Analysis and Application of K-L Transform Information Features*

Fengxiang JIN and Shifeng DING, China P. R.

Key words: Indicating Entropy, Information Feature, Information Function, Information Quantity, K-L Transform.

SUMMARY

This paper makes a brief introduction about the Orthogonal Transform with a minimum of Mean Square Error (MSE), called K-L Transform. Then the information features of K-L transform is analyzed from the viewpoint of information theory. A new idea is presented about using Indicating Entropy to measure the information quantity. Some active results are shown following a practical example.

RESUME

Cet article donne une présentation de la Transformation Orthogonale par la méthode des moindres carrés (MSE), appelée Transformation K-L. Puis ses caractéristiques sont analysées du point de vue de la théorie de l'information. Une nouvelle idée est presentée par l'entropie de l'indication pour mesurer la quantité d'information. Quelques resultats actifs sont donnés par des exemples réels.

* The project supported by Natural Science Foundation of China（40074001) and Education Ministry of China

Analysis and Application of K-L Transform Information Features*

Fengxiang JIN and Shifeng DING, China

1. INTRODUCTION

Information Pattern Recognition is a new idea in data analysis, which was developed by the author. The main aim of it is to study the useful information included in the digital data accumulated historically, and to deal with the problems, such as Information Evaluation, Information Extraction, Information Measurement and Information Description[1] [2]. Pattern recognition theory is new scientific field developing from the recent 30 years. Its theory and method are widely used in almost every field. But its application in surveying and mapping field is just at the beginning. Some research work about its use in surveying was done by some experts[1] [2]. With the development of GIS, geography data is becoming richer, the information in it is becoming more and more. At the same time the quantity of it is becoming huger in volume and complex in relation. In order to effectively excavate and use the information contained in the data, a correspondent theory and method should be established. It can be used to realize the information analysis and information excavation. It will become a very important field in the digital data processing theory.

Generally, a large quantity of digital data from characteristics of object studied is obtained with the purpose of the problem study. Sometimes, a huge data set forms through historically accumulating. There are two problems with the analysis of it: the first is the huge number of characteristics brings some difficult to the data analysis; the second is the correlation among the characteristics makes data very complex. Usually, it is necessary that a data pre-processing procedure is realized before the data analysis, such as Feature Compression[3][4]. Before pre-processing an information analysis should be done in order to chose or determine the best data processing method. This information analysis includes Information Evaluation, Information Measurement and Information Feature Analysis. In addition, in order to establish the most reasonable mathematical model of the useful information between the information quantity and the information losing, the information quantity and the information structure should be studied. Therefore, this paper presents some parts of research work done in this field by our research group.

Many people have done some work with quite similar problems based on some methods, such as correlation analysis[5], principal component analysis[6], rough sets[7]. But it is far to

* The project supported by Natural Science Foundation of China（40074001）and Education Ministry of China

completely solve the problems. Especially, there is little work done from the viewpoint of information theory. This paper makes a brief introduction about the Orthogonal Transform with the Minimum of Mean Square Error(MSE), called K-L Transform. Then the information Features of K-L transform is analyzed from the viewpoint of information theory. A new idea is presented about using Indicating Entropy to measure the information quantity. Some active results are shown following a practical example.

2. K-L TRANSFORM AND FEATURES

Suppose there be a characteristic vector of n-dimension, $x=(x_1,x_2,\cdots,x_n)'$, its covariance matrix be $\Sigma_x = E[(x-\bar{x})(x-\bar{x})']$, and its expectation is $\bar{x}=E(x)$, $T=(\phi_1,\phi_2,\cdots,\phi_n)'$ be an orthogonal transform matrix, where ϕ_i is the corresponding eigenvector of the $i-th$ eigenvalue λ_i of Σ_x. The λ_i and ϕ_i are satisfied with

$$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_m \ge \cdots \ge \lambda_n \qquad \phi_i'\phi_j = \begin{cases} 1, & i=j \\ 0, & i \ne j \end{cases} \tag{1}$$

Based on the orthogonal transform, the vector x can be changed into $y=(y_1,y_2,\cdots,y_n)'$, that is

$$y = T'x = (\phi_1,\phi_2,\cdots,\phi_n)'x \tag{2}$$

Therefore

$$x=(T')^{-1}y = T\,y = (\phi_1,\phi_2,\cdots,\phi_n)y = \sum_{i=1}^{n} y_i\phi_i \tag{3}$$

With the condition of MSE, the transform $y=T'x$ is called K-L transform and $x=T\,y$ is called K-L expansion of x. The covariance of y is

$$\Sigma_y = E(y-\bar{y})(y-\bar{y})' = T'\Sigma_x T = diag(\lambda_1,\lambda_2,\cdots,\lambda_n) \tag{4}$$

In practice of data analysis, Σ_x is usually replaced by its estimation from the data. In addition, for the trace of the covariance is of certainty. It could be proved as follows.

[Proof] Suppose there be another orthogonal matrix $V=(v_1,v_2,\cdots,v_n)$, the orthogonal transform from it $z = V'x$. The variance of z_i is

$$m_i = \mathrm{var}(z_i) = E(z_i - \bar{z}_i)^2 \qquad (5)$$

According to orthogonal basis $\{u_j\}$, every orthogonal vector v_i could be expanded as follows

$$v_i = \sum_{j=1}^{n} a_{ij} u_j, \quad i = 1,2,\cdots,n \qquad (6)$$

Where coefficient is $a_{ij} = u_j' v_i, \quad i, j = 1,2,\cdots n$

Let

$$a_i = (a_{i1}, a_{i2}, \cdots, a_{in})' \qquad (7)$$

From formula (7) and orthogonal $v_i, v_j (i \neq j)$, we know that $a_1, a_2, \cdots, a_n$ are still orthogonal vectors. So, there is

$$m_i = \mathrm{var}(z_i) = v_i' \Sigma_x v_i = (\sum_{j=1}^{n} a_{ij}\phi_j)' \Sigma_x (\sum_{j=1}^{n} a_{ij}\phi_j) = (\sum_{j=1}^{n} a_{ij}\phi_j)'(\sum_{j=1}^{n} a_{ij} \Sigma_x \phi_j)$$

$$= (\sum_{j=1}^{n} a_{ij}\phi_j)'(\sum_{j=1}^{n} a_{ij}\lambda_j\phi_j) = \sum_{j=1}^{n} a_{ij}^2 \lambda_j = \sum_{j=1}^{n} a_{ij}^2 \,\mathrm{var}(y_j) \qquad (8)$$

Therefore, considering $\sum_{j=1}^{n} a_{ij}^2 = 1$, we have

$$\sum_{i=1}^{n} m_i = \sum_{i=1}^{n}[\sum_{j=1}^{n} a_{ij}^2 \,\mathrm{var}(y_j)] = \sum_{j=1}^{n}(\sum_{i=1}^{n} a_{ij}^2)\,\mathrm{var}(y_j) = \sum_{j=1}^{n} \mathrm{var}(y_j) = \sum_{j=1}^{n} \lambda_j \qquad (9)$$

So, the conclusion is correct.

3. INFORMATION FEATURE ANALYSIS AND INFORMATION MEASUREMENT

With the K-L transformation procedure following some rules, the information in x is totally transmitted to the new variable vector y. This transform make the internal information come out. It could be more easily for us to well understand the information distribution, information structure and information magnitude, and to get the information features about the data.

3.1 Indicating Features on Information

K-L transform indicates the all information in an orthogonal space. New variables are independent each other. Eigenvalue illustrates the information magnitude. The bigger the eigenvalue is, the more the information has. So the indicating features on information could be discussed from that.

1. K-L transform makes the original correlative information become independent information. And they are presented in an orthogonal space so that it makes the information category and magnitude study quite possible.

2. λ indicates the information magnitude. The bigger the eigenvalue is, the more the information has. When λ_i =0, it means there is not any information at i-dimension.

3. K-L transform presents the information distribution features at space and the effectiveness of the information transmission.

4. Because K-L transform is unique, the presented information structure is unique too. It makes the information measurement become possible.

3.2 Indicating Entropy and its information features

A transform of λ_i is carried out as follows

$$\rho_i = 1 - \lambda_i \Big/ \sum_{i=1}^{n} \lambda_i , \quad \sum_{i=1}^{n} \rho_i = 1, \quad 0 \le \rho_i \le 1 \tag{10}$$

Therefore, ρ_i is of the numerical properties of probability and similar with the definition of entropy function, we get the quasi entropy function and the indicating entropy as follows

$$I(\lambda_i) = -\log \rho_i , i = 1,2,\cdots,n) \tag{11}$$

$$H(T) = H(\rho_1, \rho_2, \cdots, \rho_n) = -\sum_{i=1}^{n} \rho_i \log \rho_i \tag{12}$$

$H(T)$ reflects unevenness of ρ_i or $\mathrm{var}(y_i)$. According to formulae (11) and (12), we have

1. $I(\lambda_i)$ denotes information magnitude in λ_i. The bigger λ_i is, the bigger $I(\lambda_i)$ is.

2. When all of λ_i are equal to each other, the value of the indicating entropy reaches maximum value.

3. Compared with other orthogonal transform, K-L transform gives the smallest value of the indicating entropy. It illustrates that the information description from K-l transform is the simplest one (best one). It could be proved as follows.

[Proof] Let formula (5) be transformed as follows

$$\sigma_i = 1 - m_i \Big/ \sum_{j=1}^{n} m_j , \quad \sum_{i=1}^{n} \sigma_i = 1, \quad 0 \le \sigma_i \le 1 \tag{13}$$

Consequently, the indicating information entropy is got

$$H(V) = H(\sigma_1, \sigma_2, \cdots, \sigma_n) = -\sum_{i=1}^{n} \sigma_i \log \sigma_i \tag{14}$$

From formulae（8）and（9）, we can obtain

$$\sigma_i = 1 - m_i \Big/ \sum_{j=1}^{n} m_j = 1 - \sum_{j=1}^{n} a_{ij}^2 \lambda_j \Big/ \sum_{j=1}^{n} \lambda_j = \sum_{j=1}^{n} a_{ij}^2 - \sum_{j=1}^{n} a_{ij}^2 \lambda_j \Big/ \sum_{j=1}^{n} \lambda_j$$

$$= \sum_{j=1}^{n} a_{ij}^2 (1 - \lambda_j \Big/ \sum_{j=1}^{n} \lambda_j) = \sum_{j=1}^{n} a_{ij}^2 \rho_j \tag{15}$$

Using Jensen inequality for lower convex function, we can get

$$\sigma_i \log \sigma_i = (\sum_{j=1}^{n} a_{ij}^2 \rho_j) \log(\sum_{j=1}^{n} a_{ij}^2 \rho_j) \le \sum_{j=1}^{n} a_{ij}^2 \rho_j \log \rho_j \tag{16}$$

According to formulae (14) and（16）, we can get

$$H(V) = -\sum_{i=1}^{n} \sigma_i \log \sigma_i \ge -\sum_{i=1}^{n} (\sum_{j=1}^{n} a_{ij}^2 \rho_j \log \rho_j)$$

$$= -\sum_{j=1}^{n} (\sum_{i=1}^{n} a_{ij}^2)(\rho_j \log \rho_j) = -\sum_{j=1}^{n} \rho_j \log \rho_j = H(T)$$

Because of arbitrariness of V, we have

$$H(T) = \min_{V} \{H(V)\} \tag{17}$$

So, the conclusion is correct.

4 Example

The data came from the samples of 7 characteristics, $x = (x_1, x_2, \cdots, x_7)'$ [10]. The estimation of Σ_x is R.

$$R=\begin{pmatrix} 1.0000 & -0.2912 & -0.5705 & -0.5676 & 0.8513 & -0.3886 & 0.0000 \\ & 1.0000 & 0.2725 & 0.2593 & -0.3444 & 0.1386 & 0.0000 \\ & & 1.0000 & 0.9869 & -0.7485 & 0.6695 & 0.0000 \\ & & & 1.0000 & -0.7356 & 0.7779 & -0.5855 \\ & & & & 1.0000 & -0.4729 & 0.2793 \\ & & & & & 1.0000 & -0.8875 \\ & & & & & & 1.0000 \end{pmatrix}$$

The eigenvalues of R, information rates, indicating entropy, accumulated information rates and contribution rates are listed in table 1.

Table 1

variable	eigenvalue (λ)	information rates (ρ)	$\rho_i \log \rho_i$	accumulated information rate (AIR)	Contribution rate (TCR)
1	4.2444	0.3937	0.5294	68.74	60.63
2	1.2513	0.8212	0.2333	83.26	78.51
3	0.9202	0.8685	0.1766	93.66	91.65
4	0.4384	0.9374	0.0874	98.43	97.91
5	0.1136	0.9838	0.0231	99.63	99.54
6	0.0314	0.9951	0.0070	99.99	99.99
7	0.0007	0.9999	0.0001	100.00	100.00
indicating entropy			1.0569		

From Table 1, some conclusions could be summarized as□

The information quantity of the data is 1.0569 bits. Compared with its maximum quantity $\log_2 6 = 2.8074$ (under the uniform distribution of information), it is small. So its information distribution is not uniform at each dimension. The main information is concentrated on the first 3 dimensions, i.e. the first new variables. It tells us the original data is strongly correlative.

REFERENCES

[1] Jin Fengxiang,Wang Tongxiao,Du zhixing. Pattern recognition on surveying data information analysis[J],The Chinese Journal of Nonferous Metals,1999,9(2):425-430 (in chinese)

[2]Du Zhixing,Jin Fengxiang,Wang Tongxiao et al. Pattern Recognition Method for Periodical Error Test[J], Sichuan Surveying and Mapping. 1999,22(1):11-14 (in chinese)

[3]Bian Zhaoqi,Zhang Xuegong. Pattern Recognition[M]. Beijing: Tsinghua University Publishing House,2000 (in chinese)
[4]Sun Jixiang. Modern Pattern Recognition[M]. Changsha: The Defence University of Science and Technology Publishing House,2002 (in chinese)
[5]Ji Xiaojun,Li shizhong,Li Ting. Application of the correlation analysis in feature selection[J], Journal of Test and Measurement Technology,2001,15(1):15-18 (in chinese)
[6]Zeng Huanglin,Yu Juebang,Zeng Qian. System feature reduction on principal component analysis [J],Journal of Sichuan Institute of Light Industry and Chemical Technology,1999,12(1):1~4 (in chinese)
[7]Zeng Huanglin,Yuan Zengren. About a new approach of selection and reduction on system feature in pattern recognition [J], Journal of Sichuan Institute of Light Industry and Chemical Technology 1999,12(4):1-5 (in chinese)
[8]Shannon,C.E. A Mathematical Theory of Communication[J]. BSTJ,1948,379-423.
[9]Jiang Dan. Theory of Information[M].Hefei:The Publishing House of China University of Science and Technology,1987(in chinese)
[10]Li Jinzong. Introduction of Pattern Recognition[M]. Beijing:High Education Press,1994 (in chinese),442
[11]Wang Xueren. Multivariance Statistics Analysis of Geology Data[M]. Beijing: Science Press,1982 (in chinese),298-300
[12]Jin Fengxiang etc. Patter Recognition Problem in Deformation Monitoring and Analysis[J], The Chinese Journal of Nonferrous Metals. 1997(3)

CONTACS

Prof. Fengxiang Jin, Dr-Eng.
Geoscience Dept
Shandong University of Science and Technology
Taian,Shandong Prov. (271019)
P. R. CHINA
Tel +86 538 622 6998 (H)
Tel +86 538 622 7099 (O)
Fax +86 538 622 7415
Email: Fxjin@sina.com and Jinfengxiang@yahoo.com
Web site: http://myprofile.cos.com/jinfengxiang