MULTI-SPECTRAL IMAGE ANALYSIS BASED ON DYNAMICAL EVOLUTIONARY PROJECTION PURSUIT

YU Changhui [a], MENG Lingkui [a], YI Yaohua [b],

[a] School of Remote Sensing Information Engineering , Wuhan University, 39#,Luoyu Road, Wuhan,China,430079, yuchhuis@263.net

[b] Dept. of Printing and Packing, Wuhan University , 39#,Luoyu Road, Wuhan,China,430079.

WG VII /6

KEY WORDS: multi-spectral image, projection pursuit, dynamical evolutionary algorithm

ABSTRACT:

Principal component analysis (PCA) is usually used for compressing information in multivariate data sets by computing orthogonal projections that maximize the amount of data variance. PCA is effective if the multivariate data set is a vector with Gaussian distribution. But multi-spectral images data sets are not probably submitted to such Gaussian distribution. The paper proposes a method based on Projection Pursuit to find a set of projections that are "interesting", in the sense that they deviate from Gaussian distribution. Also a dynamical evolutionary algorithm was developed in order to find the optimal projection index. The effectiveness of this method is demonstrated through simulated data and multi-spectral image data.

1. INTRODUCTION

Remote sensing is an indispensable tool in many scientific disciplines. It is one of the major tools in monitoring our earth environment in a cost-effective way. As a fully new technique, multi-spectral & hyper-spectral remote sensing has special characters such as narrow bands, multi channels and integrating of image and spectral compared to traditional remote sensing. Imaging spectral remote sensing, which integrate the techniques of imaging and spectrality, can get mass multi-spectral & hyper-spectral images about earth resources and environment.

Compared to traditional remote sensing, Multi-spectral and Hyper-spectral remote sensing image could provide more information. So new algorithms and software are required to processing and extracting the information in hyper-spectral remote sensing. Principal component analysis (PCA) is usually used for compressing information in multivariate data sets by computing orthogonal projections that maximize the amount of data variance. PCA is effective if the multivariate data set is a vector with Gaussian distribution. But multi-spectral images data sets are not probably submitted to such Gaussian distribution. This paper presents an algorithm based on Projection Pursuit to find a set of projections that are "interesting", in the sense that they deviate from Gaussian distribution. In multi-spectral & hyper-spectral imagery analysis applications, the general used method to processing high-dimension data is the Exploratory Data Analysis (EDA). While as a new method for analysis multivariate data projection pursuit is attracting more and more attentions.

2. PROJECTION PURSUIT

2.1 Title

Projection Pursuit is applied to explore the potential structures and characters of the multi-dimension data through projecting the high dimensional data set into a low dimensional data space while retaining the information of interest. Projection Pursuit was first presented by Kruskal in 70th of 20 century. He found the clustering structure and settled the classification problem of fossil through projection high dimension data into low dimension space. Then Friedman and Tukey presented a new projection index which combined the wholly clustering structures and local clustering to carry out the classification problems. They coined the concept of projection pursuit formally. The kernel idea of projection pursuit is to project a high dimension data set into a low dimension space while reminding the information interest. The fitness of the projection was evaluated by designing a projection index. So projection pursuit means to search the optimal projection index to get the maximum projection. In nature it is a problem of maximum optimization.

Projection pursuit is different from PCA in that it can help us to find the "interesting" linear projection. From the point of projection pursuit, PCA is just a special case of projection pursuit. Let the data set is $y_{m\times n}$, we can get $x = w^{T}_{m\times m} y$, in which w is the projection vector and x is the corresponding projection value. If made the maximum variance the projection index, and the columns of w are perpendicular to each other. Then the projection transform becomes the PCA transform. The core of projection pursuit is to find an appropriate projection index. This paper use information divergence as the projection index.

Given two continuous probability distributions $f(x)$ and $g(x)$, the relative entropy of $f(x)$ with respect to $g(x)$ can be defined as

$$d(f \| g) = \int_{-\infty}^{+\infty} g(x) \log \frac{f(x)}{g(x)} dx. \quad (1)$$

And their absolute information divergence can be calculated by the follow formul.

$$j(f,g) = |d(f \| g)| + |d(g \| f)| \quad (2)$$

Then we get that $j(f,g) = j(g,f)$ and $j(f,g) \geq 0$. When $f = g$ the divergence will be zero. If let $g(x)$ submits to Gaussian distribution, then we can calculate the divergence of $f(x)$ relative to Gaussian distribution. To do this, we must evaluate the $f(x)$. But the general used method Parzen Windows is very difficult. A simple and effective method is approximate the continuous quantity $f(x)$ and $g(x)$ with the discrete counterparts p and q respectively. Then their relative entropy can write as

$$D(p \| q) = \sum_i p_i \log \frac{p_i}{q_i} \quad (3)$$

in which p_i and q_i denote the ith components of the p and q separately. The absolute information divergence becomes

$$J(p,q) = D(p \| q) + D(q \| p) \quad (4)$$

The projection index can be calculated as the follows:

(1) Quantized the standard normal distribution into n - dimentional vector by selecting the step length Δx.

(2) Standardize the projection data using the formula $(x - \bar{x}) / \sigma_x$.

(3) According to Δx calculating the histogram p from the standardized data and normalize p to let $\sum_i p_i = 1$.

(4) Calculating the probability of Gaussian distribution by the formula

$$q_i = 1/\sqrt{2\pi} \int_{i\Delta x}^{(i+1)\Delta x} e^{-t^2/2} dt, \; i = -n/2 \sim n/2.$$

(5) Combine formula (3) and (4) to compute the projection index $J(p,q)$.

In this paper let $\Delta x = 1$ and the discrete range be $[-10\sigma, 10\sigma]$ to find the data points that different from normal distribution by searching the projection index which can maximum the projection vectors.

3. DYNAMICAL EVOLUTIONARY ALGORITHM

This paper proposes a dynamical evolutionary algorithm (DEA) based on the theory of dynamics evolvement system to searching the appropriate projection index. The algorithm can be implemented by the follow four main steps (1) select the code scheme, (2) decide the selection strategy, (3) define the parameter and variant to control the algorithm, (4) determine the method and the stop rule. Through the above four process, the evolutionary algorithm can found the optimal projection.

(1) Code scheme

For traditional binary process the method may be simple but if needs higher quantize precision the string length by binary coding will be extremely longer. Then the time of coding and decoding of chromosome and solution will occupy lots of computing time. Because the column vector of projection vector w is unit vector that is each component has the value between $[-1,1]$. We can apply a coding scheme that is assigned a space of double byte for each component of column vector. The value of the double byte is a real number of $[-1,1]$. So the step length of each real number is 2/65536=0.00003. Experiments show that the above precision satisfied the practical condition. Then one column can be represented by such double byte of size m. Such a string can be called one chromosome and the correspondence projection tracing value of some projection index of this vector is the adaptive value of the chromosome.

(2) Selection strategy

The problem of projection tracing can be simplified as a maximum problem:

$$\max_{w_i \in s} f(w_i)$$

$$s.t. \qquad w_i \cdot w_j = \begin{cases} 0, i \neq j \\ 1, i = j \end{cases} \; i, j = 1,2,...,m$$

let $x_1, x_2, ..., x_N$ be the N particles and they composed a dynamic system. Their function values are $f(x_1), f(x_2), ..., f(x_N)$. Different from traditional generic algorithm the iterative step t in DEA is called time.

The momentum of a particle x_i at the time t is defined as

$$p(t, x_i) = f(t, x_i) - f(t-1, x_i)$$

where $f(t, x_i)$ denotes the function value of the x_i at the time t.

The activity of particle x_i at time t is defined as

$$a(t, x_i) = a(t-1, x_i) + 1$$

If at time t the particle x_i is selected then the activity will be changed as the above formula, otherwise if keeps unchanged.

Introduction the above two quantities we define the selection operator in DEA as:

$$slct(t, x_i) = \sum_{k=0}^{t} | p(k, x_i) | + a(t, x_i),$$

Assign a weight coefficient $\lambda \in (0,1)$ according to the importance of the above quantities. Then the above formula changed to :

$$slct(t, x_i) = \lambda \sum_{k=0}^{t} | p(k, x_i) | + (1-\lambda) a(t, x_i)$$

In this algorithm, we sort the $slct(t, x_i)(i = 1,2,...,N)$ from small to large and it can denote by $slct(t)$. From these preparations, we can describe the DEA as follows.

(A) Initialize a dynamic evolutionary system，$t \leftarrow 0$

a. Random generate $\Gamma = (x_1, x_2, ..., x_N)$;

b. Calculate the function values of particles in Γ ,and set $p(t, x_i) = 0, a(t, x_i) = 0, x_i \in \Gamma$;

c. Save the best particle and its function value;

d. Calculate and sort $slct(t)$.

(B) Perform Iteration

Repeat:

(a) $t \leftarrow t + 1$;

(b) select N particles from the forefront of $slct(t-1)$;

(c) Implement cross and variation operator on N particles;

(d) Modify the function values, momenta and activities of the N particles;

(e) Save the best particle and its function value;

(f) Modify and sort $slct(t)$;

Until:

{stopping criterion}

The selection strategy of DEA is that the smaller variations of function value between two generations and the fewer selected of a particle, the more possible it is selected. This ensures every particle can be selected with enough long time. That is the selection strategy drives all particles moving just like molecules move all the time and everywhere. That is the reason why we called the algorithm the dynamical evolutionary algorithm.

(3) Select the control parameter

The selection of parameter has important effect on the convergence and convergence pace. The parameters include intercross probability p_c , differentiation probability p_m and the population size etc. Intercross is the key process in GA. The reasonable of intercross probability selection is vital to optimal result and optimal efficiency. All of the part of the particles can participate in the intercross that means let $p_c = 1.0$ or $p_c < 1.0$. The particle number is determined by the intercross probability p_c . The new generated particles which corporate parts of chromosomes from father and another part of chromosome from mather maybe inferior to their parent and will be eliminated or they maybe excellent than their parent for inherit the fine genes. In any case the new particle will be fall into the possible solution areas. Generally speaking the intercross probability p_c should select bigger. Data shows the value should be 0.6~1.0. In this paper the intercross probability p_c is 0.65.

Differentiation probability p_m is also an important parameter. If the p_m is too small then the differentiate ability of the unit will not enough and result in all the population evaluated to a unitary unit model. This model may correspond to a local extremum and not a global extremum. However if the p_m is too large then the unit will differentiate frequently and lead to the matching improved slowly and lower convergence pace. A optimal differentiation probability p_m can get the optimization population and model and not fall into a local extremum. Data shows the appropriate Differentiation probability p_m is ranging from several in a thousand to several in a hundred. This paper we define the p_m as 0.01.

Considering the selection of population size, if the size is small we can retain the diversity of genes and if it is too large the time of iterative compute will increase and cannot get the good searching efficiency. Moreover the population size is respect to the problem space. Volumes of data tell us if the size be controlled in 40~120 then the result will good. So this paper the size is 80.

(4) Stop criterion

In the DEA we can use dynamic to get two stop criterions. The first one is when $\sum_{i=1}^{N} | p(t, x_i) | < \varepsilon$ the evolvement will over. In which ε is a given positive constant. The intuitive explanation of this criterion is that if all the particles cannot be improved the evolvement will be stopped. From the point of statistics when all of the molecule exhaust its energy then the system will attain a low energy state. In fact the left of the formula equals to the energy norm defined at statistics.

The second stop criterion is when $\max_{x_i \in \Gamma} \sum_{k=0}^{t} | p(k, x_i) | > M$ the process stopped. Where M is a given large positive constant. In here M is only on theory and in practice the M is hard to found. The first Stop criterion was used in this paper。

4. EXPERIMENT AND CONCLUSION

4.1 Simulated Data Experiments

To demonstrate our approach we constructed a non-Gaussian data set. Generating 1000 ten-dimensional sample which has zero-mean. The variance of the ten-dimensional feature are geometrically decreased with the value is 1, 1/2, 1/4 ,1/8 ,…,1/512. In which 1％(ten)of the sample were regenerate by Gaussian random generator in the third and fourth features. Because there are little relative variance exists in the data features. The PCA cannot find the unnormal 10 points. Whereas using the method of divergence index we can detect the unnormal points. Figure 1 is the result of PCA and figure 2 denotes the result of divergence index.

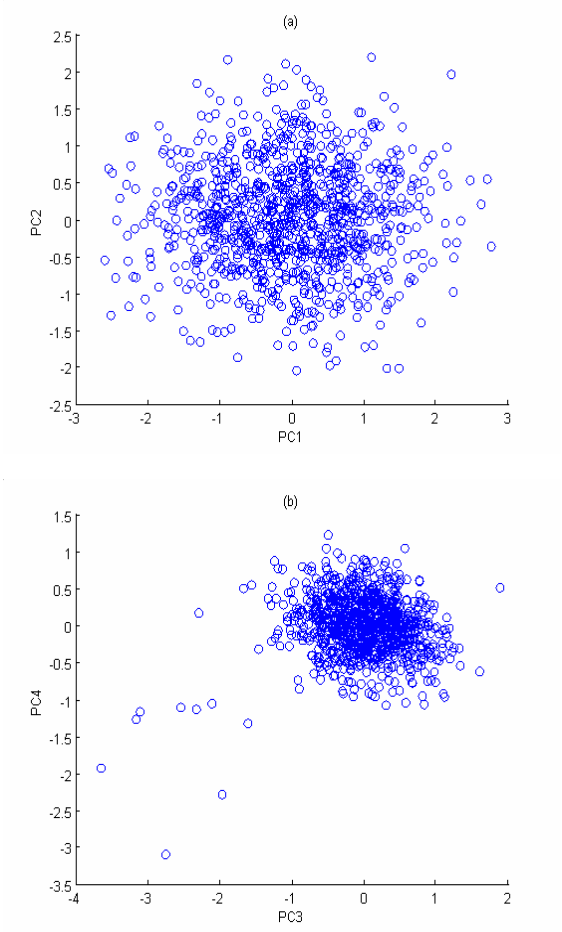
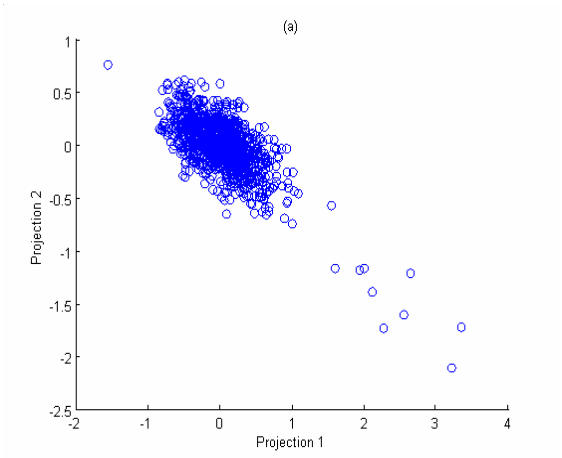


Figure 1. Result of Principal Component Analysis



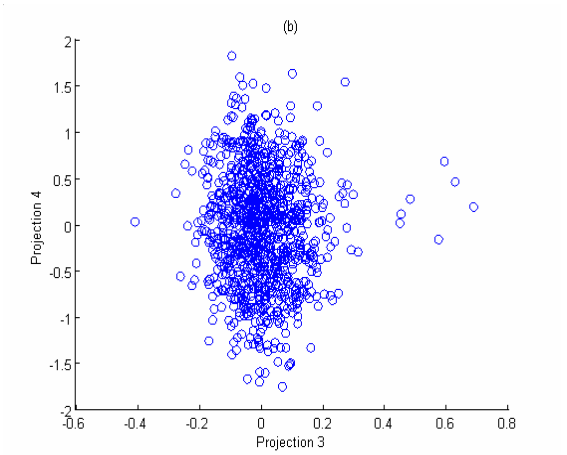


Figure2. Result of Dynamical Evolutionary Projection Pursuit

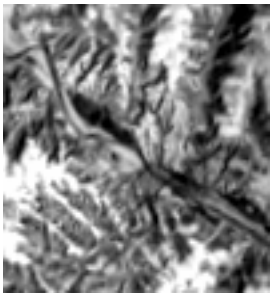
Figure1 is scatter point of the PCA to simulate data. (a) the first and the second component. (b) the third and the fourth component. Figure 2 is the scatter point of divergence index to simulate data. (a) the first and the second projection. (b)the third and the fourth projection. The first and the second component in figure 1 show the non-Gaussian distribution of data. Just as we expected the third and the fourth component displayed the information of the abnormal points. Figure 2 shows the method of projection pursuit. The distribution of the abnormal points of the first and the second component can be found exactly by using convergence index as the projection index.

4.2 Multi-spectral Image Data Experiments

The 7-band ETM+ data used in the experiments is imaged in the June twenty-ninth 2000.It is a sub-scene of 100*100 pixels extracted from HEKOU town. Only 6 bands (including visible and infrared bands) were chosen, that is band 1~5 and 7. Its space resolution is 30meter. Figure3 is the composite color image by R-3、G-4、B-2。

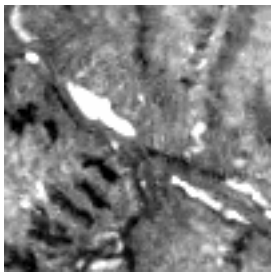
Figure3. ETM+ multi-spectral image data

Figure4 shows the first three principal component images and figure5 is the first three projection images after Dynamical Evolutionary Projection Pursuit (DEPP) by divergence index.



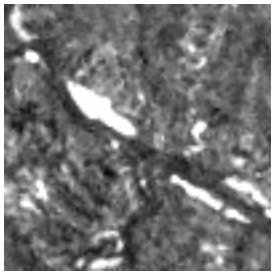
(1)

(2)

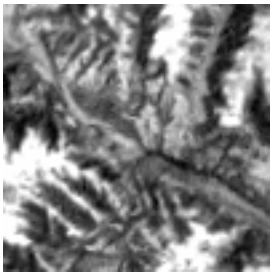


(3)

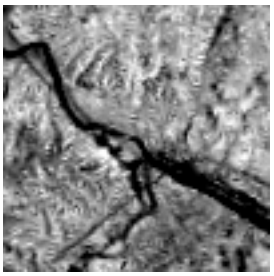
Figure4. The fore three Component images from PCA



(1)



(2)



(3)

Figure5. The fore three Projection image from DEPP

Component/Projection	Max $J(p,q)$ PCA	Max $J(p,q)$ DEPP
1	0.090679	1.359902
2	0.158729	0.762566
3	0.227217	0.213851
4	0.064051	0.104878
5	0.029244	0.085720
6	0.006726	0.079304

Table1. Max $J(p,q)$ Values for the PCA and DEPP

Table 1 is the divergence value of the each principal component image and projection image. It can be concluded that the divergence value after PCA processing is much more small than that after DEPP processing from table.1. Simulated Data and Multi-spectral Image Data Experiments show that the proposed DEPP method provides an effective means for finding the anomaly structure character from multi-dimensions dataset.

REFERENCES

Cheng Ping, Li Guoying. 1986. Projection Pursuit - a New Statistical Method. *Application of probability statistic* 2(3): pp.8-12 .

A.Ifarragaerri and C.-I.Chang 2000. Multispectral and hyperspectral image Analysis with projection pursuit. *IEEE Trans.Geosci. Remote Sensing.* vol.38 pp.2529-2538.

S.-S.Chiang and C.-I.Chang 2001. Unsupervised Target Detection in Hyperspectral Images Using Projection Pursuit, *IEEE Trans,Geosci Renote Sensing.* vol.39,pp.1380-1391.

Li YuanXiang, Zou XiuFen, Kang Lishan and Zbigniew Michalewicz. 2003. A New Dynamical Evolutionary Algorithm Based on Statistical Mechanics, *J.Comput. Sci & Technol*, vol.18 pp.361-368.