Neurocomputing 147 (2015) 395–402

Contents lists available at ScienceDirect

Neurocomputing

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

Wavelet kernel entropy component analysis with application to industrial process monitoring

Yinghua Yang *, Xiaole Li, Xiaozhi Liu, Xiaobo Chen

College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110819, China

ARTICLE INFO

Article history:
Received 25 February 2014
Received in revised form
10 May 2014
Accepted 16 June 2014
Communicated by W. Yu
Available online 28 June 2014

Keywords:
Kernel entropy component analysis
Wavelet transform
Process monitoring
Fault identification
TE process

ABSTRACT

Aiming at the features that modem industrial processes always have some characteristics of complexity and nonlinearity and the process data usually contain both Gaussion and non-Gaussion information at the same time, a new process performance monitoring and fault detection method based on wavelet transform and kernel entropy component analysis (WT-KECA) is proposed in this paper. Unlike other kernel feature extraction methods, this method chooses the best principal component vectors according to the maximal Renyi entropy rather than judging by the top eigenvalues and eigenvectors of the kernel matrix simply. Besides, it can denoise and anti-disturb due to the application of wavelet transform. The proposed method is applied to process monitoring in the Tennessee Eastman (TE) process and the fault identification is realized. The simulation results indicate that the proposed method is more feasible and efficient in comparing to KPCA method.

1. Introduction

The monitoring of industrial processes and the detection of faults in those processes are especially significant procedures as the industrial process control system tends to be large-scale and complicated. The actual production data, however, inevitably has properties of noises, random disturbances. There exist some data transformation methods to tackle these problems. Among them, Principal Component Analysis (PCA) is a powerful technique for extracting structure from possibly high-dimensional data sets and widely utilized for process monitoring and fault diagnosis on account of its ability to handle high-dimensional, noisy, and highly correlated data. MacGregor and Kourti [1] established a PCA model from the training data and detected the abnormal behavior of online processes. However, PCA is a linear transformation method whose performance would degrade the monitoring performance greatly for some complicated cases in chemical industry process with particularly nonlinear characteristics [2].

To solve the problem posed by nonlinear data, Scholkopf proposed kernel principal component analysis (KPCA) method [3]. The main idea of KPCA is to map the input space into a feature space via nonlinear mapping and then perform PCA in a higher dimensional feature space. It has proven to be a very effective fault diagnostic method in process monitoring because of the main advantage that it does not involve nonlinear optimization [4]. However, due to the complexity of industry, the application of KPCA in process monitoring is not very good for some complicated industrial processes fault and may cause false alarms.

Kernel entropy component analysis (KECA) is a new method of data transformation and dimensionality reduction which has been proposed by Robert Jenssen recently [5]. KECA is founded on information theory and tries to preserve the maximum Renyi quadratic entropy estimated via Parzen window rather than depend on the second order statistics of the data set [6]. As a result, no limitation of Gaussian-like assumption is involved before we apply the KECA method. KECA may produce strikingly different transformed data sets whose data transformation is achieved by projecting onto the kernel PCA axes that contribute to the maximum entropy estimate of the input space data set in comparing to KPCA method with data transformation corresponding to the top eigenvalues of the kernel matrix. Some scholars have applied KECA on face recognition, audio emotion recognition, data clustering and denoising techniquewhich lead to better results than PCA and KPCA [7–9]. However, there were few reports about the reseach on the KECA applied in process monitoring. In addition, the actual production data of chemical process inevitably contain random and gross errors due to sensor noise, disturbances, instrument degradation, and human errors. So when we only apply KECA for process monitoring and fault detection, it will impact the effect of process information treatment or analysis and reduce the confidence degree of outcome by using these contaminated data. Hence straightly using KECA is not fit for fault detection very well.

In order to improve the effect of KECA method application in the field of fault identification, wavelet transform (WT) have

* Corresponding author.
E-mail address: yhyang@mail.neu.edu.cn (Y. Yang).

http://dx.doi.org/10.1016/j.neucom.2014.06.045
0925-2312/

shown that wavelet is an efficient tool for noise removal of noisy signal and it has found applications in a variety of fields in biomedical signal processing. Wavelet transform is a kind of time–frequency analysis and it provides a useful alternative to Fourier methods in the enhancement of nonlinear data. Wavelet has characteristics of well time frequency localization, special denoising ability and convenient extracting weak signals for signal processing [10] so it has found applications in a variety of fields in signal analysis, image processing, data compression, and process modeling.

In this paper, we apply KECA method to dynamic nonlinear process monitoring and it shows a better monitoring performance than other approaches in the simulation of TE process. What's more, according to the above discussion, this paper modifies KECA with wavelet (WT-KECA), and proposes a novel process monitoring algorithm. This method takes both advantaged of KECA and the wavelet method and it also shows a better performance.

The remaining of this paper is organized as follows: Section 2 explains the KECA and wavelet analysis algorithms. In Section 3, further discussion of WT-KECA on Process monitoring is described and Section 4 is the simulation process of TE and discussion. Finally, Section 5 concludes the paper.

2. Wavelet kernel entropy component analysis

According to the discussion in the previous section, the improved method based on wavelet analysis and kernel entropy component analysis is proposed in this paper, in which the original signal is firstly decomposed by wavelet analysis, then the kernel entropy component analysis method is applied to the preprocessed data. In the following section, the basic principle of the KECA and wavelet analysis will be introduced respectively.

2.1. Kernel entropy component analysis

The Renyi quadratic entropy [11] is given by

$$H(p) = -\log \int p^2(x)dx \tag{1}$$

where $p(x)$ is the probability density function of the data set, or sample, $D = x_1, \dots, x_N$. Because the logarithm is a monotonic function, we can consider the following quantity

$$V(p) = \int p^2(x)dx \tag{2}$$

In order to estimate $V(p)$, and hence $H(p)$, we may invoke a Parzen window density estimator described as

$$\hat{p}(x) = \frac{1}{N}\sum_{i=1}^{N} K_\sigma(x, x_i) \tag{3}$$

here, $K_\sigma(x, x_i)$ is the so-called Parzen window, or kernel, centered at x_i and with a width governed by the parameter σ.

Using the sample mean approximation of the expectation operator, we have

$$\hat{V}(p) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j}^{N} K_\sigma(x_i, x_j) = \frac{1}{N^2}\mathbf{1}^T\mathbf{K}\mathbf{1} \tag{4}$$

Here, the element (i, j) of the $N \times N$ kernel matrix $\mathbf{K}$ is $K_\sigma(x_i, x_j)$ and $\mathbf{1}$ is an $(N \times 1)$ vector containing all ones.

The Renyi entropy estimator may be expressed in terms of the eigenvalues and eigenvectors of the kernel matrix, which may be decomposed as $\mathbf{K} = \mathbf{E}\mathbf{D}_\lambda\mathbf{E}^T$, where $\mathbf{D}_\lambda$ is a diagonal matrix storing the eigenvalues $\lambda_1, \dots, \lambda_N$ and $\mathbf{E}$ is a matrix with the corresponding eigenvectors $e_1, \dots, e_N$ as columns. Rewriting (4), we then have

$$\hat{V}(p) = \frac{1}{N^2}\sum_{i=1}^{N}(\sqrt{\lambda_i}e_i{}^T\mathbf{1})^2 \tag{5}$$

Each term in Eq. (5) will contribute to the entropy estimate. This means that certain eigenvalues and eigenvectors will contribute more to the entropy estimate than others since the terms depend on different eigenvalues and eigenvectors. The eigenvalues and eigenvectors selected are the first l largest contribution to the entropy estimate in KECA so that the cumulative contribution rate of the selected Renyi entropy reaches 85% of all the Renyi entropy.

In KPCA, the nonlinear map from input space to feature space is given by $\phi : R^d \to F$ such that $x_t \to \phi(x_t), t = 1, \dots, N$.Let $\Phi = [\phi(x_1), \dots, \phi(x_N)]$ and u_i is the feature space principal axe. The projection of Φ onto theith principal axis u_i in the kernel feature space is defined as $P_{u_i}\phi = \sqrt{\lambda_i}e_i^T$.Eq. (5) therefore reveals that the Renyi entropy estimator is composed of projections onto all the kernel PCA axes. Certainly, only a principal axis e_i for which $\lambda_i \neq 0$ and $e_i^T\mathbf{1} \neq 0$ contributes to the entropy estimate. Hence, e_i is composed of a subset of KPCA axes but not necessarily those corresponding to the top l eigenvalues.

The kernel entropy component analysis procedure, as described above, is summarized as follows:

(1) Select the kernel function and get the kernel matrix $\mathbf{K}$;
(2) Compute the eigen-decomposition of $\mathbf{K} : \mathbf{K} = \mathbf{E}\mathbf{D}_\lambda\mathbf{E}^T$;
(3) Select the first l largest contribution to the entropy estimate in KECA;
(4) Calculate the kernel feature space data points, $\Phi_{eca} = \mathbf{D}_l^{1/2}\mathbf{E}_l^T$ and the components $\mathbf{T}_{N\times l} = \mathbf{K}_{N\times N}\mathbf{E}_{N\times l}$.

2.2. Wavelet transform for denoising

Wavelet analysis is a new method of time–frequency analysis, it has the characteristic of multi-resolution analysis and focus on the details of signal points to make the time–frequency domain analysis thus it has been called the “mathematical microscope”. In this section, we will introduce the basic theory of wavelet.

2.2.1. Discrete wavelet transform

In discrete wavelet analysis, $x(t)$ is decomposed on different scale s as follows:

$$x(t) = \sum_{j=1}^{K}\sum_{k=-\infty}^{\infty} d_j(k)\psi_{j,k}(t) + \sum_{k=-\infty}^{\infty} a_K(k)\phi_{K,k}(t) \tag{6}$$

where $\psi_{j,k}(t)$ are discrete analysis wavelets and $\phi_{K,k}(t)$ are discrete scaling functions, $d_j(k)$ are the detailed signals (wavelet coefficients) at scale 2^j and $a_K(k)$ is the approximated signal (scaling coefficients) at scale 2^K. The idea of discrete wavelet analysis is presented by means of a wavelet decomposition tree.

The discrete wavelet transform can be implemented by the scaling and wavelet filters

$$h(n) = \frac{1}{\sqrt{2}}\langle\phi(t), \phi(2t-n)\rangle$$

$$g(n) = \frac{1}{\sqrt{2}}\langle\psi(t), \psi(2t-n)\rangle = (-1)^n h(1-n) \tag{7}$$

The estimation of the detail signal at level j will be done by convolving the approximate signal at level $j-1$ with the coefficients $g(n)$. Convolving the approximate signal at level $j-1$ with the coefficients $h(n)$ gives an estimate for the approximate signal at level j. The decomposition scheme involves retaining every other sample of the filter output.