Computers and Chemical Engineering 96 (2017) 42–54

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Multiple adaptive mechanisms for data-driven soft sensors

Rashid Bakirov*, Bogdan Gabrys, Damien Fay

Data Science Institute, Bournemouth University, Fern Barrow, Poole House P256, Talbot House, BH12 5BB Poole, UK

ARTICLE INFO

Article history:
Received 29 March 2016
Received in revised form 12 July 2016
Accepted 31 August 2016
Available online 26 September 2016

Keywords:
Soft sensors
Adaptive mechanisms
Streaming data
Ensemble methods

ABSTRACT

Recent data-driven soft sensors often use multiple adaptive mechanisms to cope with non-stationary environments. These mechanisms are usually deployed in a prescribed order which does not change. In this work we use real world data from the process industry to compare deploying adaptive mechanisms in a fixed manner to deploying them in a flexible way, which results in varying adaptation sequences. We demonstrate that flexible deployment of available adaptive methods coupled with techniques such as cross-validatory selection and retrospective model correction can benefit the predictive accuracy over time. As a vehicle for this study, we use a soft-sensor for batch processes based on an adaptive ensemble method which employs several adaptive mechanisms to react to the changes in data.

1. Introduction

Modelling industrial processes typically involves estimating a finite set of physical quantities. Certain necessary measurements in the process industry are often excessively costly or time consuming. Soft sensors have proven to be useful tools in these situations, providing information about the quantity to be estimated without directly performing the measurements. There are two main families of soft sensors: physical model based and data-driven soft sensors (Kadlec et al., 2009). Physical model-driven soft sensors estimate the quantity using chemical and physical laws behind the process. For many complex processes this is impossible as accurate first principle models are not known or evolution of the process is not taken into account.

In this work we focus on data-driven soft sensors. In particular we scrutinise and explore the multiple adaptive mechanisms applied to soft sensors in a streaming data scenario. The streaming data scenario itself introduces some interesting questions. For the batch processes where data arrives in large segments called batches, which are common in the process industry, especially in the chemical, microelectronics and pharmaceutical areas (Cinar et al., 2003), the models are typically adapted when a new batch of data is observed. This can be done with or without historical data (which may have been jettisoned, is not readily available or would be computationally costly to include).

The underlying assumptions of a soft-sensor model may only hold for a certain period of time (Gallagher et al., 1997). It has been shown that many changes in the environment which are no longer being reflected in the model contribute to the deterioration of predictive model's accuracy over time (Schlimmer and Granger, 1986; Kadlec et al., 2011). Factors such as sensor/measurement deterioration, addition of new sensors, changes in the process flow or input materials, etc. can result in alternate models explaining the process better. This requires constant manual retraining and readjustment of the soft sensors which is often expensive, time consuming and in some cases impossible – for example when the historical data is not available any more.

To avoid outright retraining and development of soft-sensor models for an evolving process, many soft sensors with adaptive mechanisms (AMs) (Kadlec et al., 2011) have been proposed, starting with the recursive Principal Component Analysis (PCA) and Partial Least Squares (PLS) approaches presented in Dayal and MacGregor (1997), Li et al. (2000) and developing as discussed later (Section 2). Often adaptation operates by reducing the weight applied to irrelevant parts of the historical data, which may be implemented in a variety of ways. In addition, recent adaptive soft sensors use, not one, but multiple AMs. Employing multiple AMs is more versatile than a single approach and can lead to superior prediction performance (e.g. Kadlec and Gabrys, 2011; Jin et al., 2015a) because given an evolution in the process some AMs are more appropriate than others at different times. However, in practice, most research has concentrated on AM's deployed in a prescribed fixed order set at the model design time. The common choice is to deploy all of the AMs at the same time; however, this can lead to undesirable results with some AMs cancelling the effect of others or by overcompensating for change in the process.

* Corresponding author.
E-mail addresses: rbakirov@bournemouth.ac.uk (R. Bakirov), bgabrys@bournemouth.ac.uk (B. Gabrys), dfay@bournemouth.ac.uk (D. Fay).

http://dx.doi.org/10.1016/j.compchemeng.2016.08.017
0098-1354/

In this paper we are providing a deeper analysis of a soft sensor with multiple AMs, concentrating on the choice and order of AMs' deployment. For this purpose we use the Simple Adaptive Batch Learning Ensemble (SABLE) (Bakirov et al., 2016), on which the analysed soft sensor is based. SABLE is an ensemble method, meaning that the final prediction is calculated by combining the predictions of different models (experts). As such, it uses three different popular AMs to deal with changing data: (i) Recursive Partial Least Squares (RPLS) (Joe Qin, 1998) is used to discount older data, (ii) adapting the combination weights targets the ensemble mix, and (iii) addition/merge/removal of experts adapts the structure of the model ensemble. This allows us to relate to many other soft-sensors which use similar AMs for their adaptation, making our exploration relevant to a very broad class of similar approaches.

As a result of our analysis based on three real world process industry datasets we provide strong empirical evidence that deploying AM's in a flexible order without a predetermined sequence leads to better prediction accuracy. Two methods in particular were effective for the choice of the AM – cross-validatory selection and retrospective model correction (see Section 3.2). Cross-validatory AM selection involves selecting the AM to deploy based on the performance on the current data. Once the subsequent data has been fully observed, the AM which would indeed have been the best for previous batch becomes known. Retrospective model correction is reverting the model to the state, which the deployment of this best AM would have created.

The paper is structured as follows: Section 2 introduces the related work, concentrating on soft sensors with one or more adaptive mechanisms. Section 3 presents mathematical formulation of the framework of a system with multiple adaptive elements in batch streaming scenario. Section 4 introduces the algorithm for the soft-sensor, which was used for the experimentation, including its AMs and a description of RPLS. Experimental methodology, description of the datasets and results of our experiments are covered in Section 5. We conclude by giving our final remarks in Section 6.

2. Related work

Recently, many soft sensors and other regression methods for industrial processes, which explicitly consider the adaptation of the model, such as Kadlec and Gabrys (2011), Grbić et al. (2013), Kaneko and Funatsu (2014, 2015), Gomes Soares and Araújo (2015a,b), Souza and Araújo (2014), Shao et al. (2014), Ni et al. (2014), Jin et al. (2014, 2015a,b), Shao and Tian (2015) and Shao et al. (2015a,b), have been proposed. Many of the algorithms are examples of incremental learning, with the exception of Jin et al. (2015a,b) which are specifically targeted at batch processes. Adaptivity is usually achieved by building a predictive model using (a) the latest historical data; and/or (b) the historical data which is the most similar to the current data. Adaptive methods often use multiple models to make the final prediction, either by the weighted combination of their outputs (Kadlec and Gabrys, 2011; Grbić et al., 2013; Kaneko and Funatsu, 2014, 2015; Gomes Soares and Araújo, 2015a,b; Souza and Araújo, 2014; Jin et al., 2015a,b; Shao and Tian, 2015; Shao et al., 2015b) – these are known as ensemble models, or, more rarely, selecting one of them (Jin et al., 2014, 2015b; Shao et al., 2015a). Most of the models, or experts, are built on the subsets of historical data which represent different degrees of relation to the current data.

Ensemble methods date as far back as 1960s, when it was shown that combining multiple predictive models may give better results than using single models (Bates and Granger, 1969). One of the advantages of ensemble methods is the ability to model local dependencies in the data, a classical example being an adaptive mixture of local experts presented in Jacobs et al. (1991). This is achieved by weighting the models' predictions on a data instance by the location of this instance in the input space. Soft sensors using local ensembles are described in Kadlec and Gabrys (2011, 2009b, 2010), Shao and Tian (2015), Shao et al. (2015b) and Jin et al. (2015a). These methods first identify the disjoint segments of the historical input space where the process produced outputs described by a common model, sometimes also called *receptive fields*. Then they build a model for each receptive field using Partial Least Squares (PLS) (Wold, 1966) or Support Vector Regression (Drucker et al., 1996). The models therefore describe different regions of the process. The final prediction is a weighted average of all of the experts. Here, for each new data instance, the weights of experts depend on the location of the observed instance and in some cases the prediction. The AM used in Kadlec and Gabrys (2009b) is based on change of models' local weights depending on their error. This model was extended in Kadlec and Gabrys (2011) to include adaptation of the base models using the RPLS forgetting. Kadlec and Gabrys (2010) further extends the model to include creation of additional experts. Shao and Tian (2015) and Shao et al. (2015b) use adaptation of base models and adaptive weighting with Jin et al. (2015a) additionally introducing adaptive offset correction. Another soft sensor based on local ensemble with a moving window and weights change AMs is described in Grbić et al. (2013).

Also popular in the literature are global regression ensembles (Kaneko and Funatsu, 2014, 2015; Gomes Soares and Araújo, 2015a,b). These typically assign weights to experts based on their general performance, not considering the local aspects of data. Global ensemble methods use similar AMs. For instance, Kaneko and Funatsu (2014) adapt to changes by creating new experts and changing their weights. Gomes Soares and Araújo (2015b) includes AMs such as adaptation of base models via a moving window strategy, changing experts weights and adding new experts. Gomes Soares and Araújo (2015a) additionally employs a boosting like *instance weighting* mechanism resampling the training data. Both Gomes Soares and Araújo (2015a,b) may remove experts as well. A method which uses an ensemble of univariate regressors for multivariate regression is described in Souza and Araújo (2014). It includes weighting of models and forgetting factor AMs. Kaneko et al. (2014) uses *time difference* ensemble based on the distance between the current input and historical inputs. This method can use either moving window or just-in-time (creation of a model from most relevant instances) approaches for adaptation. Kaneko et al. (2014) also use just-in-time model creation with global performance based adaptive weighting.

From the analysis above we can see that there are a host of adaptation mechanisms which can be applied with ensemble methods. A review of these mechanisms for soft sensors is given in Kadlec et al. (2011). The mechanisms target different characteristics of the model: the error, the current location in the input space (or output space), and the temporal distance. The SABLE framework chosen here also includes such functionality. Most of the described work above have a common characteristic that whatever the AMs, they are applied at every time step in the same manner. In contrast the approaches proposed in Gomes Soares and Araújo (2015a,b), Jin et al. (2015b), Kadlec and Gabrys (2010), Kaneko et al. (2014) and Kadlec and Gabrys (2009a) change the order of the adaptation. This research is perhaps the most relevant to the current paper. In particular, Kadlec and Gabrys (2010) creates new experts when existing ones are not built on the relevant data, Gomes Soares and Araújo (2015a,b) create new experts when the predictive error on an instance is above a set threshold. In Kaneko et al. (2014) the predictive accuracy is assessed to switch between two predictive models. Again the, predictive accuracy is used to choose between just-in-time model creation and offset update in Jin et al. (2015b). Kadlec and Gabrys (2009a) present a plug and play architecture for preprocessing, adaptation and prediction which foresees the