A Novel Multi-resolution Kernel Principle Component Analysis Method

Jianjun Wu[a], Weijun Gong[b] and Zhen Shang[c]

College of Information Technology and Communication, Hexi University, Zhangye, 734000, China.

[a]hxuwjj@163.com, [b]gongweijun214@163.com, [c]shangzhen2006@126.com

Keywords: Multi-resolution, principle component analysis, kernel; machine learning

Abstract: Focusing on the traditional kernel principle component analysis (KPCA) can not provide principle component information under the condition of multi-resolution, this paper proposes a novel multi-resolution kernel principle component analysis (MKPCA) method, by combining KPCA with high dimensional multi-resolution analysis theory. MKPCA can explore sample data's characters and principle component information on the different resolution. The experiments demonstrate the feasibility of the proposed method.

1. Introduction

Since Spearman proposed the principal component analysis (PCA) method in 1904, it has been widely used due to its simplicity and efficiency. The improvement of PCA algorithm is mainly classified in two aspects: on one hand, the robustness of PCA has been investigated from different perspective, and various improved algorithms are proposed. On the other hand, in the traditional PCA algorithm, the principal component is determined only by the second-order statistics of the data. For a better description of the non-Gaussian distribution data, Karhunen et.al introduce appropriate nonlinear processing based on the known input sample distribution. And they proposed some algorithms for nonlinear PCA [1]. The nonlinear PCA algorithm can also be roughly divided into two categories. The one category starts from the distribution of samples, and hopes to find a probability model that best describes the internal structure of the data, such as independent component analysis [2]. The other category is represented by the kernel principal component analysis proposed by Scholkopf [3].

KPCA introduces a kernel function to obtain any high-order correlation of the input variables, and finds the required principal components by the inner product of the input data [4]. In recent years it has become one of the research hotspots in the field of machine learning, and has been widely used in many domains [5]. In this paper, KPCA is combined with the high-dimensional multi-resolution analysis theory in the separable case. Based on this combination, the kernel principal component analysis can be performed at the different resolutions, and a multi-resolution kernel principal component analysis method is proposed.

2. High-dimensional multi-resolution analyses in separable situations

Taking a two dimensional function as an example, let $f(x_1, x_2)$ be a two-dimensional function in the $L^2(R^2)$ space. Now, the function space is divided into two parts by the power series $[a = 2^j, j \in z]$ of the resolution a. The required segmentation meets the following conditions [6]:

(1) Gradual inclusion: $V_j^{(2)}(x_1, x_2) \supset V_{j+1}^{(2)}(x_1, x_2)$;

(2) Gradual replacement: $V_j^{(2)}(x_1, x_2) = V_{j+1}^{(2)}(x_1, x_2) \oplus W_{j+1}^{(2)}(x_1, x_2)$;

(3) Completeness: $\bigcap_{j\in Z} V_j^{(2)} = \langle 0 \rangle$ and $\bigcup_{j\in z} V_j^{(2)} = L^2(R^2)$;

(4) Two scale characteristic: if $f(x_1, x_2) \in V_j^{(2)}$, then $f\left(\frac{x_1}{2}, \frac{x_2}{2}\right) \in V_{j+1}^{(2)}$

(5) Displacement invariance: if $f(x_1, x_2) \in V_j^{(2)}$, then $f(x_1 - k_1, x_2 - k_2) \in V_j^{(2)}$, $(k_1, k_2 \in z^2)$. It supposes that the two-dimensional space $V_j^{(2)}(x_1, x_2)$ can be separable, that is, it can be decomposed into the tensor product of two one-dimensional spaces $V_j^{(1)}(x_1)$ and $V_j^{(1)}(x_2)$:

$$V_j^{(2)}(x_1, x_2) = V_j^{(1)}(x_1) \otimes V_j^{(2)}(x_2)$$

Let $\phi(x_1, x_2)$ and $\psi(x_1, x_2)$ be the integer-distributed orthogonal normalized basis of $V_0^{(2)}(x_1, x_2)$ and $W_0^{(2)}(x_1, x_2)$, then $\phi(x_1, x_2)$ must be decomposed into:

$$\phi(x_1, x_2) = \phi(x_1)\phi(x_2)$$

and $\int \phi(x_1)\phi(x_1 - k)dx_1 = \delta_{k1}$, $\int \phi(x_2)\phi(x_2 - k)dx_2 = \delta_{k2}$, where $(k_1, k_2 \in z^2)$

Similarly

$$\psi(x_1, x_2) = \psi(x_1)\psi(x_2)$$

and $\int \psi(x_1)\psi(x_1 - k)dx_1 = \delta_{k1}$, $\int \psi(x_2)\psi(x_2 - k)dx_2 = \delta_{k2}$, where $(k_1, k_2 \in z^2)$.

3. Multi-resolution Kernel Principle Component Analysis

For the KPCA, the key role is the kernel function. If we use the high-dimensional multi-resolution analysis theory to construct a new kernel and replace the existing one, we can construct a new principal component analysis method. The proposed MKPCA is based on this idea.

The kernel function can be divided into two categories, the one is the translation invariant kernel and the other is the dot product kernel. Zhang Li used wavelet to construct a kernel function [7]. However, the kernel of Zhang is a kind of translation invariant kernel, and its scaling function has an explicit expression. In this paper, the wavelet is used to construct a new dot product kernel. Considering the scale function does not have an explicit expression, the numerical simulation is more difficult than the translation invariant kernel. The construction steps of new kernel are as follows:

Firstly, we construct a one-dimensional kernel based on the Mercer theorem:

Mercer theorem [8]: Let the symmetric and real-valued functions K make the integral operator $(T_k f)(x) = \int K(x, y) f(y) dy$ positive. And let $\lambda_k > 0$, $\varphi_k(x)$ be the eigenvalues of the integral operator $(T_k f) f(x)$ and the normalized Eigen functions (it satisfies $\int K(x, y)\varphi_k(y)dy = \lambda_k \varphi_k(x)$). The following conclusions are established:

(1) $\sup_k \|\varphi_k\| < \infty$;

(2) $\varphi : x \to \left(\sqrt{\lambda_k}\varphi_k(x)\right)_k$;

(3) $K(x, y) = \sum_k \lambda_k \varphi_k(x)\varphi_k(y)$;

Following that, we construct a high-dimensional wavelet kernel based on a one-dimensional kernel. If an n-dimensional space $F(x^1, x^2, \cdots x^n)$ can be decomposed into tensor products of n one-dimensional spaces, that is, $F(x^1, x^2, \cdots x^n) = F(x^1) \otimes F(x^2) \cdots F(x^n)$, and the kernel defined on the i-th one-dimensional space $F(x^i)$ is $K(x^i, y^i)$. The multi-dimensional kernel defined in the n -dimensional space can be represented by

$$K(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{n} K(x^i, y^i), \ \mathbf{x} = (x^1, x^2, \cdots x^n), \ \mathbf{y} = (y^1, y^2, \cdots y^n).$$

4. Numerical experiments

A toy data set is shown in Fig. 1. The data set consists of three clusters of data, each of which is subject to a Gaussian distribution. In the next experiment, it is analyzed by the KPCA and MKPCA respectively. The analysis results are shown in Fig. 1 and Fig. 2. The red dots in the figure represent data points, and the blue lines represent principle component contours. Comparing Fig. 1 to Fig. 2, it is seen that MKPCA can provide principal component information at different resolutions.

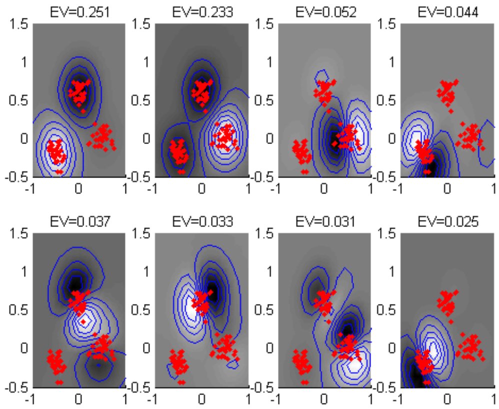
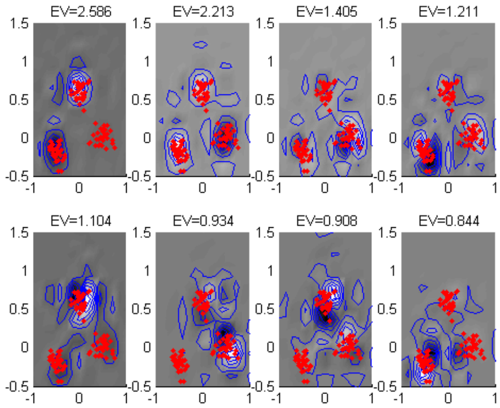
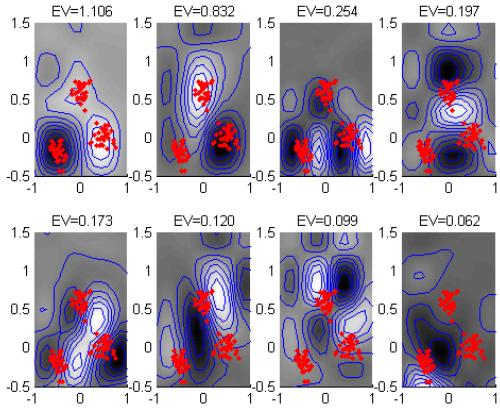


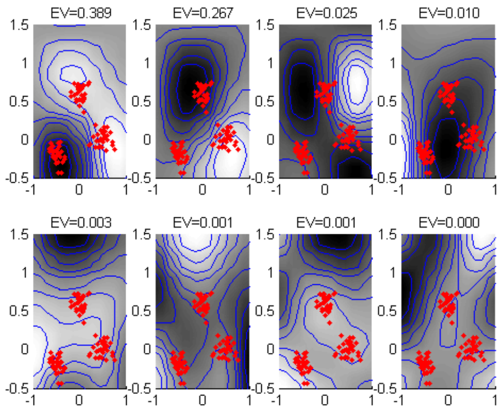
Fig. 1 Traditional kernel principal component analysis (KPCA) results



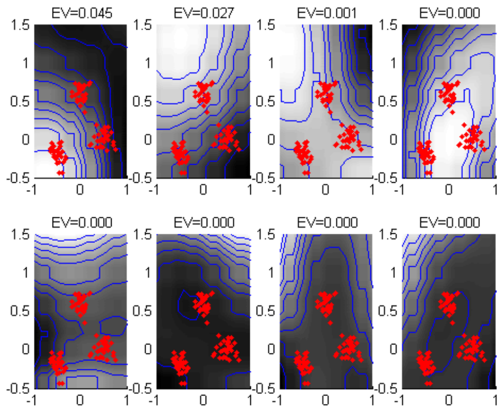
(a)Resolution j=-2



(b)Resolution j=-1



(c)Resolution j=0



(d)Resolution j=1

Fig. 2 Multi-resolution kernel principle component analysis (MKPCA) results

5. Conclusion

Traditional KPCA can not provide principal component information in multi-resolution situations. This paper proposes a novel MKPCA that solves this problem. The experimental results verify the feasibility of the proposed method, and output the principal component information under the conditions of resolution equal to -2, -1, 0, 1, etc. The next step includes using MKPCA for feature extraction in machine learning systems, etc.

References

[1] J. Karhunen, and P. Pajunen. Blind source separation and tracking using nonlinear PCA criterion: a least-squares approach. International Conference on Neural Networks (ICNN), 1997, Vol.4, 2147-2152.

[2] Daniel G. Silva, and Romis Attux. Simulated annealing for independent component analysis over galois fields. IEEE Signal Processing Letters, 2018, 25 (4): 516-520.

[3] B. Schölkopf, et al. Nonlinear component analysis as a kernel eigenvalue problem, Neural Computation, 1998, 10: 1299-1317.

[4] K. R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf. An introduction to kernel-based learning algorithms, IEEE Trans Neural Networks, 2001, 12 (2): 181-201.

[5] A. Ruiz and P. E. L. Teruel. Nonlinear kernel-based statistical pattern analysis. IEEE Trans Neural Networks, 2001, 12 (1): 16-32.

[6] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. PAMI, 1989, 11: 674-693.

[7] Zhang Li. Scaling Kernel Function Support Vector Machines. ACTA Electronica SINICA, 2002, Vol.30 No.4 April.

[8] J. Mercer. Functions of positive and negative type and their connection with the theory of integral equations. Philos Trans Roy Soc London. 1909, A209: 415-446.