Control Engineering Practice 72 (2018) 19–28

Contents lists available at ScienceDirect

Control Engineering Practice

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

Adaptive soft sensors for quality prediction under the framework of Bayesian network

Ziwei Liu, Zhiqiang Ge *, Guangjie Chen, Zhihuan Song

State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, Zhejiang, PR China

ARTICLE INFO

Keywords:
Model degradation
Adaptive soft sensor
Bayesian network
Quality prediction

ABSTRACT

Soft sensor is widely used to predict quality-relevant variables which are usually hard to measure timely. Due to model degradation, it is necessary to construct an adaptive model to follow changes of the process. Adaptive models—moving windows (MW), time difference (TD), and locally weighted regression (LWR) under the framework of Bayesian network (BN) are proposed in this paper. BN shows great superiorities over other traditional methods, especially in dealing with missing data and the ability of learning causality. Furthermore, the acquisition of variances in BN makes it possible to perform fault detection, on the basis of 3-sigma criterion. A debutanizer column and CO_2 absorption column are provided as two industrial examples to validate the effectiveness of our proposed techniques. In a debutanizer column, RMSE of MW-BN is decreased by 40% in comparison to MW-PLS. In a CO_2 absorption column, the largest absolute prediction error of TD-BN is reduced by approximate 7% when compared with that of TD-PLS. Furthermore, about 38% improvements of prediction precision can be achieved in LW-BN in contrast to LW-PLS.

1. Introduction

Soft sensor aims at constructing a model to estimate quality variables that are usually difficult to measure or have a long time delay measurement. It is essential to predict these quality-relevant variables accurately and timely in order to guarantee product quality. Mechanism-based and data-based methods are two major techniques to build a soft sensor model. With computer science and data technology developing rapidly, data-driven methods have gained more and more attention from both researchers and engineers (Ge, Song, & Gao, 2013; Ge & Chen, 2016; Zhu, Ge, & Song, 2017; Ge, Song, Ding, & Huang, 2017; Ge, 2017).

To date, various data-based models have been applied to describe the mathematical relationship between quality-relevant and easy-to-measure variables. Principal component regression (PCR) and partial least squares (PLS) are traditional linear models (Geladi & Kowalski, 1986; Tang, Yu, Chai, & Zhao, 2012). Due to their simplicity and practicability, they are widely used in industrial applications. Taken the uncertainty into consideration, a probabilistic form is formulated for both PCR and PLS (Ge, Gao, & Song, 2011; Zheng, Song, & Ge, 2016). Support vector machine, as a nonlinear modeling approach, is applied to soft sensor field and achieves high prediction accuracy (Jin et al., 2015; Yan, Shao, & Wang, 2004). Gaussian mixture model (GMM) shows great advantages at coping with multiphase processes, and the feasibility of GMM is validated in the literature (Yu & Qin, 2008; Yuan, Ge, & Song, 2014). Other nonlinear methods such as artificial neural networks (ANN) have also been employed to estimate those hard-to-measure variables (Himmelblau, 2008; Wu & Chai, 2010).

Unfortunately, model degradation is unavoidable due to process drifts and catalyst performance loss. That is to say, the initial model is unfit to describe the current process state. To solve this problem, the terminology of adaptive soft sensor was proposed. There are three mainstream techniques to deal with the failure of models—moving window (MW), time difference (TD), and locally weighted regression (LWR) (Cheng & Chiu, 2004; Fujiwara et al., 2009; Hazama & Kano, 2015; Jin et al., 2015; Kaneko & Funatsu, 2011a, b; Kaneko & Funatsu, 2015; Kim et al., 2013; Shao & Tian, 2015). It is hard to say which one is best owing to different applicability environments (Kaneko & Funatsu, 2013). MW is apt at managing changes of the slope between X and Y. TD can handle the drifts of both X and Y well. Moreover, one of the prominent merits with TD model is maintenance-free. Nevertheless, it is unable to cope with changes of slope. LWR is a particular case of just-in-time learning. Unlike MW and TD, the performance depends heavily on definition of similarity and weights of the training samples.

* Corresponding author.
E-mail address: gezhiqiang@zju.edu.cn (Z. Ge).

https://doi.org/10.1016/j.conengprac.2017.10.018
Received 8 July 2017; Received in revised form 3 September 2017; Accepted 30 October 2017

LWR performs quite well in disposing shifts of X but can do nothing to changes of Y and slopes. To sum up, all of three techniques can solve model degradation to some degree.

Bayesian network (BN) is regarded as one of the most powerful way for probabilistic modeling and analytics (Boudali & Dugan, 2005; Friedman, Linial, Nachman, & Peer, 2000). It has attracted more and more researchers and become a hot topic in artificial intelligence. BN is a kind of unique graphic expression of probability distribution, which is visual and easily-understood. Particularly, it has a gift for dealing with uncertainties. Consequently, BN has been applied to classification, clustering, prediction and causal reasoning successfully (Mori, Mahalec, & Yu, 2014; Nikovski, 2000).

In view of the properties, Bayesian network can be introduced for soft sensor modeling. Three favorable merits of BN stands out among traditional methods for predicting:

- BN is capable of coping with missing data. Since parameter learning of BN adopts Expectation Maximum algorithm (EM) and gives the maximum likelihood estimation through frequent iterations. It makes BN equipped with a natural advantage of dealing with missing values.
- Achieve the expectation and variance of y simultaneously. It is assumed that each node obeys a Gaussian distribution. Variance information and mean values of y can be obtained together by BN inference. Based on 3-sigma criterion, approximate 99.7% of samples fall in the interval $[\mu - 3\sigma, \mu + 3\sigma]$. Therefore, if the measured values exceed the 3-sigma interval continuously, a fault is more likely to happen in a process.
- Possess a desirable ability of learning causality. BN is a combination of probability, causal reasoning and graph theory. It incorporates structure learning and parameter learning, depicting causal relationships qualitatively and quantitatively, respectively.

With the purpose of giving a full play to the strengths, BN is combined with adaptive techniques to solve the model degradation problem in the present paper. Bayesian network with moving windows (MW-BN), Bayesian network with time difference (TD-BN), and locally weighted Bayesian network (LW-BN) are proposed. The feasibility of the developed methods will be verified through real industrial data in a debutanizer column and a CO_2 absorption tower. In addition, compared with traditional methods such as partial least squares with moving windows, those BN based adaptive soft sensors could realize higher prediction accuracy.

The remainder of the paper is structured as follows. In Section 2, a brief review of Bayesian network is presented, followed by detailed description of the proposed models. Section 4 is devoted to two industrial examples and comparative results with PLS based methods are also demonstrated. Finally, a conclusion is drawn in Section 5.

2. Bayesian network

Nowadays, researches of Bayesian network (BN) have focused on inference, learning of a network, and applications. Inference pays attention to a variety of efficient algorithms, and it comprises exact reasoning and approximate reasoning. Junction tree algorithm aimed at conditional Gaussian BN is developed in the form of information parameters under the framework of Hugin (Lauritzen, 1992). While the Lauritzen–Spiegelhalter junction tree for exact inference is also designed for Gaussian BN (Lauritzen & Jensen, 2001). However, computation efficiency is often restricted by the size of a junction tree. To solve the problem, a hierarchy of junction tree has been proposed for the restriction reduction (Wu & Wu, 2007). A junction tree algorithm for fault diagnosis is put into practice in an industrial tanks system (Ramírez Julio, Muñoz, & Gutierrez, 2009). Besides, a model of dynamic BN has been utilized to predict a possible future outcome in tissue engineering (Komarlu, Shao, Akar, & Bayrak Elif, 2017).

BN is a graphical expression form of mathematical model. It is made up of a graphic structure and a probabilistic description of dependences of practical variables. The essence of BN is a directed acyclic graph (DAG) consisting of observable nodes and hidden nodes.

Assume a Bayesian network $G = \langle N, A, \Theta \rangle$, every node $n \in N$ represents a physical variable, every edge $a \in A$ means dependent relationship between a parent node and a child node. Θ denotes conditional probability distributions related to nodes. Similar to other graphic models, the definition of a parent and child node is determined by the direction of an arrow which also indicates cause and effect.

Theory: Given a joint probability distribution $P\left(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\right)$, there is a Bayesian network corresponding to P, satisfying the expression

$$P\left(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\right) = \prod_{i=1}^{n} P\left(\mathbf{x}_i | \boldsymbol{\pi}_{\mathbf{x}_i}\right) \tag{1}$$

among which, $\boldsymbol{\pi}_{\mathbf{x}_i}$ denotes the parent node set of $\mathbf{x}_i$.

The theory gives the decomposition of a joint probability distribution, which can greatly decrease the complexity of a probabilistic model, especially when the number of nodes is so large. $\boldsymbol{\pi}_{\mathbf{x}_i} = \varnothing$ is an exception, meaning marginal distribution $P\left(\mathbf{x}_i\right)$.

The structure of BN illustrates probabilistic relations of nodes. Most often, in a soft sensor model, parent nodes are observable variables while a quality-relevant variable is hidden. Generally, there are three ways to determine a structure of BN: by expert experience, by learning from huge data, by combining data and expert experience (Chickeringn, 1996; Friedman & Koller, 2003). The structure determines the precision of prediction largely and changes of a structure may lead to a completely different result. Nevertheless, relying on data learning totally may not be always satisfying. Usually, a network is constructed by learning from training samples first and then necessary adjustments are made according to experience knowledge.

Parameter learning is performed right away after the structure of a BN model is determined. The purpose is to calculate the probabilistic distribution density (continuous variables) or probabilistic distribution table (discrete variables) under a certain structure. BN adopts expectation maximum algorithm (EM) to estimate optimal parameters. EM algorithm is an effective way to give the maximum likelihood estimation (MLE) through several iterations in spite of missing data existing. This may be a desirable merit for modeling with missing values. The estimation of parameters relies on relationships among missing data and unknown model parameters. It works by continual iterations until the estimated parameters converge to their MLE. Thanks to EM algorithm, BN is capable of predicting quality variables with a portion of missing values.

After a Bayesian network is set up, inference is a crucial step to obtain useful information. In brief, inference is to acquire variables' marginal distribution or maximum probability with newly increased evidences. Results are obtained by inference, and it is an application of the network. During inferring, newly observed evidence which incorporates further information should be added into the constructed BN.

One of the most widely-used inference algorithm is called junction tree. It belongs to the exact inference category, and has been popular for accuracy and efficiency. It functions by transforming BN into an undirected junction tree and computing probabilities through defining new patterns of passing information. During the procedure, messages will spread to every node so as to make the tree global consistent. After the implementation of message passing, the marginal distribution of a subset of variables can be achieved. Furthermore, a probability distribution of single variable is known through simple calculation. More detailed derivation of junction tree can be referred to Appendix A.1.

In this paper, junction tree inference is employed to acquire maximum posterior probability distribution in BN. In order to follow changes of a practical industrial process, BN can be cooperated with adaptive techniques for a better performance, which is called adaptive soft sensor.