The determination of kernel function parameter initial value of the KPCA based on adaptive contribution rate

Xinghuo Wan, Meilin Zhang, Yili Tan

College of Science, North China University of Science and technology, Tangshan 063210, China

Abstract: In this article, we start from the samples and the kernel principal component contribution rate adaptive point of view. Our aim is to research that how to choose the initial parameters of the mixed kernel function in comprehensive evaluation of KPCA. To some extent, it will overcome the shortcomings that the kernel function parameter is difficult to determine, and it provides an effective method that KPCA applies to determine the kernel function initial parameter of the multi-indexes comprehensive evaluation. We believe that it will be more widely applied to feature extraction with its advantages appear in specific issues, and makes the evaluation method more and more scientific and effective.

1. Introduction

Evaluation and summary evaluation is the extremely important activity of understanding human society, almost all activities can make comprehensive evaluation. With the increasing of people's awareness levels, the evaluation objects that we see become more and more complex. Due to the factors that affect the evaluation tend to be numerous and complex, it is unreasonable that we use a single indicator to evaluate the objects, so we can get a comprehensive index through collecting the multi-information of the objects. We can reflect the overall situation of objects. This is the multi-indexes comprehensive evaluation method.

In recent years, there have been appeared many methods that using modern methods to research multi-indexes comprehensive evaluation method[1-3]. There have some new discipline methods, such as the fuzzy mathematics[6], the artificial neural networks[8], the gray system theory[34] and the kernel principal component analysis(KPCA)[5] and so on. They are also be introduced to the comprehensive evaluation in the past.

Kernel principal component analysis is the effective method which combine the kernel function and principal component analysis, its aim is to solve nonlinear problems and provide more information. It lets the original variable space X mapped into a high dimensional feature space F by a nonlinear transformation φ, we use linear principal component analysis in F by nuclear techniques. We do not know the specific form φ, but we can calculate just in the original space. We can compare with the principal component analysis, the evaluation method has a widely application and the dimension reduction effect is significant. Finally, it has practical value.

Specific evaluation steps of KPCA method is as follows:

1) We use the standardized approach to initialize the input samples, and then use the mapping of the kernel function to obtain a matrix $K$.

2) We want to solve the matrix $\tilde{K}$, in which $\tilde{K} = K - I_iK - KL_i + I_iKL_i$, $(I_i)_{ij} = \frac{1}{i}$. 
3) We can find the matrix of eigenvalues \( \lambda_i (i = 1, 2, \cdots, I) \) and eigenvectors \( v_i (i = 1, 2, \cdots, I) \).

4) We can identify the eigenvalues \( \lambda_r \) and eigenvectors \( v_r (r = 1, 2, 3, \cdots, m) \) corresponding with the \( m \) main elements.

5) We can find the contribution rate and cumulative contribution rate of the eigenvalues.

6) Finally, we can use the extracted kernel principal component analysis to make comprehensive evaluation.

Choosing kernel function plays a key role on the effect of comprehensive evaluation. Each kernel function has its own advantages and disadvantages, and different kernel functions exhibit the different characteristics. The kernel function can be divided into two categories now: global and local kernel function. Global kernel function has global feature that allows the distant data points have an influence on the value of the kernel function. But the local kernel function has locality that allows the close data points have an influence on the value of the kernel function. Local kernel function has strong learning ability, but the generalization performance is weak. On the other hand, the generalization performance of the global kernel function is strong, and learning capability is weak. Therefore, we can combine the two kinds of kernel function, and we can use the respective advantages of each function adequately. So the mixed kernel function not only has good learning ability, but also has good generalization ability. The result of the comprehensive evaluation is more significant when we use the mixed kernel function on the KPCA.

However, there is no proven theory to guide which kernel function we can select and how to determine the initial kernel of the kernel function parameters of the mixed kernel function, so we must make several tests to finalize the kernel functions in practical applications. And how to select the initial kernel parameters is very important in this process. Because it directly affects the number and the test results of the comprehensive evaluation. The time of the kernel principal component analysis which provided is not long, and few people study this. In this paper, we start from the samples and the kernel principal component contribution rate adaptive point of view. Our aim is to research that how to choose the initial parameters of the mixed kernel function in comprehensive evaluation. I believe that with its advantages embodied in specific issues this method will be more widely applied in feature extraction.

2. Determination of the kernel function initial value based on samples and kernel principal component contribution rate adaptive

2.1 Analyze the original data

When we collect the original data, we don’t use all multivariate data to make kernel principal component analysis and comprehensive evaluation. The basic idea of kernel principal component is let the combination of the original value which has a certain correlation into a new variable, we should calculate the correlation coefficient between the variables. If the correlation coefficient is small, we will don’t use the kernel principal component analysis. When the absolute value of the correlation coefficient is large, we can use this method to analyze, and then we can make a comprehensive evaluation.

Then we analyze the original data of the 2007 year energy consumption (we choose data from China Energy Statistical Yearbook 2007), the original data is normalized as follows:
We can find from the figure that the absolute value of the ten variable correlation coefficient is greater than 0.5, so we can use kernel principal component analysis. We use the Matlab program to calculate and analyze the 2007 year energy consumption.

2.2 Determination of the Gaussian kernel function initial parameters

We need to give a fixed initial value to the parameter of the kernel function when we make comprehensive evaluation of KPCA. Then we can discuss the conclusion of the comprehensive evaluation. We should to determine the range of the kernel function parameters through the above data and the aim is the first principal component. First we need to determine the $\sigma^2$ of the Gaussian radial basis kernel function.

$$K(x,x_i) = \exp\left(-\frac{|x-x_i|^2}{\sigma^2}\right)$$

| $\sigma^2$  | The first principal component |
|------------|-----------------------------|
| 0.0001     | 0.1141                      |
| 0.001      | 0.1770                      |
| 0.01       | 0.2230                      |
| 0.1        | 0.3377                      |
| 1          | 0.6875                      |
| 10         | 0.8031                      |
| 100        | 0.8165                      |
| 1000       | 0.8179                      |
| 10000      | 0.8180                      |

We can know from the above table that the contribution rate of the first principal component tends to balance when $\sigma^2 > 10000$. So the kernel matrix of the Gaussian kernel function is a unit matrix. The Gaussian kernel loses its original role in a mixed kernel function, so we can choose $\sigma^2 = 100$.

2.3 Determination of the polynomial kernel function initial parameters

Polynomial kernel function: $K(x,x_i) = [c(x \cdot x_i) + m]^\beta$

In the polynomial kernel function, when the $m = 1$, the changing result of $\beta$ is as follows.
### Table 2: The values of $\beta$

| Parameter | $m$ | $\beta$ | The first principal component |
|-----------|-----|---------|-------------------------------|
| 1.0       | 1   | 0.8180  |
| 1.0       | 2   | 0.7445  |
| 1.0       | 3   | 0.7103  |
| 1.0       | 4   | 0.7095  |
| 1.0       | 5   | 0.7284  |
| 1.0       | 6   | 0.7548  |
| 1.0       | 7   | 0.7818  |
| 1.0       | 8   | 0.8066  |
| 1.0       | 9   | 0.8283  |
| 1.0       | 10  | 0.8466  |

As can be seen from the table, the contribution rate of the first principal has a good effect when $\beta = 10$. But when the value of $\beta$ is too large, the generalization ability of the kernel function is too large and it will lose their role. And after reaching a certain extent, the kernel function matrix will not exist. So in order to ensure the kernel matrix exists the largest characteristic root, the value of $\beta$ is ten.

When the value of $\beta$ is ten, the value of $m$ is as follows.

### Table 3: The values of $m$

| $\beta$ | $m$ | The first principal component |
|---------|-----|-------------------------------|
| 10      | 1   | 0.8466                        |
| 10      | 2   | 0.8225                        |
| 10      | 3   | 0.8061                        |
| 10      | 4   | 0.7887                        |
| 10      | 5   | 0.7734                        |
| 10      | 6   | 0.7601                        |
| 10      | 7   | 0.7486                        |
| 10      | 8   | 0.7390                        |
| 10      | 9   | 0.7309                        |
| 10      | 10  | 0.7241                        |

When $m > 1$, the first principal component will be decrease during the change of $\lambda$.

As we can see from the above table, the polynomial kernel function become linear kernel function when $m = 0$, so $m \neq 0$. With the increasing of $m$, the first principal component shows a decreasing trend. So at that time, the first principal component contribution rate is the best when $m = 1$.

After all of the parameters are determined, we can obtain two kernel functions which parameter values are fixed:

$$K(x, x_i) = [(x, x_i) + 1]^{10}$$

$$K(x, x_i) = \exp\left(-\frac{|x - x_i|^2}{100}\right)$$
3. Conclusion
Kernel function method has solid theoretical foundation and has been successfully applied in many fields, so a large number of researchers focus on this method. Kernel function need to get rid of the trouble of dimension through quoting the inner product kernel function and solving the dual problem, and this method will solve the dual problem effectively. We will face with the problem that a larger amount of memory capacity and computing too long when the size of the training set is so large. How to determine the form of kernel function and select the parameter is the difficult problem to solve.

Although the researchers do a lot of work on the choice of kernel function parameters, the selection of kernel function parameters still lack a unified approach. The selection of kernel function parameters play a decisive role on how to improve the performance of the kernel function, so it is necessary to study it. Particularly, it is a good method which based on principal component analysis that can be the tool which can solve the nonlinear problems and it has been widely used in various fields. The problem of the selection of kernel function parameters are rarely involved. In this article, we infer the method of the determination of the kernel function initial value based on the sample and the contribution rate adaptive of kernel principal component, to some extent, this method overcome the difficult to determine the kernel function parameters. And this method provides an effective evaluation methods to the multi-index comprehensive evaluation method. And this method will make the evaluation methods more scientific and effective.

References
[1] Genting Yan. Hybrid kernel function SVM [J] Journal of Harbin Institute of Technology, 2007,39 (11): 1704-1706.
[2] Jolliffe I T. Principal component Analysis (2nd) [M] New York:. Springer, 2002.
[3] Yonghong Hu, Sihui He. Comprehensive Evaluation Method [M]. Beijing Science Press, 2000.
[4] Smitsgf, Jordanen. Improved SVM regression using mixtures of kernels [A] IEEE Proceeding of IJCNN pon Neural Networks [C] 2002: 2785-2790.
[5] Fuqin Zhu. Comprehensive evaluation of the level of urbanization based on the KPCA [J] decision making, 2004 (1): 40-41.
[6] Aili Zhang, Guangli Liu, Qingshui Liu. Scientific journals comprehensive evaluation model -KPCA [J] Computer Engineering and Applications, 2003 (24): 200-201.
[7] Muller KR, Mika S, Ratsch S, et al An introduction to kernel -based learning algorithms [J] IEEE Transactions on Neural Network, 2001, 12 (2): 181.
[8] Xuemin Wang. Query on the comprehensive scoring method of principal components analysis [J]. Statistics and Decision, 2007,23 (8): 31-32.