suspended Solids (SS) in WW, (vi) Volatile suspended solids (VSS) in WW, (vii) Sediments in WW, (viii) Conductivity of WW. Similarly input variables to secondary treatment include (i) pH, (ii) BOD, (iii) COD, (iv) SS, (v) VSS, (vi) SED, (vii) COND. These variables are stacked horizontally to get matrix X. Hence the X data matrix had 15 variables and 527 observations.

0.4 Objectives

The primary objective of this study is to predict process upsets using the 16 input parameters. First, a fixed PCA model will be employed and its performance will be evaluated on a validation set (monitoring set). Next, on the same training and monitoring set, an adaptive model will be employed and results will be compared. In the following section methods used to build adaptive model is discussed in detail.

0.5 Methods

A fixed PCA model was built using first 50 normal observations using ProMV software and labeled as Model #1. Next observation was then monitored using this model #1 and if found within SPE, T^2 limits, this observation was incorporated in the training data to build an updated model. If the observation is found outside SPE or T^2 limits, it was incorporated in the training set only after the alarm is identified as a false alarm due to normal product quality.

Now, a crude way to go about updating the model would be to add the new observation in the raw data and make another model using ProMV software. Using recursive techniques to come up with the updated model is elegant and more efficient in terms of computation required. Processing large amounts of data with sampling intervals withing seconds requires such techniques to adapt the model online.

Offline model adaptation is feasible for this case because observations are one day apart, however when the training data set becomes too large to store and sampling intervals are within seconds, models need to be updated online using minimal past information required to update the model. The following subsection 5.1 talks about the mathematics behind updating the model recursively and building the new model. Subsection 5.2 describes the methods used to define the confidence limits for SPE and T^2. Also shown are the mathematical expressions used to calculate SPE and T^2 for a new observation using the current model.

0.5.1 Recursive PCA

Let $X^0 \in R^{n\times m}$ denote the raw training dataset available. This matrix becomes X after mean centering and scaling using mean vector $b \in R^{1\times m}$ and standard deviation vector $\sigma \in R^{1\times m}$. m is equal to 16 for our dataset. Let $X_1 \in R^{n+1\times m}$ denote the centered and scaled matrix corresponding to the first updated model. Similarly, $b_1 \in R^{1\times m}$ is the mean vector and $\sigma_1 \in R^{1\times m}$ is the standard deviation vector for first model update. $x^0 \in R^{1\times m}$ denotes raw data vector of a new observation. Using these notations, recursive updation of mean vector and standard deviation vector is shown below in equations (1) and (2) respectively.

$$b_{k+1} = \Big(\frac{n}{n+k}\Big)b_k + \Big(\frac{1}{n+k}\Big)x^0 \qquad \forall \quad k = 0 \ \text{ to } \ k, \qquad \text{where} \qquad b_0 = b \tag{1}$$

Let $\sigma_{1.i}$ denote the standard deviation of ith variable in the vector σ_1. Then

$$(n+k)\sigma^2_{k+1.i} = (n+k-1)\sigma^2_{k.i} + (n+k)\Delta b^2_{k+1} + (x^0_i - b_{k+1.i})^2 \qquad \forall \quad i = 1 \text{ to } 16 \tag{2}$$

$$\text{where,} \qquad \Delta b_{k+1} = b_{k+1} - b_k \tag{3}$$

Once the new mean and standard deviation vectors are calculated, we go ahead with calculating the correlation matrix recursively. Let R denote the correlation matrix corresponding to initial training matrix X, then

$$R = \frac{1}{n-1} X^T X \tag{4}$$

Let R_{k+1} denote the correlation matrix for $k+1^{th}$ model updation.

$$\begin{aligned} R_{k+1} &= \frac{1}{n+(k+1)-1} X_k^T X_k \\ &= \frac{1}{n+k} X_k^T X_k \end{aligned}$$

$$R_{k+1} = \frac{n+k-1}{n+k} \Sigma_{k+1}^{-1} \Sigma_k R_k \Sigma_k \Sigma_{k+1}^{-1} + \Sigma_{k+1}^{-1} \Delta b_{k+1} \Delta b_{k+1}^T \Sigma_{k+1}^{-1} + \frac{1}{n+k} x_{k+1}^T x_{k+1} \tag{5}$$

$$\text{where,} \qquad \Sigma_j = \text{diag}(\sigma_{j.i}, ..., \sigma_{j.m})$$

$$x_{k+1} = (x_{k+1}^0 - b_{k+1}) \Sigma_{k+1}^{-1}$$

Above equation no. (5) can be derived easily. The derivation steps are not shown in this report. The reader can refer to **cite1** for the proof. Once the correlation matrix has been calculated recursively, eigen value decomposition using 'eig' command in matlab is carried out and the highest 5 eigen values and corresponding vectors represent the new model.

0.5.2 SPE, T^2 calculation and Adaptation of Thresholds

Calculation of SPE and T^2 for a new observation using the current model is performed in the following way

$$SPE = x_k^T \big(I - P_a P_a^T\big) x_k$$
$$T^2 = x_k^T P_a \Lambda_a^{-1} P_a^T x_k$$

where, P_a is the eigen vector matrix for the current model corresponding to a highest eigen values calculated using the correlation matrix, $\Lambda_a = \mathrm{diag}(\lambda_1, ..., \lambda_a)$. The number of principle components used in this study was 5, hence $a = 5$.

Values calculated above shall be compared against the threshold values for SPE and T^2. These thresholds were updated with every new observation and the next observation was predicted to be 'normal' or 'abnormal' using the latest threshold values. SPE limit was calculated using the following equation derived by **cite4**:

$$SPE_\alpha = \theta_1 \left[\frac{\eta_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}} \tag{6}$$

$$\text{where,} \qquad \theta_1 = \sum_{j=a+1}^{m} \lambda_j \quad ; \theta_2 = \sum_{j=a+1}^{m} \lambda_j^2 \quad ; \theta_3 = \sum_{j=a+1}^{m} \lambda_j^3 \quad ; h0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2}$$

η_α is the normal deviate corresponding to upper $(1 - \alpha)$ percentile. Similarly limit for T^2 is calculated using the equation

$$T^2_{a,\alpha} = \frac{(N-1)(N+1)a}{N(N-a)} F_\alpha(a, N-a) \tag{7}$$

where, $N = n + k$ for k^{th} model updation, $a = 5$, $F_\alpha(a, N - a)$ is the inverse F-distribution function calculated in matlab using $finv(0.01, 5, N - 5)$.

0.6 Results and Discussion

A fixed model using first 50 normal observations was made and tested on the next 100 observations. Both SPE and T^2 plots for the testing set are shown in Figures 3 and 4 respectively. SPE and T^2 for the training set can be found in figure 5 and 6 (observation 1 to 50). From historical information only one observation was known to be abnormal and the rest 99 proceeded normally with good output quality of water. Since, the first 50 observations do not cover all the variation in the influent quality that can be handled by the plant normally, we see a lot of false alarms by looking at the SPE and T^2 plots of the testing set. Please refer to Table 1 at this point.

This behaviour of a large number of points falling outside limits from observation 51 to 150 can be attributed to a significant change in influent water quality and volumes. This hypotheses is confirmed by looking at the raw data plot of influent BOD (refer to Figure 7) where a general trend can be seen in the values of BOD. Variation in BOD around the mean value is quite narrow from days 1 to 50 which is increased in days 51 to 100 with almost negligible variation in the mean. Days 101 to 150 experience an increase in the mean BOD value after which the mean again starts decreasing until day 200. Other quality parameters also show rough trends over the days shown in figures 8,9,10. Since this data was taken in a chronological manner, it can be seen that water quality parameters experience a drift with time due to seasonal changes.

In order to adapt the model to new influent conditions which can be handled by the plant normally, an adaptive scheme was implemented where a new observation in the testing set (51 to 150 in this case) was monitored using the current model and if found to be operating normally was incorporated in the training set to build the new model before testing the next observation. Values of SPE and T^2 both for the training and testing set using an adaptive model are shown in figures 5 and 6 respectively. The one process upset is identified by the model along with numerous false alarms too. However the number of false alarms has decreased by 50% when compared to the fixed model case discussed above. Please refer to Table 1 for further details.

A similar comparison was done with an initial training set of 250 normal days this time. The previous study included only small amount of variation in wastewater due to less number of observations, hence a bigger dataset was chosen this time to cover most of the influent variation and check if the adaptive modeling scheme provides any advantages over fixed modeling. SPE and T^2 of the testing batch (255 to 527) for fixed and adaptive model are given in figures 11,12,13,14. Number of alarms generated and identified process upsets are listed in Table 2. 4 less false alarms were generated using the adaptive model hence it is better than using a fixed model.

0.6.1 Evolution of loadings

Evolution of loading values (adaptive model #1) for 4 variables was plotted for the first component. Refer to figures 15,16,17,18 and 19. However, a clear understanding of the trends in relation to their raw data trends could not be identified in this study.

0.7 Conclusions and Future Work

In conclusion, an adaptive modeling strategy was found to be better than fixed modeling in cases where monitoring needs to start as early as possible and a few false alarms in the initial phase of monitoring can be taken care of suitably.

Such adaptive strategies can be merged with a forgetting factor approach for processes where past data is forgotten while model building because it does not represent the process operating state anymore due to process drift. This is seen mostly in steady state processes and hence a forgetting factor was not implemented in this study.

Future work can include interpretation of the evolution of loading values with time and more robust adaptive techniques to identify the false alarms in the adaptive scheme shown in this report.