ADAPTIVE ROBUST KERNEL PCA ALGORITHM

Congde Lu, Taiyi Zhang, Ruonan Zhang, Chunmei Zhang

Dep. of Information and Communication Eng., Xi'an Jiaotong University, Xi'an 710049, P. R. China
{congdelu, ruonan}@mailst.xjtu.edu.cn {tyzhang, cmzhang}@mail.xjtu.edu.cn

ABSTRACT

A novel algorithm, robust kernel principal component analysis (robust KPCA) is proposed in this paper based on the research of KPCA algorithm and its robustness. This algorithm generalizes the minimum error criteria of signal reconstruction to feature space, which can automatically recognize the outliers in the training sample set, and exterminates their effects to the accuracy of the KPCA algorithm via iterative computing. The robust KPCA algorithm not only remains non-linearity property of KPCA but gets better robustness and improves the accuracy of KPCA. The simulation experiments show that the robust KPCA algorithm developed is better than the KPCA algorithm.

1. INTRODUCTION

Principal component analysis (PCA) algorithm is a multivariable statistical analysis technique of data compression and feature extraction. It has been widely used in statistical analysis, pattern recognition and image processing. However, There exist serious problems in robustness of traditional PCA algorithm based on eigenvalue decomposition and PCA only describes input, which subordinate to normal distribution.

Many different improved algorithms of PCA have been introduced, and the research of these improved algorithms of PCA are focused on two aspects. One is how to reduce the calculation error and poor convergence of the algorithm for the outliers in the training set [1][2]. Another is how to reach the independence between the output principal components when the input do not subordinate to normal distribution. Usually, the irrelevance between the principal components can be acquired easily in the case two-order characteristics are considered. We know that the independence is equal to irrelevance only when the input subordinate to normal distribution. The problems of how to select the nonlinear function and how to combine it to the PCA have already been studied by some researchers [4][5].

Although the KPCA algorithm is excellent, it is instable when input samples include outliers, because there is a lack of considering how to recognize the outliers in the training sample set, and how to exterminate their effects to the accuracy of the KPCA algorithm. Robust KPCA algorithm is proposed in this paper based on the analysis of KPCA algorithm and its robustness.

In the next section, we will fist review the kernel PCA algorithm [4]. In section 3, we discuss the principle of the robust KPCA algorithm in particular. The experiments results on robust KPCA algorithm are given in section 4, and we have a conclusion in section 5.

2. THE PRINCIPLE OF KPCA

We map the input x to a high dimensional feature space F via a nonlinear function: $\phi : R^N \to F$, $x \to X$. Assume the mapped data is centered, i.e. $\sum_{k=1}^{M} \phi(x_k) = 0$ in the feature space F , then the covariance matrix in F is: $C^F = \frac{1}{M}\sum_{i=1}^{M} \phi(x_i)\phi(x_j)^T$. We now have to find eigenvalues $\lambda^F \geq 0$ and eigenvectors $W^F \in F \setminus \{0\}$ satisfying $C^F \lambda^F = \lambda^F W^F$. All solution W^F with $\lambda^F \neq 0$ lies in the span of $\phi(x_1), \cdots, \phi(x_M)$. There exist coefficients $\alpha_i (i = 1, \cdots, M)$ such that $W^F = \sum_{i=1}^{M} \alpha_i \phi(x_i)$. Thus we can obtain:

$$\lambda^F \sum_{i=1}^{M} \alpha_i (\phi(x_k) \cdot \phi(x_i))$$
$$= \frac{1}{M}\sum_{i=1}^{M} \alpha_i (\phi(x_k) \cdot \sum_{j=1}^{M} \phi(x_j))(\phi(x_j) \cdot \phi(x_i))$$
$$k = 1, \cdots, M \tag{1}$$

Defining a $M \times M$ matrix K by

$$K_{ij} = (\phi(x_i) \cdot \phi(x_j)) . \tag{2}$$

So the equation (1) can be rewritten as $M\lambda^F K\alpha = K^2\alpha$. This is equivalent to

$$M\lambda^F \alpha = K\alpha \quad (3)$$

Let $\lambda_1^F \geq \lambda_2^F \geq \cdots \geq \lambda_M^F$ denote the eigenvalues, and $\alpha_1, \alpha_2, \cdots, \alpha_M$ the corresponding complete set of eigenvectors of the equation (3). We need to normalize the eigenvectors in the feature space F. Suppose $\lambda_k^F \neq 0, k = 1,2,\cdots,l$, by virtue of (1), (2) and (3), $(W_k^F \cdot W_k^F) = 1$ is equivalent to

$$(W_k^F \cdot W_k^F) = \sum_{i,j=1}^{M} \alpha_i^k \alpha_j^k (\phi(x_i) \cdot \phi(x_j))$$

$$= \sum_{i,j=1}^{M} \alpha_i^k \alpha_j^k K(x_i, x_j) = \lambda_k^F (\alpha^k \cdot \alpha^k) = 1$$

$$k = 1, \cdots, l \quad (4)$$

For the purpose of principal component extraction, we need to compute projections onto the eigenvectors W_k^F in F. Let x be a test point whose projection is $\phi(x)$ in F, then the projection of $\phi(x)$ onto the eigenvectors W_k^F are the nonlinear principal components corresponding to

$$(W_k^F \cdot \phi(x)) = \sum_{i=1}^{M} \alpha_i^k (\phi(x_i) \cdot \phi(x)) \quad (5)$$

3. ROBUST KPCA ALGORITHM

The KPCA is a of the most excellent nonlinear PCA algorithm. But KPCA dose not consider the situation of the outliers in the input data, so the KPCA does not eliminate outliers when used in principle component analysis in feature space. It is relatively easy to eliminate outliers in input space, and many scholars have proposed several algorithms [1][2], but eliminating outliers in feature space is much more difficult, because we can not obtain the explicit form of the non-linear mapping function ϕ. In order to eliminate the effect of outliers to the algorithm, we must consider two problems: First, whether or not the outliers in input space remain outliers in feature space? Second, how to eliminate outliers in feature space?

As to the first problem, literature [5] has replied that if the nonlinear mapping is smooth and continuous then the topographic ordering of the data in input space will be preserved in feature space. To solve the second problem is actually to establish the criterion for recognizing outliers in feature space. The following first introduces the criterion based on minimum error criterion of signal reconstruction in input space, then generalizes the criterion to feature space, and at last, establishes the robust KPCA on the basis of the criterion.

3.1 Minimum error criteria of signal reconstruction in input space

General principle of minimum error criteria of signal reconstruction is: Set $y = W^T x$ as the principle components obtained from the input n-dimension random vector x, and $u = Wy$ as the signal reconstruction of input x, then $e = x - u$ is the reconstruction error. Define the error function $J(W)$ as

$$J(W) = E\|e\|^2 = E\|x - u\|^2 \quad (6)$$

As to training sample set, the estimation of the error function is

$$J(W) = \frac{1}{N} \sum_{i=1}^{N} \left\| x_i - WW^T x_i \right\|^2 \quad (7)$$

where the column vectors of W is unit vector and linear irrelative to each other. So the goal of optimizing error function is to at most reduce the loss to signal in the dimension reduction. It has been proved that the W, which the minimal $J(W)$ corresponds to, equal to m-dimension PCA subspace of input random vector x. That is, the subspace made up of column vectors of W equals to that made up of the principle components of x [1]. Therefore, the reconstruction error criteria can be utilized to recognize outliers.

3.2 Minimum error criteria of signal reconstruction in feature space

In the input space, set the W made up of the first m principle components of input random vectors x, and $\varepsilon > 0$ as a given threshold. An input variable x_i here is recognized as outliers, if

$$e(x_i) = \left\| x_i - WW^T x_i \right\|^2 > \varepsilon \quad (8)$$

This is the criterion for recognizing outliers in input space. The criterion is generalized to the feature space as following.

Map the input space to feature space with a non-linear function ϕ. then the signal reconstruction error in feature space is presented by

$$e(\phi(x_i)) = \left\| \phi(x_i) - WW^T \phi(x_i) \right\|^2 \quad (9)$$

However, for the explicit form of the non-linear function ϕ is unknown, we can't calculate the signal reconstruction error directly. In order to compute the error of signal reconstruction, we rewrite (9) in terms of kernel functions $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$. In fact:

$$e(\phi(x_i)) = \left\| \phi(x_i) - WW^T \phi(x_j) \right\|^2$$

$$= \phi(x_i)\cdot\phi(x_i) - 2WW^T\phi(x_i)\cdot\phi(x_i)$$
$$+ (WW^T)(WW^T)\phi(x_i)\cdot\phi(x_i)$$
$$= K(x_i, x_i) - 2WW^T K(x_i, x_i)$$
$$+ (WW^T)^2 K(x_i, x_i) \quad (10)$$

Here gets the signal reconstruction error in feature space via computing (10). On the basis of the formula, the new robust KPCA algorithm is proposed. Although the explicit form of non-linear function ϕ is not presented, we can compute out directly the dot products in feature space via the kernel function, e.g. polynomial kernel.

3.3 Robust KPCA algorithm

Utilizing the above criterion, we can only give out the estimation of W, rather than the precise value of W. Then we utilize the estimation of W to recognize and eliminate the outliers, and get the more accurate estimation of W, and repeat the procedure above. At last, here gets the relatively accurate W via such iterative computing. Additionally, compute the reconstruction error of the whole training set, and set the samples with relatively more reconstruction error as outliers.

In summary, for a given training sample set $X = \{x_1, x_2, \cdots, x_M\}$, the robust KPCA algorithm proposed is described as following:

Step1: Initialize the number of steps of the iterative $k = 0$, and set the sample set to be processed $Y = X$, that is the number of outlier samples $O(k) = 0$;

Step2: Center the training sample set in the feature space;

Step3: Analyze the sample set with kernel PCA, and get the estimation matrix $W(k)$;

Step4: Normalize the estimation matrix $W(k)$ in the feature space;

Step5: According to $W(k)$, utilize (10) to calculate reconstruction error of the training samples, that is $e(\phi(x_i)) = K(x_i, x_i) - 2W(k)W^T(k)K(x_i, x_i) + (W(k)W^T(k))^2 K(x_i, x_i), i = 1,2,\cdots,M$;

Step6: Set the step number of iterative $k = k+1$, eliminate the sample of the most reconstruction error from the sample set X in the last step, and set the number of outlier samples as $O(k+1) = O(k)+1$. Recompose the new sample set Y to be processed with the remained samples;

Step7: If $W(k+1)$ satisfies convergence condition, the iterative computing ends; otherwise, jump to the step3.

The robust KPCA above mainly consider two problems: first, which rule is selected to recognize the outliers? It is difficult to ascertain ε, If utilize threshold ε to recognize the outliers. Therefore, we predefine the number of the outliers in the training sample set in the experiment. In practice, although the number of outliers is unknown, they always occupy only a little percentage in the training sample set, such as 2%-5%. Setting the number of outliers a little more than that of practice is generally acceptable. Because the number of non-outlier samples is much more, they will not affect much the accuracy of final result even though some of non-outlier samples are recognized as outliers and are ignored. Second, How to select kernel function? Experiments show that different kinds of kernel functions represent approximately the same performance [6]. Presently, the polynomial kernels, radial basis function kernels and sigmoid kernels are all widely used.

4. SIMULATION EXPERIMENTS

The simulation experiments in this paper respectively adopt KPCA algorithm and the robust KPCA algorithm to analyze the input sample set. And two sample sets are involved in the experiment: one has outliers and the other dose not. The contrast results will evaluate robustness of the KPCA and the robust KPCA. In order to get more contrastive results, this experiment adopts the artificial data in literature [4] and polynomial kernel.

Generate randomly 300 normally distributed 2-dimesion sample set, which is composed of 3 clusters, indicated as Λ (Fig.1); Insert randomly 3% outliers whose distribution is quit different from the normal distribution into the sample set Λ, indicated as sample set Ξ (Fig.2).

First apply KPCA to sample set Λ, Ξ. From Fig.1, KPCA algorithm has very high accuracy on sample set without outliers, and the portrayed contour reflects the data structure quit well. The first two principle components succeed in separating the 3 clusters, and the following 3 principle components again separate the 3 clusters with more detail. At the same time, the experiment illustrates that KPCA can provide abundant principle component contrast to PCA. Fig.2 shows that outliers have great effect on the accuracy of principle components in KPCA. Therefore, for the data with outliers, KPCA is instable. Also from Fig, 2, only 3% outlier samples result in the problems that the first two components can not separate the three clusters well, and the effect to the following three components is more serious.

For the sample set without outliers, the robust KPCA can obtain quit accurate principle components too. For the sample set Ξ, utilize the improved KPCA algorithm as Fig.3. This illustrates that for the first two components, the improved KPCA algorithm has already eliminate the effect of noise. Therefore, the robust KPCA can obtain the approximately precise principle components. In the following three components, the effect of noise is also

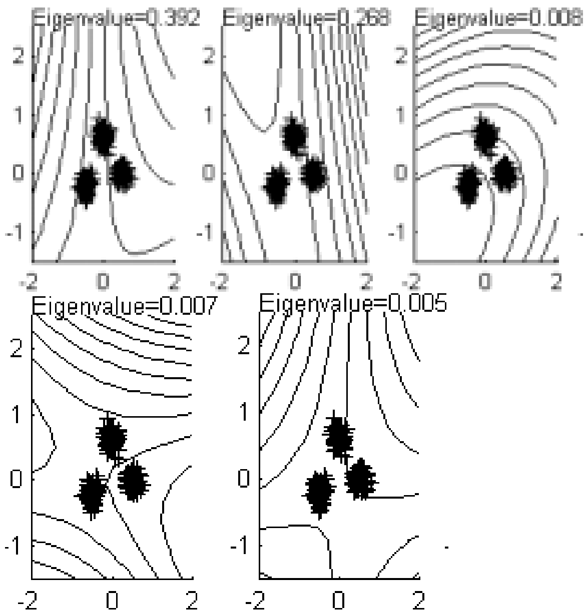

Fig. 1 analysis results of non-outliers sample set with the KPCA algorithm

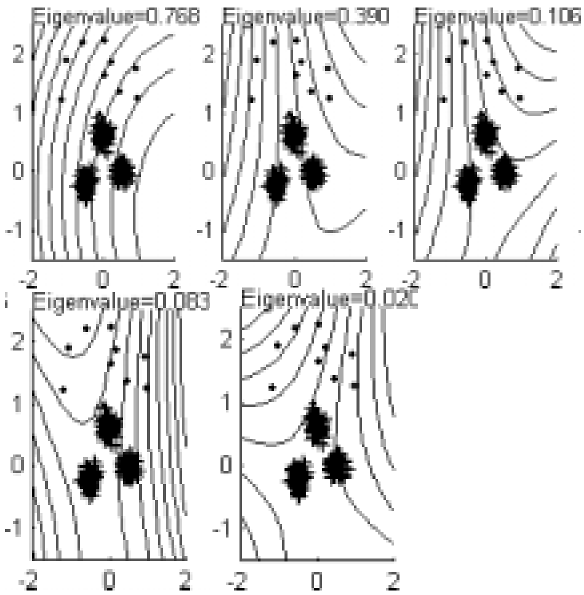

Fig. 2 analysis results of sample set including outliers with the KPCA algorithm

slight. In summary, the improved KPCA algorithm has better robustness contrast to KPCA. In addition, the simulation experiment also demonstrates that the robust KPCA algorithm succeeds in finding out the outliers that affect the accuracy severely, which has important significance for signal and data processing.

5. CONCLUSION

This paper, based on the analysis of KPCA and its robustness, generalizes minimum error principle of signal reconstruction to the feature space and proposes the adaptive robust KPCA algorithm. This algorithm remains non-linear property of KPCA and possesses better robustness. The simulation experiments illustrate that the improved KPCA algorithm, contrast to the KPCA algorithm, enhances obviously the robustness and gives out the approximately accurate principal components. This algorithm may be applied in data compression, feature extraction and image processing.

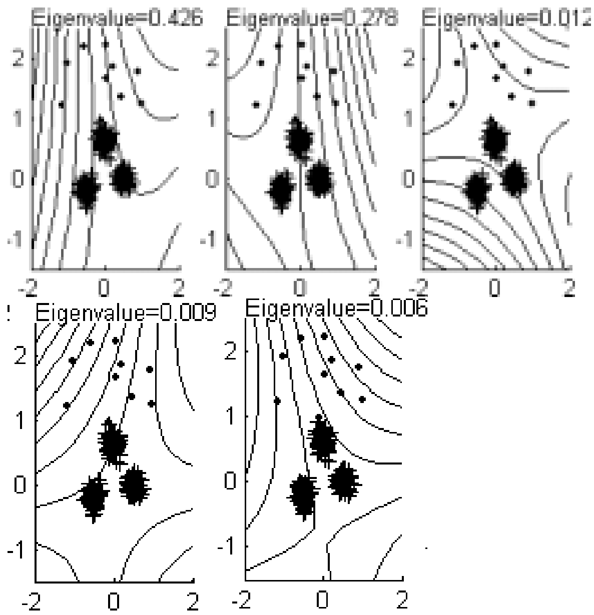

Fig. 3 analysis results of sample set including outliers with the robust KPCA algorithm

6. REFERENCES

[1] S. Wang, Y.L. Liang etc. "Robust PCA Based on Neural Networks", Proceedings of the 36th IEEE Conference on Decision and Control, Vol. 1, pp.503-508, 1997.
[2] L. Xu and Alan L. Yuille, "Robust principal component by Self-organizing Rules Based on Statistical Physics Approach", IEEE Trans. Neural Networks, Vol. 6, No. 1, pp.131-143, 1995.
[3] J. Karhunen, J. Joutsensalo, "Generalizations of principal component analysis, optimization problems, and neural networks", Neural Networks, Vol. 8, No. 4, pp.549-562, 1995.
[4] B. Scholkopf, A. Smola and K. R. Muller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem", Neural Computing., Vol. 10, No. 5, pp.1299-1319, 1998.
[5] M. Girolami. "Mercer Kernel Based Clustering in Feature Space", IEEE Trans. Neural Networks, Vol. 13, No. 5, pp. 780 –784, 2002.
[6] V. N. Vapnik, "The Nature of Statistical Learning Theory", Springer-Verlag, 1995.