G Model
JJPC-2152; No. of Pages 13
ARTICLE IN PRESS
Journal of Process Control xxx (2017) xxx–xxx
Contents lists available at ScienceDirect
Journal of Process Control
journal homepage: www.elsevier.com/locate/jprocont

Hybrid online sensor error detection and functional redundancy for systems with time-varying parameters

Jianyuan Feng[a], Kamuran Turksoy[b], Sediqeh Samadi[a], Iman Hajizadeh[a], Elizabeth Littlejohn[c], Ali Cinar[a,b,*]

[a] Department of Chemical and Biological Engineering, Illinois Institute of Technology, Chicago, IL, United States
[b] Department of Biomedical Engineering, Illinois Institute of Technology, Chicago, IL, United States
[c] Biological Sciences Division, University of Chicago, Chicago, IL 60637, United States

ARTICLE INFO

Article history:
Received 7 October 2016
Received in revised form 9 April 2017
Accepted 30 April 2017
Available online xxx

Keywords:
Functional sensor redundancy
Sensor error detection
Kalman filter
Locally weighted partial least squares regression
Glucose concentrations
CGM

ABSTRACT

Supervision and control systems rely on signals from sensors to receive information to monitor the operation of a system and adjust manipulated variables to achieve the control objective. However, sensor performance is often limited by their working conditions and sensors may also be subjected to interference by other devices. Many different types of sensor errors such as outliers, missing values, drifts and corruption with noise may occur during process operation. A hybrid online sensor error detection and functional redundancy system is developed to detect errors in online signals, and replace erroneous or missing values detected with model-based estimates. The proposed hybrid system relies on two techniques, an outlier-robust Kalman filter (ORKF) and a locally-weighted partial least squares (LW-PLS) regression model, which leverage the advantages of automatic measurement error elimination with ORKF and data-driven prediction with LW-PLS. The system includes a nominal angle analysis (NAA) method to distinguish between signal faults and large changes in sensor values caused by real dynamic changes in process operation. The performance of the system is illustrated with clinical data continuous glucose monitoring (CGM) sensors from people with type 1 diabetes. More than 50,000 CGM sensor errors were added to original CGM signals from 25 clinical experiments, then the performance of error detection and functional redundancy algorithms were analyzed. The results indicate that the proposed system can successfully detect most of the erroneous signals and substitute them with reasonable estimated values computed by functional redundancy system.

1. Introduction

Sensor failure is a frequent problem in monitoring and control systems due to the characteristics and work conditions of sensors. Erroneous sensor signals can cause an online control system to suggest wrong manipulated variable values that may deteriorate system performance and cause dangerous situations.

Sensor errors can be divided into two main categories: hard failure (complete failure) and soft failure such as bias, drift and outliers [21]. It is easier to detect hard failures while it is difficult to detect and reconcile some soft failures, since it may be difficult to distinguish between sensor values reporting changes in process operation under normal situations and soft faults that appear intermittently or have small deviations. Several approaches have been proposed for sensor fault detection and diagnosis (FDD) and data reconciliation. Technologies for FDD are grouped into three categories: quantitative model based, qualitative model based, and process history based (data-driven) techniques [28]. For model-based techniques, prior knowledge (either quantitative or qualitative) about the process is needed to develop the model for computing residuals between measured and predicted values and compare the residual with a threshold to determine the presence of an error. The models can be first principles, empirical or expe-

Abbreviations: ANN, artificial neural networks; BGC, blood glucose concentration; CGM, continuous glucose monitoring; DIP, decrease followed by increase period; DP, decrease period; EAP, error appearance percentage; EDRF, error detected but reconciliation failed; EDRS, error detected and reconciled successfully; FDD, fault detection and diagnosis; FDR, false detection ratio; GC, glucose concentration; IDP, increase followed by decrease period; IP, increase period; LW-PLS, locally-weighted partial least squares; LWR, locally weighted regression; MPR, model prediction residual; NAA, nominal angle analysis; ORKF, outlier-robust Kalman filter; PCA, principal component analysis; PLS, partial least square; PISA, pressure-induced sensor attenuations; S, sensitivity; SED&FR, sensor error detection and functional redundancy; SGF, Savitzky-Golay filter; SJCC, signal jump caused by calibration; SRR, successful reconciliation rate; SP, steady period.

* Corresponding author at: Department of Chemical and Biological Engineering, Illinois Institute of Technology, Chicago, IL, United States.
E-mail address: cinar@iit.edu (A. Cinar).

http://dx.doi.org/10.1016/j.jprocont.2017.04.004
0959-1524/

riential. For process history based techniques, a large amount of historical process data is needed to create a database of fault patterns, then compute statistical limits that indicate the significance of deviations in sensor readings [22]. Qualitative model-based techniques have also been integrated with data-driven techniques to leverage the power of multivariate statistical approaches and knowledge-based systems [27]. A different paradigm is to develop robust control systems that can tolerate sensor errors [15,20]. A limitation of a robust control approach is the need for prior knowledge of sensor error to build the controller. Error detection methods based on Kalman filter [19,30], artificial neural networks [23,25] (ANN) and principal component analysis [9,29] (PCA) are some of the model-based and data-driven techniques that have been used in many applications.

For online control systems, sensor error detection is only part of the work, the erroneous sensor reading must be modified before it is used in control decisions. Data reconciliation [2] techniques have been proposed to build models for providing estimates of the erroneous or missing sensor signal. Complex systems with time-varying parameters and/or nonlinearities necessitate either fixed reliable models that can represent the process over wide ranges of operating conditions or adaptive models. The development of detailed models can be challenging and adaptive simple models provide an attractive and feasible option. However, the use of erroneous data in updating model parameters would yield a model that includes the effects of sensor errors, and affect error detection and data reconciliation. In order to address this problem, we propose a hybrid online sensor error detection and functional redundancy (SED&FR) system based on two technologies, the outlier-robust Kalman filter [12,26] (ORKF) and the locally-weighted partial least squares [12,18] (LW-PLS).

The two technologies use different approaches for fault detection and analysis. ORKF builds a Kalman model by using signals collected in recent past sampling times. LW-PLS builds a partial least squares (PLS) model by using the signal samples in the database of historical data that are most similar to current signal samples. The fault reported by ORKF indicates that the current signal is different from the signal trace indicated by recent past values of signals, but it may not be a fault if compared in LW-PLS because the signal samples in the database have similar behavior. Conversely, if a new signal pattern does not match any type of true signal trace in the database, LW-PLS may not able to provide the correct detection. Hence, the use of both algorithms simultaneously provides a more robust fault detection and diagnosis effort.

The final estimate based on functional redundancy relies on the same two different methods to build the models and provide estimates for erroneous or missing sensor measurements. The hybrid system increases the probability of rapid detection of sensor faults and improves the accuracy of the estimated values. The results from each algorithm can be compared or combined to find the optimal reconciled estimated value. When a continuous sensor error is detected that lasts for a long period, LW-PLS would be more advantageous in making a prediction.

The remainder of the paper is structured as follows. The two technologies (ORKF and LW-PLS) and the nominal angle analysis algorithm for discrimination between sensor readings affected by large changes in process operation and sensor faults are be described in Section 2. The results of the system performance with data from clinical experiments are given in Section 3. The discussion of results and conclusions are provided in Section 4 and Section 5, respectively.

2. Methods

For an online control system, the incoming sensor data must be reconciled before being used to update the model equations or calculate a value for manipulate variables. Hence, data reconciliation techniques that require access to future data samples, such as the Kalman smoother [3,17] or Savitzky-Golay filter (SGF) [24] are not appropriate. The techniques that will be used should be capable of both error detection and data reconciliation in order to maintain the function of the control system. The erroneous or missing measurements should be also replaced by appropriate model-based estimates. The ORKF [26] and the LW-PLS [18] provide two complementary technologies. ORKF provides predictions of sensor readings based on the recent behavior of the sensor. LW-PLS uses a model developed by historical data and updates its parameters with current sensor information. Consequently, its predictions take into account the historical information about the expected system performance.

2.1. Outlier-robust Kalman filter

The ORKF method has been used to solve many sensor outlier-related problems in various fields such as global positioning system (GPS) data analysis [1] and robotic systems [26]. Sensor data $\boldsymbol{y}_k \in \Re^{d_1}$ can be described by Kalman filter system equations with hidden states $\boldsymbol{x}_k \in \Re^{d_2}$ where d_1 and d_2 are the dimensions of outputs (number of sensors) and state variables:

$$\boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{x}_k + \boldsymbol{v}_k \tag{1}$$

$$\boldsymbol{x}_k = \boldsymbol{A}\boldsymbol{x}_{k-1} + \boldsymbol{s}_k \tag{2}$$

where $\boldsymbol{C} \in \Re^{d_1 \times d_2}$ is the observation matrix, $\boldsymbol{A} \in \Re^{d_2 \times d_2}$ is the state transition matrix, $\boldsymbol{v}_k \in \Re^{d_1 \times 1}$ is the observation noise at time step k, and $\boldsymbol{s}_k \in \Re^{d_2 \times 1}$ is the state noise at time step k. We assume that v_k and $\boldsymbol{s}_k$ are uncorrelated mean-zero Gaussian noise: $\boldsymbol{v}_k \sim Normal(0, \boldsymbol{R})$, $\boldsymbol{s}_k \sim Normal(0, \boldsymbol{Q})$. Covariance matrices $\boldsymbol{R} \in \Re^{d_1 \times d_1}$ and $\boldsymbol{Q} \in \Re^{d_2 \times d_2}$ are diagonal matrices, with $r \in \Re^{d_1 \times 1}$ and $q \in \Re^{d_2 \times 1}$ on their diagonal, respectively. For the standard Kalman filter, for a data set of length N and $k = 1, \ldots, N$, the filter propagation and update equations are:

$$\boldsymbol{x}'_k = \boldsymbol{A}\langle \boldsymbol{x}_{\mathbf{k-1}} \rangle \tag{3}$$

$$\boldsymbol{\Sigma}'_k = \boldsymbol{A}\boldsymbol{\Sigma}_{k-1}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{Q} \tag{4}$$

$$\boldsymbol{S}'_k = \left(\boldsymbol{C}\boldsymbol{\Sigma}'_k\boldsymbol{C}^{\mathrm{T}} + \boldsymbol{R}\right)^{-1} \tag{5}$$

$$\boldsymbol{K}'_k = \boldsymbol{\Sigma}'_k\boldsymbol{C}^{\mathrm{T}}\boldsymbol{S}'_k \tag{6}$$

$$\langle \boldsymbol{x}_k \rangle = \boldsymbol{x}'_k + \boldsymbol{K}'_k\left(\boldsymbol{y}_k - \boldsymbol{C}\boldsymbol{x}'_k\right) \tag{7}$$

$$\boldsymbol{\Sigma}_k = \left(\boldsymbol{I} - \boldsymbol{K}'_k\boldsymbol{C}\right)\boldsymbol{\Sigma}'_k \tag{8}$$

where $\langle \boldsymbol{x}_k \rangle$ is the posterior mean vector of the state $\boldsymbol{x}_k$, $\boldsymbol{\Sigma}_k$ is the covariance matrix of $\boldsymbol{x}_k$, and $\boldsymbol{S}'_k$ is the covariance matrix of the residual prediction error at step k. As a model used to predict the measurements, the data used to train the model has to be robust to outliers. Hence, the coefficient matrices representing the system dynamics ($\boldsymbol{C}$, $\boldsymbol{A}$, $\boldsymbol{R}$ and $\boldsymbol{Q}$) are unknown and need to be adaptive to fit dynamic changes in a system with time-varying coefficients. Consequently, the standard Kalman filter with constant coefficients that considers all data samples to be part of the data cloud would not be appropriate and could generate erroneous estimates.

To overcome these limitations, a Bayesian algorithm that treats the weights associated with each data sample probabilistically is introduced. A scalar gamma-distributed weight w_k is assigned for each observed data sample, $\boldsymbol{y}_k$. Gamma distribution is chosen for the weights to ensure that they remain positive. The resulting prior distributions are Normal, with means and variances as:

$$\boldsymbol{y}_k | \boldsymbol{x}_k, w_k \sim Normal\left(\boldsymbol{C}\boldsymbol{x}_k, \boldsymbol{R}/w_k\right)$$