Computers and Chemical Engineering 46 (2012) 94–104

Contents lists available at SciVerse ScienceDirect

Computers and Chemical Engineering

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

# The optimization of the kind and parameters of kernel function in KPCA for process monitoring

Mingxing Jia*, Hengyuan Xu, Xiaofei Liu, Ning Wang

*College of Information Science and Engineering, Key Laboratory of Integrated Automation of Process Industry, Northeastern University, Ministry of Education, No. 11, Lane 3, WenHua Road, HePing District, Shenyang 110819, Liaoning, China*

ARTICLE INFO

*Article history:*
Received 23 August 2011
Received in revised form 24 May 2012
Accepted 20 June 2012
Available online 30 June 2012

*Keywords:*
Chemical processes
KPCA
Kernel function
Optimization
Genetic algorithm
Fermentation

ABSTRACT

Kernel principal component analysis (KPCA) has been widely used in chemical processes monitoring due to its simple principle. However, how to select the kind and parameters of kernel function still limits the application of the method until now. In this paper, an optimization method based on genetic algorithm is developed to choose proper kind and parameters of kernel function. In this method, kernel kind and parameters are seen as decision variables of optimization, using correct monitoring rate, number of principal components, and statistical control limit of square prediction error (*SPE*) as multi-objective. For this specific problem, the fitness function, the algorithm of genetic selection, crossover and mutation are designed to ensure the diversity of kernel function and more selected chances of optimal individual in evolution process. A simple example and penicillin fermentation process are used to investigate the potential application of the proposed method; simulation results show that the proposed method is effective.



## 1. Introduction

In the field of process performance monitoring and fault diagnosis in chemical processes, multivariate statistical process control (MSPC), or process performance monitoring, has been extensively researched over the last decade as an alternative to knowledge-based approaches. One of cornerstones of MSPC is the principal component analysis (PCA) (Wold, Esbensen, & Geladi, 1987), which has played a significant role in dimensionality reduction, noise removal, and feature extraction from the original data set as a pre-processing. It divides data information into systematic part and noisy part. The systematic part contains the most variation in the data, while the noisy part has the least variation. For process monitoring, PCA uses two statistics, represented by Mahalanobis and Euclidean distances, to detect changes in the systematic part and the noisy one, respectively.

However, PCA is a linear method and most practical problems are nonlinear, thus using linear PCA in nonlinear problems can sometimes be inadequate (Palus & Dvorak, 1992). To handle the problem posed by nonlinear data, two types of methods are designed. One of methods is neural network. Kramer (1991) developed a nonlinear PCA method based on auto-associative neural networks. However, the network proposed by Kramer is difficult to train because it has five layers. Moreover, it is difficult to determine the number of nodes in each layer. Tan and Mayrovouniotis (1995) developed a nonlinear principal component analysis method based on input-training neural network, which has three layers. However, the input and weight training are received by the output, and the training is difficult. Dong and MaAvoy (1996) developed a nonlinear PCA approach based on principal curves and two three-layer neural networks. Saegusa, Sakano, and Hashimoto (2004) developed a hybrid neural network approach. The method uses *m* sub neural network; it can determine the number and order of principal components (PCs), but each neural network still has five layers and the output of the hidden layer in each sub neural network is designed as the input of the third hidden layer in follow-up sub network, and the training is difficult. Jia, Niu, Wang, and Zhao (2007) and Jia, Zhao, Wang, Mao, and Li (2008) developed a three-layer neural network nonlinear principal component analysis approach, and it has the same interpretation with the linear principal component analysis approach.

The other method is transform space, and the most representative approach is kernel principal component analysis (KPCA) (Mika et al., 1999; Schölkopf et al., 1999; Schölkopf, Smola, & Muller, 1998). The basic idea of KPCA is firstly to map the input space into a feature space via nonlinear mapping and then compute the PCs in

* Corresponding author. Tel.: +86 24 8368 0461; fax: +86 24 8368 0461.
*E-mail address:* jiamingxing@ise.neu.edu.cn (M. Jia).

0098-1354/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compchemeng.2012.06.023

that feature space. For any given algorithm that can be expressed solely in terms of dot products, i.e., without explicit use of the variables themselves, this kernel method enables the construction of different nonlinear versions of the original algorithm (Christianini & Shawe-Taylor, 2000).

Compared with other nonlinear methods, the main advantage of KPCA is that it does not involve nonlinear optimization (Schölkopf et al., 1998); essentially it only requires linear algebra, which makes it as simple as standard PCA. KPCA requires only the solution of an eigenvalue problem, and due to its ability to use different kernels, it can handle a wide range of nonlinearities. In addition, KPCA does not require the number of components to be extracted and specified prior to modeling. Due to these advantages, KPCA has been studied extensively. Lee proposed a new nonlinear process monitoring technique based on KPCA and a fault identification method (Choi, Lee, Choi, Lee, & Lee, 2005; Choi, Lee, Lee, Park, & Lee, 2005). Lee, Yoo, and Lee (2004) developed MKPCA approach, and then KPCA was promoted to batch process. Choi and Lee (2004) developed DKPCA approach, Sun, Tsung, and Qu (2007) developed EKPCA approach, Choi, Morris, and Lee (2008) developed multiscale KPCA approach, Liu, Kruger, Littler, Xie, and Wang (2009) developed MWKPCA approach, and "Ge, Yang, and Song (2009) developed kernel PCA-based monitoring approach for nonlinear processes" is changed to "Lee, Yoo, and Lee (2004) developed multiway KPCA (MKPCA) approach", and then KPCA was promoted to batch process. Choi and Lee (2004) developed Dynamic KPCA (DKPCA) approach, Sun et al. (2007) developed evolving KPCA (EKPCA) approach, Choi et al. (2008) developed multiscale KPCA approach, Liu et al. (2009) developed moving window KPCA (MWKPCA) approach, and Ge et al. (2009) developed kernel PCA-based monitoring approach for nonlinear processes.

Kernel function is the key to KPCA approach, and the kernel function chosen by KPCA is not arbitrary. It is required to satisfy Mercer's theorem. According to Mercer's theorem of functional analysis, there exists a mapping into a space where a kernel function acts as a dot product if the kernel function is a continuous kernel of a positive integral operator. A specific choice of kernel function implicitly determines the mapping and the feature space (Lee, Yoo, Choi, Vanrolleghem, Lee, 2004). In KPCA, the selection of a kernel function is the most important since the degree of capturing nonlinear characteristic of a system is dependent on it. Representative kernel functions are as follows:

(1) Polynomial kernel: $k(x, y) = (1 + \langle x, y\rangle)^d$
(2) Sigmoid kernel: $k(x, y) = \tanh(\beta_0\langle x, y\rangle + \beta_1)$
(3) Radial basis kernel: $k(x, y) = \exp(-(||x - y||^2/c))$

where $d$, $\beta_0$, $\beta_1$ and $c$ are specified a priori by the user. The polynomial kernel and radial basis kernel always satisfy Mercer's theorem, whereas the sigmoid kernel satisfies it only for certain values of $\beta_0$ and $\beta_1$ (Haykin, 1999; Lee, Yoo, Choi, et al., 2004). The general question of how to select the ideal kernel for a given monitoring process is an open problem. Furthermore, once the kernel is set, proper kernel parameters should be set. However, there is either no theoretical framework to specify the optimal values of kernel parameters. Generally, both kernel functions and their parameters are set by trail and error. First, one initialization is empirically chosen; then the changing trend around the initial value is investigated; finally the candidate which shows best performance is obtained (Shao & Rong, 2009; Zhang, 2009). That is, the performance of KPCA-based monitoring method largely depends on its kernel function.

A specific study on penicillin fermentation process suggests that polynomial kernel function is more effective to get nonlinear characteristics of penicillin fermentation process compared with other kernel functions (Lee, Yoo, Choi, et al., 2004). Generally, the study of how to choose the kind and parameters of kernel function for the process monitoring does not be reported currently. In the past few years, the similar problem is discussed in support vector machines (SVM). Support vector machines (SVM) were first suggested by Vapnik (1995) and have recently been used in a range of problems including pattern recognition (Pontil & Verri, 1998), bioinformatics (Yu, Ostrouchov, Geist, & Samatova, 2003), and text categorization (Joachims, 1998). When using SVM, two problems are confronted: how to choose the optimal input feature subset for SVM, and how to set the best kernel parameters. To design a SVM, one must choose a kernel function, set the kernel parameters and determine a soft margin constant C (penalty parameter). The Grid algorithm is an alternative to finding the best C and gamma when using the RBF kernel function. However, this method is time consuming and does not perform well (Hsu & Lin, 2002; LaValle & Branicky, 2002). Moreover, the Grid algorithm cannot perform the feature selection task. So, Huang and Wang (2006) proposed a genetic algorithm approach for feature selection and parameters optimization. This method is applied to software effort estimation by Adriano, Petronio, Ricardo, and Márcio (2010). The asymptotic behaviors of SVM are fused with genetic algorithm and the feature chromosomes are generated, which thereby directs the search of genetic algorithm to the straight line of optimal generalization error in the super parameter space (Zhao, Fu, Ji, Tang, & Zhou, 2011). Wua, Tzeng, and Lin (2009) proposed a novel and specialized hybrid genetic algorithm for optimizing all the SVR parameters simultaneously. In the same year, a hybrid of genetic algorithm–support vector machines (HGASVM) approach is presented in digital modulation classification area for increasing the support vector machines (SVM) classification accuracy (Engin, 2009). Based on these research results, in this paper, combined with KPCA process monitoring, an optimal model is established. In this model, kernel kind and parameters are seen as decision variables; maximum correct monitoring rate (CMR), minimum number of PCs and minimum statistical control limit of squared prediction error (*SPE*) are seen as multi-objective. An improved genetic algorithm (GA) is used to solve the model. Simulation results show that the proposed method has a good effect.

## 2. Method to determine kernel function and its parameters

### 2.1. Optimization index

In industrial application, for the process under study, the kernel function and its parameters are usually determined after testing the method's monitoring performance (shown by the *T*-square and *Q*/*SPE* statistics). This is called cross-validation. Cross-validation was originally employed to evaluate the predictive validity of linear regression equations and used to forecast a performance criterion from scores on a battery of tests (Mosier, 1951). Cross-validation based on the prediction error sum of squares (PRESS) and *R* ratio firstly is a popular statistical criterion to choose the number of factors in PCA (Krzanowski, 1987; Wold, 1978). When PCA is used as a dimension reduction method for classification, cross-validation is often used to determine the number of factors in the model such as Linear Discriminant Analysis (LDA). And the correct classification rate (CCR) of classification result is seen as an indicator of cross-validation. In this paper, cross-validation is also used to determine the optimal optimization index.

#### 2.1.1. KPCA

Consider a distribution consisting of $n$ data points $x_i \in R^m$ $i = 1, \ldots, n$. To derive KPCA, these data points are mapped into a higher dimensional feature space $H$, $\Phi : x_i \rightarrow H$, and assume $\Phi(x_i)$ mean-centered and variance-scaled. Calculate the covariance matrix $C^H$