Anomaly detection based on kernel principal component and principal component analysis

Wei Wang*, Min Zhang, Dan Wang, Yu Jiang, Yuliang Li, Hongda Wu

Tianjin Key Laboratory of Wireless Communications and Power Transmission, Tianjin, China
E-mail: weiwang@mail.tjnu.edu.cn

Abstract: Nowadays, behind wall human detection based on UWB radar signal, which it had a strong anti-jamming performance, was an important problem. In this setting, principal component analysis(PCA) as an anomaly detection method was used, but PCA could only deal with linear data. Thus, we introduced the kernel principal component analysis(KPCA) for performing a nonlinear form of principal component analysis(PCA). We obtained the different state data based on UWB radar signal for the behind wall human detection. These data were used as training and testing data to calculate the squared prediction error(SPE) values that detect anomalies. The experimental results showed that the introduced approach of KPCA effectively captured the nonlinear relationship in the process data and showed superior process monitoring performance compared to linear PCA.

Keywords: anomaly detection, radar signal, KPCA, PCA.

1. Introduction

In recent years, identifying abnormal data is the basis and premise of a lot of work in some applications, such as fraud detection and risk analysis of telecommunications, insurance, bank; crime in electronic commerce; the intrusion detection in computer; the used of radar data for behind wall human detection and so on [1]. Overall, detection of volume abnormal information is becoming more and more important in our life.

In recent work, principal component analysis(PCA) is a multivariate statistical tool widely used in industry for process monitoring [2,3,4]. The PCA based process monitoring scheme mentioned above is based on the assumption that the process behaves linearly [5]. However, when a process is nonlinear, the monitoring of a process using a linear PCA model might not perform properly [6]. In order to address this issue, several nonlinear PCA methods have been proposed [7], but these methods have a poor performance on reconstruction based contribution for process monitoring [8]. In this paper, we use the kernel principal component analysis (KPCA) to detect abnormal nonlinear data. Finally, KPCA has been applied more successful than other nonlinear methods in process monitoring and construction of nonlinear active shape models. KPCA fault detection indices is expressed in terms of a kernel vector mapped by the kernel function, rather than the intangible vector in the feature space.

In this paper, we use the kernel principal component analysis(KPCA) to solve nonlinear problem. According to the theory of pattern recognition, the low dimensional space linearly inseparable pattern through nonlinear mapping to high dimensional feature space [9] can achieve linear separable by kernel function theory. The goal of this paper is in order to behind wall human anomaly detection of UWB radar signal [10] and compare the advantages and disadvantages of the two algorithms of PCA and KPCA for anomaly detection based on dimension reduction.

The arrangement of this paper is as follows. In section 2, we introduced the theoretical knowledge of PCA and KPCA. Then, the squared prediction error(SPE) and confidence limits were calculated. In section 3, we used the different state data of UWB radar signal to simulate experiment and obtain the results. Finally, we give a conclusion in section 4.

2. Theory

2.1 The anomaly detection based on PCA

Principal component analysis (PCA) [2]is essentially a theory of perfect and feasible linear dimensionality reduction algorithm. The main idea of PCA: in the sense of global minimum reconstruction error, the high dimensional observation data is projected onto the low dimensional principal subspace [11]. A few unrelated and main new variables are produced in the original variables, which contain more information.

Given training data set $X=[x_1,\cdots,x_N], x_i\in R^m, i=1,\cdots,N$ .We use the first K dimensional data instead of m dimensional data to reduce the dimensionality and debase the computational complexity. So, in practice, it is important to select a suitable value for K .We assume that the data has been centered ($\sum_i X_i=0$), and the covariance matrix C can be defined as

$$C=\frac{1}{N}\mathrm{x_i x_i}^T=\frac{1}{N}XX^T=U\Lambda U^T \tag{1}$$

Given that $U=[U_1,\cdots,U_K]$ are the principal eigenvectors of C corresponding to the largest K eigenvalues $\lambda_1,\cdots,\lambda_K$,the residual subspace projection of vector is $p=(I-UU^T)$. Thus, for any data x, the projective coordinate of the residual subspace is $z=px=(I-UU^T)x$.If z follows a multivariate normal distribution, the squared prediction error (SPE) [7] statistic is given as

$$t_{SPE}=\|z\|_2^2=\|(I-UU^T)x\|_2^2 \tag{2}$$

and follows a noncentral chi-square distribution under the null hypothesis that the data is 'normal'. Hence, rejection of the null hypothesis can be based on whether t_{SPE} exceeds a certain threshold corresponding to a desired false alarm rate β .In [12], the Q-statistic was used to compute as the threshold, it is expressed as

$$Q_\beta=\theta_1[\frac{C_\beta\sqrt{2\theta_2h_0^2}}{\theta_1}+1+\frac{\theta_2h_0(h_0-1)}{\theta_1^2}]^{\frac{1}{h_0}} \tag{3}$$

Where $\mathrm{h}_0=1-\frac{2\theta_1\theta_3}{3\theta_2^2}$, $\theta_i=\sum_{j=K+1}^{N}\xi_j^i$ for $i=1,2,3$, $c_\beta=(1-\beta)$ percentile in a standard normal distribution Q_β ,and ξ_{j}, $i=1,\cdots,M$ are the eigenvalues of C. An abnormally is detected when $t_{SPE}>Q_\beta$.

2.2 The anomaly detection based on KPCA

KPCA is the linear dimensionality reduction algorithm in the feature space, we have mapped observation data X ($x_1,\cdots,x_N\in R^m$, N is the number of samples, m is the dimension of the measurement variables) onto the feature space($x\to\Phi(x)$), then the nonlinear problem in the input space is transformed into a linear problem in the feature space. The covariance matrix of feature space [4]

$$C^F=\frac{1}{N}\sum_{j=1}^{N}\Phi_j(x)\Phi_j(x)^T=\frac{1}{N}\sum_{j=1}^{N}\Phi(X)\Phi(X)^T \tag{4}$$

We assume that $\sum_{k=1}^{N}\Phi(x)=0$ and the $\Phi(\cdot)$ is a nonlinear mapping vector from the input space to the feature space. Note that the feature space is arbitrary and infinite, to obtain the covariance matrix, the eigenvalue problem must be solved.

$$\lambda v=C^F v \tag{5}$$

Where eigenvalues $\lambda\geq 0$ and $v\in F$,we obtain the λ and v by the Eq.(5), and corresponding descending order, we take the first k principal component to represent the entire m dimensional data to reach the purpose of dimensionality reduction. The Eq. (4) is brought the formula of $C^F v$:

$$C^F v = (\frac{1}{N}\sum_{j=1}^{N}\Phi(x_j)\Phi(x_j)^T)v = \frac{1}{N}\sum_{j=1}^{N}\left\langle \Phi(x_j, v)\right\rangle \Phi(x_j) \tag{6}$$

Where $\langle x, y\rangle$ denotes the dot product between x and y .Given the $\lambda v = C^F v$ multiply both sides simultaneously $\Phi(x_k)$, the left and right sides of an equation are invariant. The equivalent to $\lambda\left\langle \Phi(x_k), v\right\rangle = \left\langle \Phi(x_k), C^F v\right\rangle, k = 1,\cdots, N$ .And the eigenvector of the characteristic equation can be expressed by linear of $\Phi(x_1), \Phi(x_2), ..., \Phi(x_n)$.

$$v = \sum_{i=1}^{N}\alpha(i)\Phi(x_i) \tag{7}$$

We define a kernel matrix K by $N * N$,Combining Eqs. (6) and (7), we can get

$$N\lambda\alpha = K\alpha \tag{8}$$

Where $\alpha = [\alpha_1, ..., \alpha_N]^T$, For nonzero eigenvalues, a justification of this procedure is given in anukool et al [7]. The main components of the sample data x are extracted by projecting $\Phi(x)$ onto eigenvectors v_k in feature space, where $k = 1,\cdots, d$.

$$y_k = \left\langle v_k, \Phi(x)\right\rangle = \sum_{i=1}^{N}\alpha_i^k \left\langle \Phi(x_i), \Phi(x)\right\rangle \tag{9}$$

Before we assume that the data has been standardized, but in the actual application, the data is not standardized. This can be done by substituting the kernel matrix K with

$$K = K - 1_N K - K 1_N + 1_N K 1_N \tag{10}$$

Where

$$1_N = \frac{1}{N}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \in R^{N\times N} \tag{11}$$

In feature space, the squared prediction error(SPE) is defined [13]:

$$SPE = \left\|\Phi(x) - \Phi_p(x)\right\|^2 = \sum_{j=1}^{N} t_j^2 - \sum_{j=1}^{p} t_j^2 \tag{12}$$

We give the expression of the SPE confidence limit:

$$SPE_\alpha \sim g\chi_h^2 \tag{13}$$

Following a chi-square distribution. In Eq. (13), g is a weighting parameter included to account for the large of SPE, and h accounts for the degree of freedom.

If a and b are the estimated mean and variance of SPE, then g and h can be obtained by $g = \frac{b}{2a}$ and $h = \frac{2a^2}{b}$. An abnormally is detected when $SPE > g\chi_h^2$.

3 Simulation and experimental results

The UWB radar signal has many merits, such as: good anti-jamming ability, high range resolution, good target recognition and so on [14]. The two algorithms of PCA and KPCA based on dimension reduction are applied in many fields [15]. In this section, the monitoring results of PCA and KPCA are compared and the proposed monitoring method was applied to fault detection in a simple example of the behind wall human detection based on UWB radar signal. In this paper, we have two main purpose of the experiment: one is the target identification of abnormal data; the other is the comparison of two algorithms.

In this setting, we use the Gaussian kernel function [13] to calculate the projection data ($K(x, y) = \exp(-\frac{\|x-y\|^2}{c})$). Here the parameter c is set up $c = rm\sigma^2$,where r is a constant that is determined by consideration of the process to be monitored, m is the dimension of the input space, and σ^2 is the variance of the data. After testing the monitoring performance for various values of c ,we found that $c = 10m\sigma^2$ is appropriate for monitoring processes with various faults. Although the value of c is dependent upon the system under study, we found that the Gaussian kernel function is the best for monitoring the nonlinear processes used as examples in the present work. And the kernel function provides a low dimensional subspace in the characteristics of the spatial distribution of vector mapping method.

Based on UWB radar signal, we use the measured unmanned data as the normal data and some people after the wall as abnormal data. We calculate the SPE value and confidence limits by two algorithms of PCA and KPCA respectively. Then compare the advantages and disadvantages of the two algorithms for data recognition. In the simulation results, the unmanned data distribution as the training set and the testing set with the normal data for the first 250 samples and the abnormal data with the interference for the after 250 samples are plotted in Fig.2 and 3. In the following figure, the red line, which the SPE value represents the normal under the red line otherwise represents abnormal, show the 95% confidence limits of the threshold. In the picture, the horizontal axis shows the number of samples data and the vertical axis represents the SPE values.

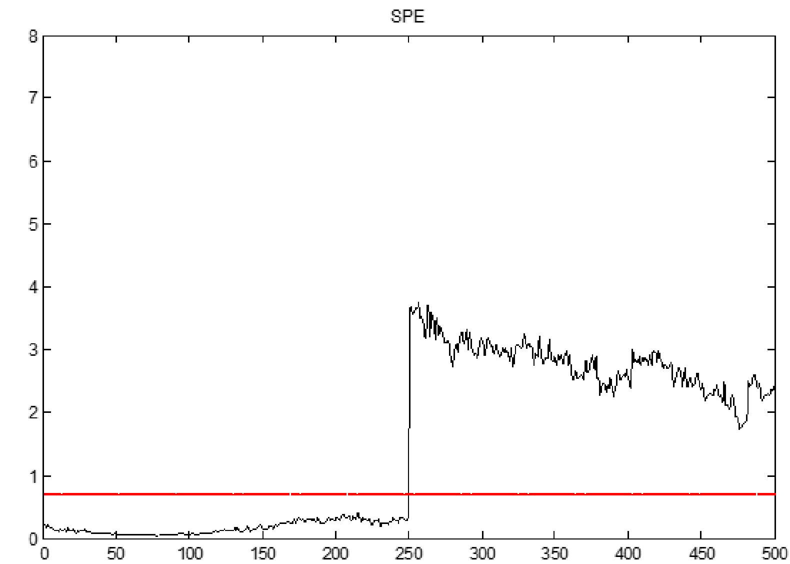

Fig.2 SPE of the big data used PCA

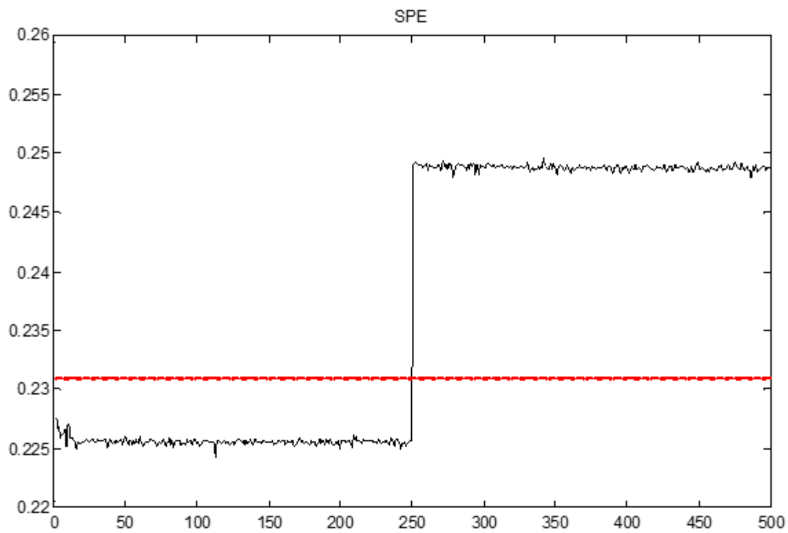

Fig.3 SPE of the big data used KPCA

The SPE chart for PCA monitoring of the process with the testing data is shown in Fig.2.The 95% confidence limits is also shown in this figure. It is evident from this chart that PCA detected data of SPE had large fluctuations and poor stability. However, applying KPCA to the same process data gives the results presented in Fig.3.KPCA shows better stability and smaller error between data samples in comparison to PCA. Overall, it shows the ability of anomaly detection in the chart of Fig.2and Fig.3, but the stability of KPCA is better than PCA in normal and abnormal data,

4. Conclusions

A framework of network anomaly detection using kernel function has been introduced. This paper introduces the approach to process monitoring that uses KPCA to achieve multivariate statistical process control. KPCA can efficiently compute compositions in high-dimensional feature spaces by means of integral operators with nonlinear kernel functions. KPCA has some main advantages with compared to other nonlinear methods: 1) the calculations in KPCA are as simple as in standard PCA; 2) no nonlinear optimization is involved; and 3) the number of compositions need not be specified prior to modeling. We note that a simple calculation of the SPE in the feature space is suggested setting similar to the paper [9]. The proposed monitoring method is applied to fault detection in anomaly detection of UWB radar data. This consequence demonstrated that the proposed approach of KPCA can effectively capture nonlinear relationships in process variables and that, when used for process monitoring, it shows better performance than linear PCA.

Acknowledge

This paper is supported by Natural Youth Science Foundation of China (61501326, 61401310), and Natural Science Foundation of China (61271411). It also supported by Tianjin Research Program of Application Foundation and Advanced Technology (15JCZDJC31500), and Tianjin Science Foundation(16JCYBJC16500).

References

[1] Ling H. In-Network PCA and Anomaly Detection. Advances in Neural Information Processing Systems 19-Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada. 2006,12.
[2] Duc-Son Pham, Svetha Venkatesh, MihaiLazarescu, Saha Budhaditya. Anomaly Detection in Large-Scale Data Stream Networks. Journal of Data Mining and Knowledge Discovery, DOI:10.1007/s10618-012-0297-3,2014:145-189.
[3] Pham, D.S., B., Lazarescu, M. and Venkates,S. Scalable network-wide anomaly detection using compressed data. Curtin University of Technology, Perth, W.A.2009.
[4] Scholkopf, Bernhard; Smola, Alexander; Muller, Klaus-Robert. Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation 10 (5),1998,1:1299–1399.
[5] Herve Abdi. Lynne J. Williams. Principal Component Analysis. Published in Encyclopedia of Statistics in Behavioral Science.2014,9.

[6] Jackson, J. Edward; Mudholkar, Govind S. Control Procedures for Residuals Associated with Principal Component Analysis. Technometrics, Vol.21(3),1979,8:341-349
[7] Lakhina, Anukool; Crovella, Mark; Diot, Christophe. Diagnosing Network-Wide Traffic Anomalies. Conference on Computer Communications,2004,8:219-230
[8] Alcala, Carlos F.; Qin, S. Joe. Reconstruction-based Contribution for Process Monitoring with Kernel Principal Component Analysis. Proceedings of the 2010 American Control Conference, 2010,6:7022-7027.
[9] Yin,Shen; Jing, Chen; Hou, Jian; Kaynak, Okyay; Gao, Huijun. PCA and KPCA Integrated Support Vector Machine for Multi-Fault Classification. Proceedings of the 42nd Conference of the Industrial Electronics Society.2016,10:7215-7220.
[10] Gang Lei; Ting Jiang. Target detection and recognition by UWB communication signal based on S-transform and matrix dimension reduction. Communications Workshops(ICC),2013 IEEE International Conference on. IEEE,2013:936-940.
[11] Zeng Xian Hua, Research on the related problems of spectral methods in popular learning. Beijing Jiaotong University.2009.6:26-30.
[12] Lee,Jong-Min; Yoo, ChangKyoo; Choi, Sang Wook; Vanrolleghem, Peter A; Lee, In-Beum. Nonlinear process monitoring using kernel principal component analysis. Published in Chemical Engineering Science.2004,1:223-234.
[13] C. J. Twining; C.J. Taylor. Kernel Principal Component Analysis and the Construction of Non-Linear Active Shape Models. Imaging Science and Biomedical Engineering.
[14] Jing Li; Zhaofa Zeng; Jiguang Sun; Fengshan Liu. Through-Wall Detetion of Human Being's Movement by UWB Radar. Article in IEEE Geoscience and Remote Sensing Letters.Vol.9(6),2012,11:1079-1083.
[15] Kramer, Mark A. Nonlinear principal component analysis using autoassociative neural networks. Sourced AIChE Journal.1991,2:233-243