Parameter Optimization in KPCA for Rotating Machinery Feature Vector Dimensionality Reduction

Lingli Jiang[1,a] Ping Li[2,b] and Siwen Tang[1,c]

[1.]Engineering Research Center of Advanced Mining Equipment, Ministry of Education, Hunan University of Science and Technology, Xiangtan 411201, China

[2.]Hunan Provincial Key Laboratory of Health Maintenance for Mechanical Equipment, Hunan University of Science and Technology, Xiangtan 411201, China

[a]linlyjiang@163.com   [b]pingyun868@163.com   [c]siw_tang@hnust.edu.cn

Key words: KPCA, kernel parameter, dimensionality reduction, rotating machinery

Abstract. This research aims at revealing the rules that the impact of kernel function and its parameters on the performance of kernel principle component analysis (KPCA) for dimensionality reduction. KPCA was performed on nine databases by using different kernel functions and a series of equal space kernel parameters. The relation charters between kernel parameters and the number of kernel principle components were constituted. It found that the Gussian kernel and its parameter above 2^5 are the best choice for rotating machinery feature vector dimensionality reduction by using KPCA. This study presents a reference and gist for the application of KPCA in rotating machinery fault diagnostic case.

Introduction

The original features extracted from test data in mechanical fault diagnosis can characterize device state[1]. Some features are valid to classification while others are invalid feature. Feature selection and extraction is necessary, which could remove invalid features and reduce feature space dimension to ensure the performance of classifiers and reduce the classification complexity[2]. PCA is a common method for feature dimension reduction. But it takes linear combination for feature extraction. The problem of random association between different model category and feature vector is handled difficultly by PCA[3-4]. Kernel principle component analysis (KPCA) is a nonlinear PCA developed by using the kernel method. It has a better performance than PCA for handling the nonlinear problem. Rotating mechanical failures often show a nonlinear behavior, the method based on KPCA is more suitable for rotating machinery fault feature dimensionality reduction [5-6].

KPCA is introduced into kernel function, so kernel function and its parameters selection must be involved. Existing studies showed that the kernel function and its parameters have a significant influence on the performance of KPCA[7-8]. According to this problem, a method that searches an approximate kernel function parameter value in a finite range based on the similarity between the given matrix and the kernel function matrix was proposed by Wang Xinfeng, Qiu Jing et al[9].

However, the kernel function and its parameters of reference matrix were assumed, while suitable reference matrix may not be found in practical application. The kernel function parameter optimized by particle swarm optimization algorithm was proposed by Wei Xiuye et al[10]. Intelligent algorithm for kernel parameter optimization was widely researched in SVM, by using which can only get the final optimal value but cannot find the rules that the impact of kernel function and its parameters on the performance of KPCA[11-12]. This paper took nine feature libraries as analysis purposes. They are time-domain statistical features, energy features and entropy features in different frequency band based on wavelet packet decomposition and empirical mode decomposition, time series features from rolling bearing, multi-sensor energy features, entropy features, correlative dimension features from gear. KPCA was performed on nine databases by using different kernel functions and a series of equal space kernel parameters. The rules of kernel function and its parameter optimization in KPCA was summarized, which can provide reference and gist for kernel function and its parameters optimization in KPCA for rotating machinery fault diagnostic case.

Theory

KPCA is one approach of generalizing linear PCA into nonlinear case using the kernel method. The idea of KPCA is to firstly map the original input vectors x_t into a high-dimensional feature space $\phi(x_t)$ and then to calculate the linear PCA in $\phi(x_t)$. The linear PCA in $\phi(x_t)$ corresponds to a nonlinear PCA in x_t. By mapping x_t into $\phi(x_t)$ whose dimension is assumed to be larger than the number of training samples l, KPCA solves the eigenvalue problem (1).

$$\lambda_i u_i = \tilde{C} u_i, \qquad i = 1, \cdots, l \qquad (1)$$

Where $\tilde{C} = \dfrac{1}{l} \sum_{t=1}^{l} \phi(x_t) \phi(x_t)^T$ is the sample covariance matrix of $\phi(x_t)$. λ_i is one of the non-zero eigenvalues of $\tilde{C}$. u_i is the corresponding eigenvector.

Eq. (1) can be transformed to the eigenvalue problem (2).

$$\tilde{\lambda}_i \alpha_i = K \alpha_i, \qquad i = 1, \cdots, l \qquad (2)$$

Where K is the $l \times l$ kernel matrix. The value of each element of K is equal to the inner product of two vectors x_i and x_j in the high-dimensional feature space $\phi(x_i)$ and $\phi(x_j)$. That is, $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$. As the calculations of the dot product $\phi(x_i) \cdot \phi(x_j)$ are all replaced with the kernel function $K(x_i, x_j)$, the mapping of $\phi(x_t)$ from x_t is implicit. The elegance of using K is that one can deal with $\phi(x_t)$ of arbitrary dimensionality without having to compute $\phi(x_t)$ explicitly. Any function satisfying Mercer's condition can be used as K. $\tilde{\lambda}_i$ is one of the eigenvalues of K, satisfying $\tilde{\lambda}_i = l\lambda_i$. α_i is the corresponding eigenvector of K, satisfying $u_i = \sum_{j=1}^{l} \alpha_i(j) \phi(x_j)$ ($\alpha_i(j), j = 1, \cdots l$ are the components of α_i). Furthermore, for assuring the eigenvectors of $\phi(x_t)$ is of unit length $u_i \cdot u_i = 1$, each α_i must be normalized using the corresponding eigenvalue by $\tilde{\alpha}_i = \alpha_i / \sqrt{\tilde{\lambda}_i}$.

Finally, based on the estimated $\tilde{\alpha}_i$, the principal components for x_t is calculated by

$$s_t(i) = u_i^T \phi(X_t) = \sum_{j=1}^{l} \tilde{\alpha}_i(j) K(x_j, x_t), \qquad i = 1, \cdots, l \qquad (3)$$

In addition, for making the sample input vectors in $\phi(x_t)$ centered $\sum_{t=1}^{l} \phi(x_t) = 0$, in Eq. (3), the kernel matrix on the training set K and on the testing set K_t are, respectively, modified by

$$\tilde{K} = (I - \frac{1}{l} 1_l 1_l^T) K (I - \frac{1}{l} 1_l 1_l^T), \qquad (4)$$

$$\tilde{K}_t = (K_T - \frac{1}{l} 1_{l_t} 1_l^T K)(I - \frac{1}{l} 1_l 1_l^T), \qquad (5)$$

where I is l-dimensional identity matrix. l_t is the number of testing data points. 1_l and 1_{l_t} represent the vectors whose elements are all ones, with length l and l_t, respectively. K_t represents the $l_t \times l$ kernel matrix for the testing data points.

From Eq. (3), it can be found that KPCA can extract more number of principal components than PCA as the maximal number of principal components in KPCA is l, instead of m. Same as PCA, the dimension of s_t can also be reduced in KPCA if only considering the first several eigenvectors. By using the kernel method to implement nonlinear PCA, the other properties of PCA as described in (1)–(3) are all retained in KPCA except these characteristics are considered in $\phi(x_t)$.

Kernel function and its parameters have a great influence on the performance of KPCA. The dimension n of vector space which obtains via nonlinear mapping from original p-dimensional vector is unknown. The l-dimensional kernel principal component is obtained by executed PCA in feature space. The PCA properties show that $l \leq n$ must be contented while $l \leq p$ may not be met. If $l \geq p$ is obtained by KPCA, superiority of KPCA that reduce feature vector dimension will not be exploited. This study performs KPCA on the nine feature libraries and investigates the impact of kernel function and its parameters on the performance of KPCA.

Nine analysis purposes are listed as follows:

Feature library 1: Time-domain statistical features of rolling bearing;

Feature library 2: Energy features of rolling bearing in different frequency band based on 3-layer wavelet packet decomposition;

Feature library 3: Entropy features of rolling bearing in different frequency band based on 3-layer wavelet packet decomposition;

Feature library 4: First eight IMFs energy features of rolling bearing based on EMD;

Feature library 5: First eight IMFs entropy features of rolling bearing based on EMD;

Feature library 6: Time series features of rolling bearing;

Feature library 7: Multi-sensor energy features of gear;

Feature library 8: Multi-sensor entropy features of gear;

Feature library 9: Multi-sensor correlative dimension features of gear;

Experiment

Experiments were performed on Machinery Fault Simulator (MFS) from SpectraQuest, Inc. shown in Fig. 1. It can simulate most of faults that commonly occur in rotating machinery, such as misalignment, unbalance, resonance, rolling bearing faults, gearbox faults, and so on. The simulator has a range of operating speeds up to 6000 rpm. In this work, the simulator is composed of a motor, a coupling, a testing rolling bearing fitted on the left of the shaft near the motor, a working rolling bearing on the other side, a bearing load and a shaft. The MFS provides a bearing fault kit

consisting of one inner race defect, one outer race defect, one with ball defect, and one combination of defects for performing experiments and studying bearing fault diagnosis, and provides one chipped tooth and one missing tooth for performing experiments and studying gear fault diagnosis.


Fig. 1.   Machinery Fault Simulator

The shaft rotating speed was obtained by a laser speedometer. Acceleration signals were measured using the Dewetron 16 channels data acquisition system and the IMI 603C01 accelerometers. The data was stored in .mat format for further Matlab operation.

Results and discussion

PCA / KPCA often take the principle components whose cumulative percentage is more than 0.85 in practical application. Because information of original feature vectors is concentrated in the principle components whose contribution ratio is high. After execute KPCA on 9 feature libraries，the relation between Gaussian and polynomial kernel parameters and the number of kernel principle component whose cumulative percentage is more than 0.85 are shown in Figure 2 and Figure 3.
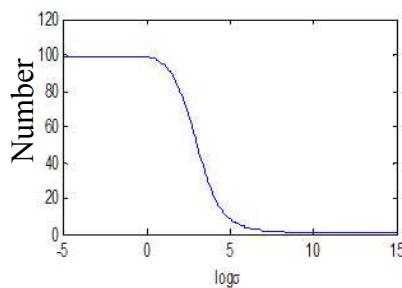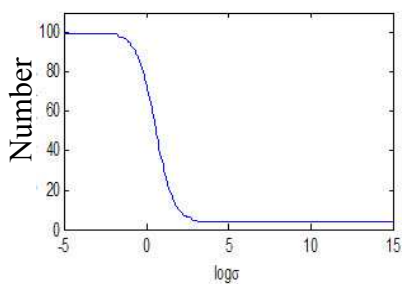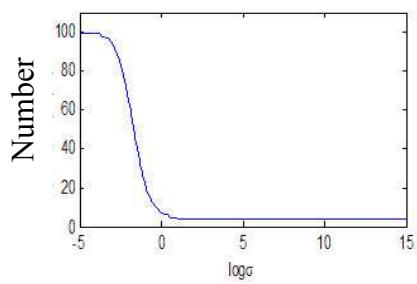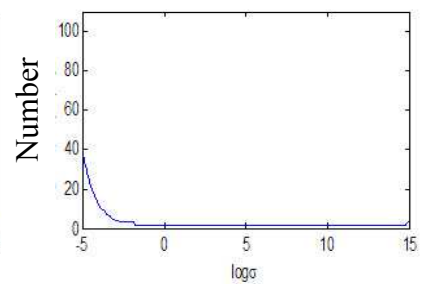

（a）Feature library 1   （b）Feature library 2   (c) Feature library 3
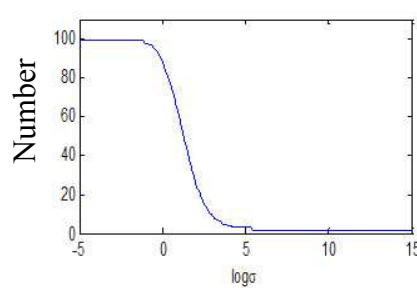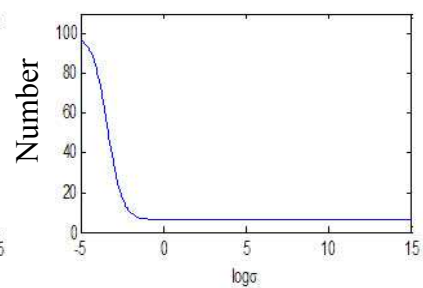
（d）Feature library   4   （e）Feature library 5   (f) Feature library 6
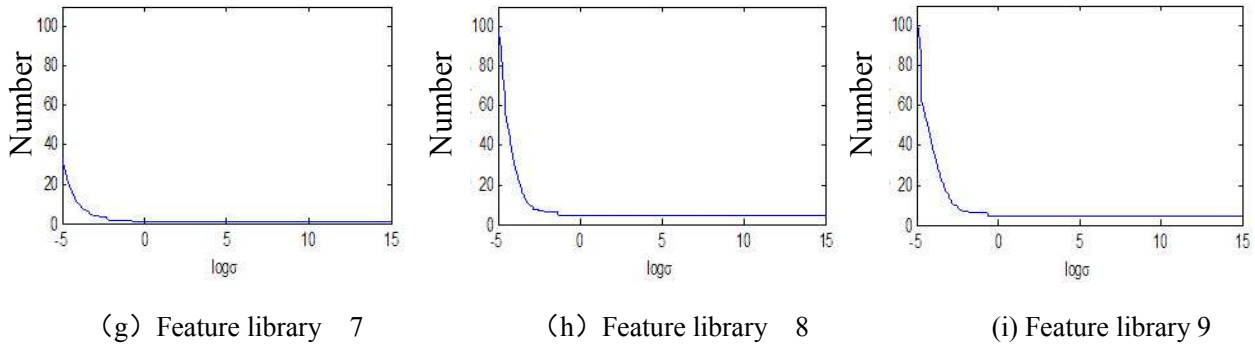
（g）Feature library 7   （h）Feature library 8   (i) Feature library 9

Fig. 2. The impact of Gaussian kernel parameters on the performance of KPCA

（a）Feature library 1   （b）Feature library 2   (c) Feature library 3

（d）Feature library 4   （e）Feature library 5   (f) Feature library 6

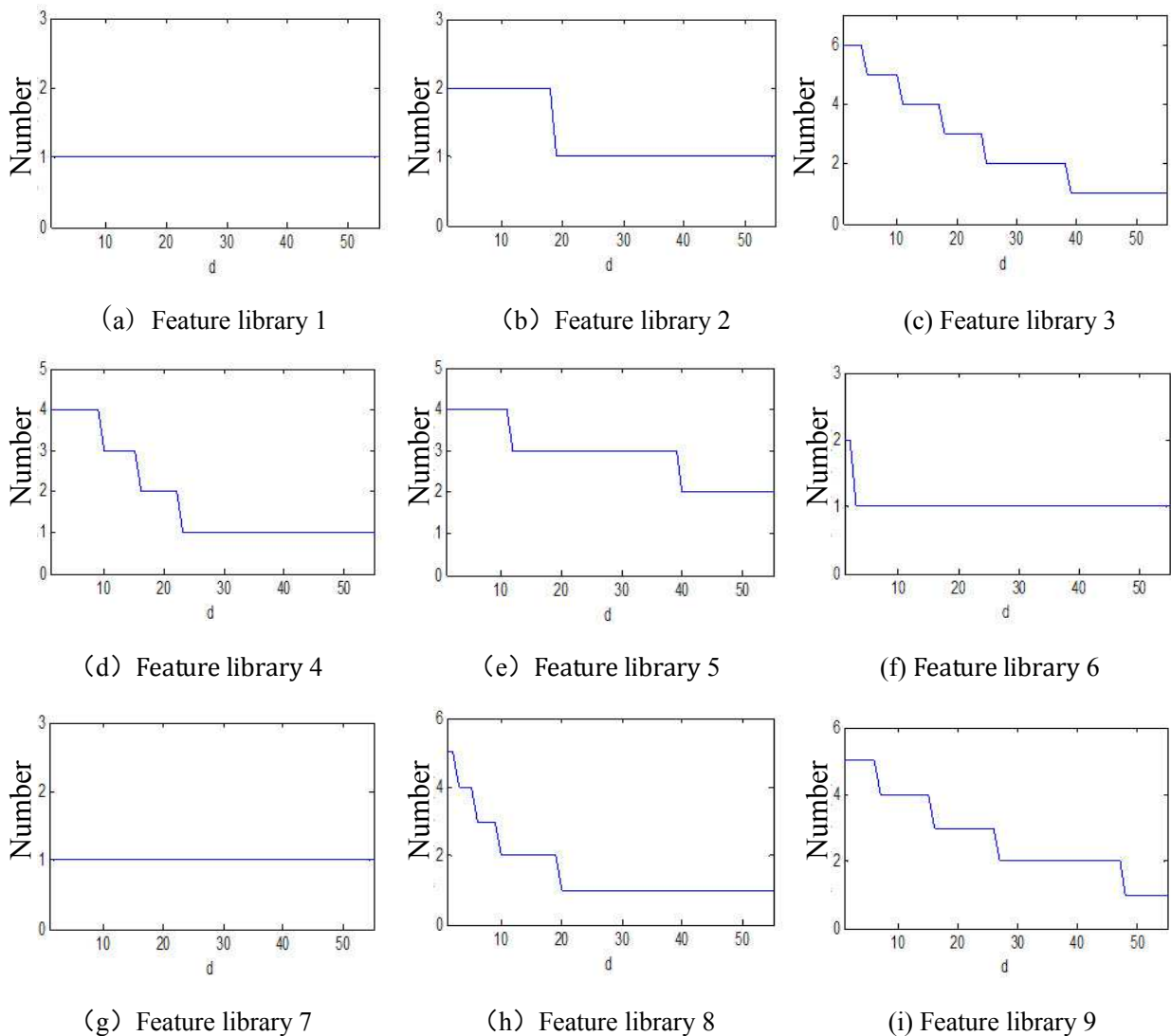（g）Feature library 7   （h）Feature library 8   (i) Feature library 9

Fig. 3. The impact of polynomial kernel parameters on the performance of KPCA

It can be found from Fig 2 and Fig3 that：

1. If it is the Gussian kernel, its parameter has a huge influence on KPCA for dimensionality reduction. The number of kernel principle components is even above 100 while the original feature vector dimension is under 10 if an inappropriate kernel parameter choice. As the kernel parameter increases, the number of kernel principle components is degressive. When the parameter is up to 2^5, the number of kernel principle components begins to convergence.

2. If it is the the polynomial kernel, its parameter has a smaller influence on KPCA for dimensionality reduction than using Gussian kernel. As the kernel parameter increases, the number of kernel principle components is degressive and convergence finally.

3.Cluster analysis found that the kernel principle component obtained a better cluster result by using the Gussian kernel than the polynomial kernel.

Conclusion

The Gussian kernel and its parameter above 2^5 are the best choice for rotating machinery feature vector dimensionality reduction by using KPCA.

Acknowledgements

Reference:

[1] A. K. S. Jardine, D. Lin and D. Banjevic: Mechanical Systems and Signal Processing 20(2006), p.1483-1450.
[2] Y. Xu, D. Zhang, F. X. Song, J. Y.   Yang et al. : Neurocomputing, 2007,70(4-6): 1056-1061.
[3] J. D. Shao, G. Rong, J. M. Lee: Chemical Engineering Research and Design, 87（11）： 1471-1480.
[4] H.X. Wang, Z.L. Hu：Neurocomputing, 71（1-3）：459-462.
[5] Y.W.   Zhang, Ch. Ma: Chemical Engineering Science, 66(1):64-72.
[6] L. J. Cao, K. S. Chua, W. K. Chong, et al.: Neurocomputing, 55(1-2): 321-336.
[7] P.L. Cui, J. H. Li, G. Z. Wang: Expert Systems with Applications, 34, (2):1210-1219.
[8] Q.B. He, F.R. Kong, R.Q. Yan: Mechanical Systems and Signal Processing, 21 (4):1755-1772.
[9] X. F. Wang, J. Q, G. J. Liu: Journal of Vibration; Measurement & Diagnosis,2007,27(1):62-64.
[10]X. Y. Wei, H. X. Pan, F. J. Wang: Journal of Vibration; Measurement & Diagnosis,2009,29(2):162-166.
[11]V. H. Nguyen, J. C. Golinval: Engineering Structures, 32(11) : 3683-3691.
[12]J. M. Lee, Ch.K. Yoo, I.B. Lee: Computers & Chemical Engineering, 28(9):1837-1847.