Fuzzy model-based detection of sensor faults in waste-water treatment plant

SKRJANC I.
Faculty of Electrical Engineering
University of Ljubljana
Tržaška 25, 1000 Ljubljana
SLOVENIA
Tel.: +38614768311, Fax: +38614264631

Abstract: In this paper detection of sensor faults in waste-water treatment plant by Gustafson-Kessel fuzzy clustering algorithm is discussed and presented. The main idea in the case of process monitoring by the use of fuzzy clustering algorithms is comparison between the fuzzy clusters for a normal operation regime and the current behavior. The detection of sensor faults was applied to the simulation model of the waste-water treatment plant, where the following measurements were obtained: influent ammonia concentration, dissolved oxygen concentration in the first aerobic reactor tank, temperature, dissolved oxygen concentration and the ammonia concentration in the second aerobic reactor. The results of fault detection based on fuzzy model are shown and discussed.

Key-Words: fuzzy clustering, monitoring, fault detection

1 Introduction

Process monitoring including fault detection and diagnosis based on multivariate statistical process control has been rapidly developed in recent years. Model based techniques, expert systems and pattern recognition have been widely used for fault detection [1]. The appearance of a range of new sensors and data gathering equipments has enabled data to be collected with greater frequency from most chemical processes. Many statistical techniques for extracting process information from massive data and interpreting them have been developed in various field [2, 3]. We are dealing with sensor faults detection in the simulated wastewater treatment plant. These types of plants are due to their nature subjected to daily, weekly and seasonal variation because of the temperature change, rain and varying process load. The theoretical modeling is in the case of wastewater treatment plant is a really demanding task with questionable results. Therefore, the methods of data mining are adopted for statistical proces monitoring. The process of wastewater treatment is highly nonlinear and needs to be treated in a nonlinear way. In our case we have applied a fuzzy clustering algorithm to preprocess the data. The false alarms due to the nonlinear behavior of the plant are avoided (if compared to the linear methods of detection which are generally used) by the use of fuzzy model which enables universal approximation of nonlinearities.

1.1 Fuzzy model based on Gustafson-Kessel clustering

In this section the methods and algorithms applied to the analysis of the data will be presented. The Gustafson-Kessel fuzzy clustering algorithm will be explained and the Takagi-Sugeno fuzzy model for wastewater treatment plant will be constructed and identified.

1.1.1 Gustafson-Kessel fuzzy clustering

The input data matrix is then given as $X \in \mathbb{R}^{n \times p}$. The input data vector in the time instant k is defined as $x_k = [x_{k1}, \ldots, x_{kp}]\,,\ x_k \in \mathbb{R}^p$. The set of n observations is denoted as $X = \{x_k \mid k = 1, 2, \ldots, n\}$, and is represented as $n \times p$ matrix:

$$X = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1p} \\ x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{np} \end{bmatrix} \qquad (1)$$

The main objective of clustering is to partition the data set X into c subsets, which are called clusters. The fuzzy partition of the set X is a family of fuzzy subsets $\{A_i \mid 1 \leq i \leq c\}$. The fuzzy subsets are defined by their membership functions, which are implicitly defined in the fuzzy partition matrix $U = [\mu_{ik}] \in \mathbb{R}^{c \times n}$. The i-th row of matrix U contains values of the membership function of the i-th fuzzy subset A_i of data set X. The partition matrix satisfies the following conditions: the membership degrees are real numbers from the interval $\mu_{ik} \in [0,1]\,,\ 1 \leq i \leq c,\ 1 \leq k \leq n$; the total membership of each s_k in all the clusters equals one $\sum_{i=1}^{c} \mu_{ik} = 1,\ 1 \leq k \leq n$; none of the fuzzy clusters is empty nor it contains all the data $0 < \sum_{k=1}^{n} \mu_{ik} < n,\ 1 \leq i \leq c$. This means that the fuzzy partition matrix U belongs to the fuzzy partition set which is defined as:

$$M = \{U \in \mathbb{R}^{c \times n} \mid \mu_{ik} \in [0,1]\,, \forall i,k; \sum_{i=1}^{c} \mu_{ik} = 1, \forall k;\ 0 < \sum_{k=1}^{n} \mu_{ik} < n,\ \forall i\}. \quad (2)$$

In our application the fuzzy partition matrix is obtained by applying the fuzzy c-means algorithm based on the Mahalanobis distance norm. The algorithm is based on the minimization of the fuzzy c-means functional given as:

$$J(X,U,V) = \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik}^{m})\, d_{ik}^{2}, \quad (3)$$

where U is the fuzzy partition matrix of X, the vector of cluster prototypes (centers) $V = [v_1, v_2, \ldots, v_c]\,,\ v_i \in \mathbb{R}^p$, which have to be determined, and

$$d_{ik}^{2} = (x_k - v_i)^T\, A_i\, (x_k - v_i)$$

is the inner-product distance norm, where

$$A_i = (\rho_i det\,(C_i))^{1/p}\, C_i^{-1}$$

where $\rho_i = 1, i = 1, ..., c$ and p is equal to the number of measured variables and where C_i is the fuzzy covariance matrix of the ith cluster defined by

$$C_i = \frac{\sum_{k=1}^{n} \mu_{ik}^{m}\, (x_k - v_i)\, (x_k - v_i)^T}{\sum_{k=1}^{n} \mu_{ik}^{m}}$$

This allows the detection of hyper-ellipsoidal clusters in the distribution of the data. If the data are distribution along the nonlinear hyper-surface, the algorithm will find the clusters that are local linear approximations of this hyper-surface. The cluster overlapping is defined by the parameter $m \in [1, \infty)$.

The number of clusters is defined by using the cluster validity measure or by iterative merging or insertion of the clusters. The overlapping factor or the fuzziness parameter m influences the fuzziness of the resulting partition; from the hard ($m = 1$) to the partition which is completely fuzzy ($m \to \infty$). In our approach the standard value $m = 2$ is used.

1.1.2 Gustafson-Kessel algorithm

For the given data set X, the number of clusters c, the chosen weighting exponent $m > 1$ and the algorithm termination tolerance $\epsilon_{end} > 0$ the algorithm will be the following:

- **initialization**

  Initialization of fuzzy partition matrix: $U \in M$ (randomly). Initialization of epoch: $r = 0$.

- **repeat**

  $$r = r + 1$$

  computation of the cluster centers:

  $$v_i^{(r)} = \frac{\sum_{k=1}^{n} \left(\mu_{ik}^{(r)}\right)^m x_k}{\sum_{k=1}^{n} \left(\mu_{ik}^{(r)}\right)^m},\ 1 \leq i \leq c. \quad (4)$$

  computation of the cluster covariance matrices and inner-product distance norm :

  $$C_i = \frac{\sum_{k=1}^{n} \mu_{ik}^{m}\, (x_k - v_i)\, (x_k - v_i)^T}{\sum_{k=1}^{n} \mu_{ik}^{m}}, \quad (5)$$

  $$A_i = (\rho_i det\,(C_i))^{1/p}\, C_i^{-1},\ 1 \leq i \leq c \quad (6)$$

  computation of the distance from the cluster centers

  $$d_{ik}^{2} = \left(x_k - v_i^{(r)}\right)^T A_i \left(x_k - v_i^{(r)}\right), \quad 1 \leq i \leq c,\ 1 \leq k \leq n \quad (7)$$

  update of the partition matrix:

  $$if\ d_{ik} > 0,\ \ \mu_{ik}^{(r)} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}} \quad (8)$$

- **until** $||U^{(r)} - U^{(r-1)}|| < \epsilon_{end}$

1.1.3 The fuzzy model in TS form

The fuzzy model, in Takagi-Sugeno (TS) form, approximates the nonlinear system by smoothly interpolating affine local models. Each local model contributes to the global model in a fuzzy subset of the space characterized by a membership function. We assume a set of input vectors $X = [x_1, x_2, \ldots, x_n]^T$ and a set of corresponding outputs which is defined as $Y = [y_1, y_2, \ldots, y_n]^T$.

A typical fuzzy model is given in the form of rules

$$R_i : \; if \; x_k \; is \; A_i \; then \; \hat{y}_k = \phi_i(x_k) \qquad i = 1, \ldots, c \tag{9}$$

The vector x_k denotes the input or variables in premise, and the variable $\hat{y}_k$ is the output of the model at time instant k. The premise vector x_k is connected to one of the fuzzy sets $(A_1, \ldots, A_c)$ and each fuzzy set A_i $(i = 1, \ldots, c)$ is associated with a real-valued function $\mu_{A_i}(x_k)$ or $\mu_{ik} : \mathbb{R} \rightarrow [0, 1]$, that produces the membership grade of the variable x_k with respect to the fuzzy set A_i. The functions $\phi_i(\cdot)$ can be arbitrary smooth functions in general, although linear or affine functions are normally used.

The model output in Eq. (9) is now described in closed form as follows:

$$\hat{y}_k = \frac{\sum_{i=1}^{c} \mu_{ik} \phi_i(x_k)}{\sum_{i=1}^{c} \mu_{ik}} \tag{10}$$

To simplify Eq. (10), a partition of unity is considered where the functions $\beta_i(x_k)$, defined by

$$\beta_i(x_k) = \frac{\mu_{ik}}{\sum_{i=1}^{c} \mu_{ik}}, \; i = 1, \ldots, c \tag{11}$$

give information about the fulfilment of the respective fuzzy rule in the normalized form. It is obvious that $\sum_{i=1}^{c} \beta_i(x_k) = 1$ irrespective of x_k as long as the denominator of $\beta_i(x_k)$ is not equal to zero (this can be easily prevented by stretching the membership functions over the whole potential area of x_k). Combining Eqs. (10) and (11) we arrive at the following equation:

$$\hat{y}_k = \sum_{i=1}^{c} \beta_i(x_k) \phi_i(x_k), \; k = 1, \ldots, n \tag{12}$$

Very often, the output value is defined as a linear combination of consequence states

$$\phi_i(x_k) = x_k \theta_i, \quad i = 1, \ldots, c,$$

$$\theta_i^T = \left[\theta_{i1}, \ldots, \theta_{i(p+q)}\right] \tag{13}$$

The vector of fuzzified input variables at time instant k is written as

$$\psi_k = \left[\beta_1(x_k)x_k, \ldots, \beta_c(x_k)x_k\right], \; k = 1, \ldots, n$$

and then the fuzzified data matrix follows as:

$$\Psi^T = \left[\psi_1^T, \psi_2^T, \ldots, \psi_n^T\right] \tag{14}$$

If the matrix of the coefficients for the whole set of rules is written as $\Theta^T = \left[\theta_1^T, ..., \theta_c^T\right]$, then Eq. (12) can be rewritten in the matrix form

$$\hat{y}_k = \psi_k \Theta \tag{15}$$

and the compact form which describes the relation from the whole set of data becomes

$$\hat{Y} = \Psi \Theta \tag{16}$$

where $\hat{Y}$ stands for the vector of model outputs $\hat{y}_k$ where $k = 1, \ldots, n$ $(\hat{Y} = [\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n]^T)$.

The fuzzy model in the form given in Eq. (15) is referred to as the affine Takagi-Sugeno model and can be used to approximate any arbitrary function with any desired degree of accuracy [4, 5, 6]. The generality can be proven with the Stone-Weierstrass theorem [7] which suggest that any continuous function can be approximated by a fuzzy basis function expansion [8].

1.1.4 Estimation of fuzzy model parameters

The estimation of the fuzzy model parameters will be done using the least square error approach. The measurements satisfy the nonlinear equation of the system

$$y_i = g(x_i), \quad i = 1, \ldots, n \tag{17}$$

According to the Stone-Weierstrass theorem , for any given real continuous function g on a compact set $U^c \subset \mathbb{R}^p$ and arbitrary $\delta > 0$, there exist a fuzzy system f such that

$$\max_{x_i \in X} |f(x_i) - g(x_i)| < \delta, \quad \forall i \tag{18}$$

This implies the approximation of any given real continuous function with a fuzzy function from class $\mathcal{F}^p$ defined in Eq. (15). However, it has to be pointed out that lower values of δ imply higher values of c that satisfy Eq. (18). In the case of the

approximation, the error between the measured values and the fuzzy function outputs can be defined as

$$e_i = y_i - f(x_i) = y_i - \hat{y}_i, \quad i = 1, ..., n \qquad (19)$$

where y_i stands for the measured output and $\hat{y}_k$ for the model output at time instant k. To estimate the optimal parameters of the proposed fuzzy function (Θ) the minimization of the sum of square errors over the whole input set of data is performed as

$$E = \sum_{i=1}^{n} e_i^2 = (Y - \hat{Y})^T (Y - \hat{Y}) = \\ = (Y - \Psi\Theta)^T (Y - \Psi\Theta) \qquad (20)$$

.

The parameter Θ is obtained as $\frac{\partial E}{\partial \Theta} = 0$ and becomes

$$\Theta = \left(\Psi^T \Psi\right)^{-1} \Psi^T Y$$

.

The idea of an approximation can be interpreted as the most representative fuzzy function to describe the domain of outputs Y as a function of inputs X. This problem can also be viewed as a problem of data reduction, which often appears in identification problems with large data sets.

2 Biological waste-water treatment process

Waste-water treatment plants are large nonlinear systems subject to large perturbations in flow and load, together with uncertainties concerning the composition of the incoming wastewater. The simulation benchmark has been developed to provide an unbiased system for comparing various strategies without reference to a particular facility. It consists of five sequentially connected reactors along with a 10-layer secondary settling tank. The plant layout, model equations and control strategy are described in detail on the www page (http://www.ensic.u-nancy.fr/costwwtp). In our approach the layout was formed where the wastewater is purified in the mechanical phase and after this phase the moving bed bio-film reactor is used. Schematic representation of simulation benchmark is shown in Fig. 1. The detection of sensor faults was applied to the simulation model where the following measurements were used to calculate the fuzzy clusters and fuzzy model: influent ammonia concentration in the inflow Q_{in}

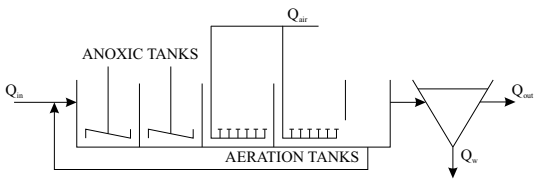

Fig. 1: Schematic representation of simulation benchmark

defined as $C_{NH4N_{in}}$, dissolved oxygen concentration in the first aerobic reactor tank $C^1_{O_2}$, dissolved oxygen concentration in the second aerobic reactor tank $C^2_{O_2}$ and the ammonia concentration in the second aerobic reactor tank $C_{NH4N_{out}}$. The fuzzy model was build to model the relation between the ammonia concentration in the second aerobic reactor tank and the other measured variables:

$$C_{NH4N_{out}}(k) = \mathcal{G}\left(C_{NH4N_{in}}(k), C^1_{O_2}(k), C^2_{O_2}(k)\right) \qquad (21)$$

where $\mathcal{G}$ stands for nonlinear relation between measured variables. First 15000 measurements (sampling time $T_s = 120s$) were used to find the fuzzy clusters and to estimate the fuzzy model parameters. At the measurement 17000 the slowly increasing sensor fault occur, which is than at time 18000 eliminated. This means that sensor to measure the ammonia concentration in the second aerobic reactor tank $C_{NH4N_{out}}$ is faulty. The signal with exponentially increasing value was added to the nominal signal. The whole set of measurements is shown in Fig. 2. The fuzzy model is obtained on the set

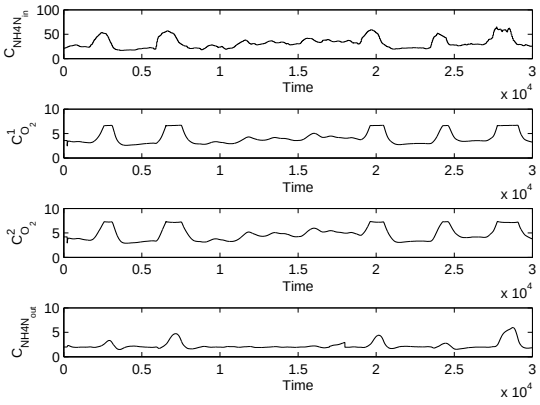

Fig. 2: The whole set of measurements

of first 15000 samples. The fuzzy model output $\hat{C}_{NH4N_{out}}$ and the process output $C_{NH4N_{out}}$ are shown in Fig. 3. The fault detection index is de-

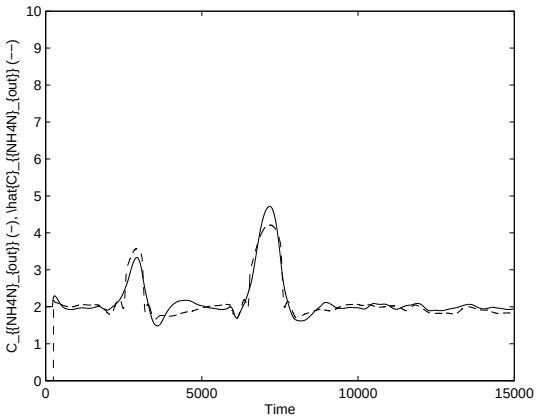

Fig. 3: The verification of fuzzy model

fined as:

$$f = \left(\frac{C_{NH4N_{out}} - \hat{C}_{NH4N_{out}}}{\hat{C}_{NH4N_{out}}} \right)^2 \qquad (22)$$

The fault tolerance index is defined as relative degree of maximal value of fault detection index in the identification or learning phase $f_{tol} = \gamma \max f$ where in our case $\gamma = 1.5$. This means that the fault detection index becomes ($f_{tol} = 0.15$). The fault which occur at the sample 17000 is detected at the sample 17556. The detection is delayed, but this is usual when the faults are slowly increasing.

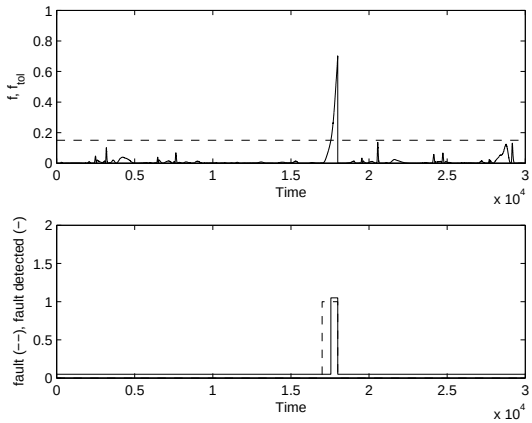

Fig. 4: The fault detection index, the fault tolerance index and the actual and detected fault

3 Conclusion

In this paper detection of sensor faults in wastewater treatment plant is discussed. It is realized by the use of fuzzy model which is obtained by Gustafson-Kessel fuzzy clustering algorithm. The detection of sensor fault was applied to the simulation model where the following measurements were used to calculate the fuzzy clusters: influent ammonia concentration, dissolved oxygen concentration in the first aerobic reactor tank, temperature, dissolved oxygen concentration and the ammonia concentration in the second aerobic reactor. The sensor fault on the sensor of the ammonia concentration in the second aerobic reactor has been detected without false alarms and with small time-delay because of the fault nature.

References

[1] Chen, J., Liao, C. M., "Dynamic process fault monitoring based on neural network and PCA". *Jour. of Process Control*, vol. 12, pp. 277-289, 2002.

[2] Johnson, R. A., Wichern, D. W., *Applied Multivariate Statistical Analysis*, Prentice-Hall, New Jersey, 1992.

[3] Daszykowski, M., Walczak, B., Massart, D. L., "Projection methods in chemistry". *Chemometrics and Intelligent Laboratory Systems*, vol. 65, pp. 97-112, 2003.

[4] Kosko, B., "Fuzzy Systems as Universal Approximators", *IEEE Transactions on Computers*, vol. 43, no. 11, pp. 1329-1333, 1994.

[5] Ying, H. GC., "Necessary conditions for some typical fuzzy systems as universal approximators", *Automatica*, vol. 33, pp. 1333-1338, 1997.

[6] Wang, L.-X., Mendel, J. M., "Fuzzy basis functions, universal approximation, and orthogonal least-squares learning", *IEEE Trans. Neural Networks*, vol. 3, no. 5, pp. 807-814, 1992.

[7] Goldberg, R. R., *Methods of Real Analysis*, John Wiley and Sons, 1976.

[8] Lin, C-H., "Siso nonlinear system identification using a fuzzy-neural hybrid system", *Int. Jour. of Neural Systems*, vol. 8, no. 3, pp. 325-337, 1997.