Computers and Chemical Engineering 33 (2009) 795–814

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Review

Data-driven Soft Sensors in the process industry

Petr Kadlec[a], Bogdan Gabrys[a,*], Sibylle Strandt[b]

[a] Smart Technology Research Centre, Computational Intelligence Research Group, Bournemouth University, Poole BH12 5BB, United Kingdom
[b] Evonik Degussa AG, 45128 Essen, Germany

ARTICLE INFO

Article history:
Received 17 March 2008
Received in revised form
27 November 2008
Accepted 30 December 2008
Available online 20 January 2009

Keywords:
Soft Sensors
Process industry
Data-driven models
PCA
ANN

ABSTRACT

In the last two decades Soft Sensors established themselves as a valuable alternative to the traditional means for the acquisition of critical process variables, process monitoring and other tasks which are related to process control. This paper discusses characteristics of the process industry data which are critical for the development of data-driven Soft Sensors. These characteristics are common to a large number of process industry fields, like the chemical industry, bioprocess industry, steel industry, etc. The focus of this work is put on the data-driven Soft Sensors because of their growing popularity, already demonstrated usefulness and huge, though yet not completely realised, potential. A comprehensive selection of case studies covering the three most important Soft Sensor application fields, a general introduction to the most popular Soft Sensor modelling techniques as well as a discussion of some open issues in the Soft Sensor development and maintenance and their possible solutions are the main contributions of this work.

Contents

* Correspondingauthor at: Bournemouth University, Computational Intelligence Research Group, Smart Technology Research Centre, Poole House, Talbot Campus, Fern Barrow, Poole BH12 5BB, United Kingdom. Tel.: +44 1202965298; fax: +44 1202965314.
E-mail address: bgabrys@bournemouth.ac.uk (B. Gabrys).

0098-1354/$ – see front matter
doi:10.1016/j.compchemeng.2008.12.012

1. Introduction

Industrial processing plants are usually heavily instrumented with a large number of sensors. The primary purpose of the sensors is to deliver data for process monitoring and control. But approximately two decades ago researchers started to make use of the large amounts of data being measured and stored in the process industry by building predictive models based on this data. In the context of process industry, these predictive models are called *Soft Sensors*. This term is a combination of the words "software", because the models are usually computer programs, and "sensors", because the models are delivering similar information as their hardware counterparts. Other common terms for predictive sensors in the process industry are *inferential sensors* (see e.g. Jordaan, Kordon, Chiang, & Smits, 2004; Qin, Yue, & Dunia, 1997), *virtual on-line analyser* as they are called in the Six-Sigma context (Han & Lee, 2002) and *observer-based* sensors (Goodwin, 2000).

At a very general level one can distinguish two different classes of Soft Sensors, namely model-driven and data-driven. The model-driven family of Soft Sensors is most commonly based on First Principle Models (FPM) but model-driven Soft Sensors based on extended Kalman filter (Welch & Bishop, 2001) or adaptive observer (Bastin & Dochain, 1990) have also been published (e.g. Chruy, 1997; Jos de Assis & Maciel Filho, 2000). First Principle Models describe the physical and chemical background of the process. These models are developed primarily for the planning and design of the processing plants, and therefore usually focus on the description of the ideal steady-states of the processes which is only one of their drawbacks which makes it difficult to base Soft Sensors on them. As a solution the data-driven Soft Sensors gained increasing popularity in the process industry. Because data-driven models are based on the data measured within the processing plants, and thus describe the real process conditions, they are, compared to the model-driven Soft Sensors, more reality related and describe the true conditions of the process in a better way. Nevertheless there is a lot of different issues which have to be dealt with while developing data-driven Soft Sensors. These issues will be discussed later on in this paper. The most popular modelling techniques applied to data-driven Soft Sensors are the Principle Component Analysis (PCA) (Jolliffe, 2002) in a combination with a regression model, Partial Least Squares (Wold, Sjstrm, & Eriksson, 2001), Artificial Neural Networks (Bishop, 1995; Principe, Euliano, & Lefebvre, 2000; Hastie, Tibshirani, & Friedman, 2001), Neuro-Fuzzy Systems (Jang, Sun, & Mizutani, 1997; Lin & Lee, 1996) and Support Vector Machines (SVMs) (Vapnik, 1998).

The range of tasks fulfilled by Soft Sensors is broad. The original and still most dominant application area of Soft Sensors is the prediction of process variables which can be determined either at low sampling rates or through off-line analysis only. Because these variables are often related to the process output quality, they are very important for the process control and management. For these reasons it is of great interest to deliver additional information about these variables at higher sampling rate and/or at lower financial burden, which is exactly the role of the Soft Sensors. The modelling methods applied to this kind of applications are either statistical or soft computing supervised learning approaches. This Soft Sensor application field is further on referred to as *on-line prediction*. Other important application fields of Soft Sensors are those of *process monitoring and process fault detection*. These tasks refer to detection of the state of the process and in the case of a deviation from the normal conditions to identification of the deviation source. Traditionally, the process state is monitored by process operators in the control rooms of the processing plants. The observation and interpretation of the process state is often based on univariate statistics and it is up to the experience of the process operator to put the particular variables into relations and to make decisions about the process state. The role of process monitoring Soft Sensors is, based on the historical data, to build multivariate features which are relevant for the description of the process state. By presenting the predicted process state or the multivariate features the Soft Sensor can support the process operators and allow them to make faster, better and more objective decisions. Process monitoring Soft Sensors are usually based on the Principle Component Analysis and Self Organizing Maps (Kohonen, 1997). It was already mentioned that processing plants embody large number of various sensors, therefore there is a certain probability that a sensor can occasionally fail. Detection of this failure is the next application area of Soft Sensors. In more general terms this application field can be described as *sensor fault detection and reconstruction*. Once a faulty sensor is detected and identified, it can be either reconstructed or the hardware sensor can be replaced by another Soft Sensor, which is trained to act as a *back-up Soft Sensor* of the hardware measuring device. If the back-up sensor proves to be an adequate replacement of the physical sensor, this idea can be driven even further and the Soft Sensor can replace the measuring device also in normal working conditions. The software tool can be easier maintained and is not subject to mechanical failures and therefore such a substitution can provide a financial advantage for the process owner.

Despite all the previously listed Soft Sensor application fields and the high number of publications dealing with Soft Sensor applications, there are still some unaddressed issues of the Soft Sensor development and maintenance. A lot of the origins of these issues are in the process data which is used for the Soft Sensor building. Common effects present in the data are measurement noise, missing values, data outliers, co-linear features and varying sampling rates. To solve these problems, there is typically a large amount of manual work needed. Another problem is that the processing plants are rather dynamic environments. Often they *develop* gradually during the operation time but there can be also sudden abrupt changes of the process, for example, if the quality of the process input changes. It is very difficult for the Soft Sensors to react to these changes which usually results in prediction accuracy deterioration. At present time, these issues are solved in a rather ad hoc manner, which leads to unnecessary high costs of the Soft Sensor development and maintenance. Further on in this work, all the aspects, which have been briefly outlined in this section, are going to be reviewed in a more comprehensive way. The rest of the paper