Neurocomputing 175 (2016) 808–815

Contents lists available at ScienceDirect

Neurocomputing

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

Research on data reconciliation based on generalized T distribution with historical data

Shengxi Wu, Qiang Ye, Cheng Chen, Xingsheng Gu *

Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, No.130 of Meilong Road, Shanghai 200237, China

ARTICLE INFO

Article history:
Received 10 November 2014
Received in revised form
18 August 2015
Accepted 9 October 2015
Communicated by Xingsheng Gu
Available online 6 November 2015

Keywords:
Data reconciliation
Maximum likelihood estimation
Generalized T distribution
Historical data
PSO algorithm

ABSTRACT

In the most of previous data reconciliation(DR) studies, process data were conventionally characterized by normal Gaussian distribution, so the optimality/validity of DR estimator is implicitly based on a main assumption that errors follow normal Gaussian distribution. When this assumption is not satisfied, conventional data reconciliation approaches will become unavailable. However, normal distribution usually does not exist in real chemical engineering practice, as it is hard to ensure the normality even for high-quality measurements. So it is necessary to propose a new DR method which can accommodate more variety of measurement error distribution. In this paper, generalized T distribution is applied to accommodate measurement error distribution, meanwhile, historical data is introduced to estimate the objective function parameters by using Particle Swarm Optimization (PSO) algorithm. A novel robust data reconciliation method is proposed based on GT distribution and historical data, at the same time, its robustness characteristics are investigated. The new method is demonstrated on a steam-metering system for a methanol synthesis unit. Based on the comparison with other DR methods, the novel robust DR method can effectively improve the reliability of reconciled data even when errors do not follow normal distribution.

1. Introduction

With the distributed control system (DCS) and programming logic controller (PLC) widely used in chemical industry, the quantity of chemical process data is greatly increased. So reliable and accurate estimations from those process data are crucial for process modeling, advanced automatic control and optimization purposes. However, the process data measured by instruments have various errors (measurement error and/or gross errors) [1]. Due to these errors we cannot expect that any set of process data will obey the laws of conservation such as mass conservation and energy balance. Therefore, data reconciliation is needed to optimally adjust measured data so that the reconciled values subject to the constraints of process model and other conservation laws [2]. In chemical industry, data reconciliation techniques are widely used, such as data reconciliation techniques were applied in a paraffin instrument [3], the catalytic cracking of heavy oil, the natural gas pipeline systems [4] and the steam turbines of boiling water reactors [5].

At present, these available data reconciliation software including Datacon (based on the process simulation model) are developed by the United States Simulation Science Co. Ltd., Advisor (based on expert system) is developed by AspenTech Ltd., Sigmafine (based on statistical data) is produced by KBC Process Technology Ltd., Datsee which is developed by Technip in France, and APC-Data pro which is developed by SUPCON Ltd. These popular software are mainly used in heavy large petrochemical enterprises with complex large-scale processes including large numbers of nodes and streams(particularly in petroleum and ethylene refining). Presently most software can only be partly used, because it is difficult to rectify process data in real time in the manufacturing execution system.

The most common method for data reconciliation is the weighted least-squares (LS), but its optimality and validity is implicitly based on an assumption that errors follow normal distribution. When this assumption is not satisfied, the conventional LS method performs poorly. Furthermore, LS is sensitive to gross errors, if gross errors exist in the measurement data, LS will lead to incorrect estimations and then severely deflect reconciliation of other measurements [6,7]. To eliminate the influence of gross errors, many gross error detection methods are proposed, such as global test, nodal test and measurement test [8]. Although these

* Corresponding author.
E-mail address: xsgu@ecust.edu.cn (X. Gu).

http://dx.doi.org/10.1016/j.neucom.2015.10.093
0925-2312/

methods can effectively eliminate the gross errors, they are still based on the normality assumption, which is not necessarily true.

As the analogy between data reconciliation and parameter regression [9], some robust identification methods can be applied to data reconciliation, such as contaminated normal estimation [10] and Huber estimation [11]. These robust DR methods greatly simplify data reconciliation procedures by combining gross error detection and conventional data reconciliation into a single optimization problem of the objective function. Even though it is insensitive to gross errors, the efficiency and optimality of the estimation is still dependent on the predefined measurement error distribution. Recently, a lot of articles have appeared to develop many other robust DR methods, Zhengjiang Zhang proposed a new quasi-weighted least squares robust estimator [12], Bethany Nicholson proposed a new Huber's fair function and Hampel's redescending estimator which can be used to get fast and accurate estimates in the presence of many gross measurement errors [13], two methods of correntropy based nonlinear dynamic data reconciliation (NDDR) as well as gross error detection and identification (GEDI) are addressed by Zhengjiang Zhang [14], Claudia E. Llanos made a comparative performance analysis of five different robust data reconciliation strategies [15].

D. Wang and J.A. Romagnoli designed a robust adaptive data reconciliation estimator based on the Generalized T distribution (GT) [16], GT distribution is applied to describe the measurement error distribution, which can accommodate different types of error distribution by changing the distributional parameters. Zhengjiang Zhang proposed a new just-in-time learning-based data reconciliation and parameter estimation (DPRE) method [17], where the historical data were used by just-in-time learning for solving large-scale DRPE problem. In this paper, a novel robust GT data reconciliation approach is proposed with historical data, meanwhile, the distributional parameters which represent the characteristics of errors are estimated by using Particle Swarm Optimization (PSO) algorithm. This work is organized as follows: conventional data reconciliation and the data reconciliation process based on historical data are introduced in Section 2. Maximum likelihood estimation is introduced in Section 3, GT data reconciliation and its robustness are introduced in Section 4, the process to estimate distributional parameters by using PSO algorithm is addressed in Section 5. Finally, the comparative advantages of the new method are illustrated in a steam-metering system for a methanol synthesis unit when compared with other DR methods.

2. Data reconciliation

2.1. Principle of data reconciliation

In this paper, we assumed that all the measurement data can be rectified and other un-measurement data can be estimated. Meanwhile, the final reconciled data should be able to obey the conservation laws, such as mass balance, energy balance and chemical elements balance. These conservation laws can be represented as a constraint equation defined as

$$F(\hat{X}, U) = 0 \tag{1}$$

where $\hat{X}$ is the vector of rectified measurement data, U is the vector of un-measurement data which should be estimated, and F represents the constraints which are usually defined by some physical and energy conserved laws. The objective function of conventional DR problem can be described as the least squares form between the measurement data and reconciled data, that is:

$$\min \sum_{i=1}^{n} (\hat{x}_i - y_i)^2 / \sigma_i^2 \tag{2}$$

where y_i is ith measured value, $\hat{x}_i$ is the ith reconciled value and σ_i is the corresponding standard deviation of the measurement error. Eq. (2) can be described as a matrix form, as shown in Eq. (3), where Q is the matrix of variance/covariance, and σ_i^2 is the diagonal element of the matrix.

$$\min\left[\left(\hat{X}-Y\right)^T Q^{-1}\left(\hat{X}-Y\right)\right] \tag{3}$$

If the conservation laws defined in Eq. (1) are linear, then the objective function of data reconciliation problem can be represented as:

$$\begin{cases} \min\left[\left(\hat{X}-Y\right)^T Q^{-1}\left(\hat{X}-Y\right)\right] \\ A\hat{X}+BU+C=0 \end{cases} \tag{4}$$

where $A\hat{X}+BU+C=0$ is the linear constraint equation, commonly, Lagrange multiplier method is applied to solve the optimization problem Eq. (4), the final results of the rectified data $\hat{X}$ and estimated data U can be obtained as the form of

$$\hat{X} = \left\{I - QA^T\left(AQA^T\right)^{-1}A\right\}Y - QA^T\left(AQA^T\right)^{-1}(BU+C)$$

$$U = \left\{B^T\left(AQA^T\right)^{-1}B\right\}^{-1}B^T\left(AQA^T\right)^{-1}(-AY-C) \tag{5}$$

where I is the unit matrix. It should be noted that we can get the solutions of Eq. (5) only when A is full row rank(the measurement data can be rectified) and B is full column rank (the un-measurement data can be estimated), otherwise, $\left(AQA^T\right)^{-1}$ and $\left[B^T\left(AQA^T\right)^{-1}B\right]^{-1}$ would be unsolvable.

2.2. Robust data reconciliation based on historical data

This work aims to improve the efficiency of data reconciliation. Fig. 1 depicts the main steps of conventional DR approach, including modeling, data classification, variance/covariance estimation, gross error detection and data reconciliation. In the conventional DR process, historical data are ignored. However, historical data always can provide more accurate error characteristics than statistical method. The framework of robust data reconciliation based on historical data is shown in Fig. 2. The greatest difference between the novel strategy and conventional DR approach is that the distributional parameters in the objective function can be adjusted with historical data. In this way, the objective function can reflect the characteristics of errors more accurately.

Another main difference between the novel strategy and conventional DR approach is that the former needs iterative calculation in the reconciliation process. The iteration will not stop until the residuals convergence reaches a certain tolerance criterion, which is equivalent for adding a self-validate link. Meanwhile, the novel strategy combines conventional data reconciliation and gross error detection into a single optimization problem of objective function. Based on these advantages, the novel strategy can significantly improve the accuracy and effectiveness of the reconciled data when compared with conventional DR approach.