GEODEZJA I KARTOGRAFIA
t. LII, z. 3, s. 133–142 (2003)

Józef Gil

Department of Land and Environment Engineering
Institute of Building Engineering
University of Zielona Góra
(65-246 Zielona Góra, ul. Z. Szafrana 2)

PCA transformation as a method of compression of experiment results

The paper presents a method of statistical analysis of multidimensional data by means of the PCA transformation (Principal Components Analysis), carried out by means of a neuron network. The PCA transformation, connected with the Karhunen-Loeve transformation is used for processing signals treated as stochastic processes. The method discussed in the paper, enables the reduction of input data on the basis of specified independent principal components with respect to their significance.

INTRODUCTION

One of the basic methods of statistical analysis of multidimensional data is the method of principal components analysis PCA, used for processing signals treated as stochastic processes. A closer look at this problem leads to the conclusion that it is a transformation of co-ordinates from a particular system to a system that is better adapted to the changeability of existing data. The PCA transformation generates new variables called principal components, which are represented by eigenvectors of a covariance matrix or correlation matrix. The use of some of the principal components enables the reduction of the size of the database with respect to its best representation as the maximum covariance matrix in terms of the trace and the determinant of the matrix. Useful information in tasks of this type is a description of the system generating existing data as a linear or non-linear filter.

The paper discusses a PCA transformation based on a correlation matrix, carried out by means of neuron networks. It deals with the problem of eliminating from the learning set, the features that have little influence on the extent of variation of particular objects.

1. *The energetic function of correlation networks*

According to the basic rule by Hebb, a change of the weight vector $\boldsymbol{w}_i = [w_{i1}, w_{i2},...,w_{iN}]^T$ is proportional to the product of the input signal $\boldsymbol{x} = [x_1,x_2,...,x_N]$ and the learning signal [4].

$$y_i = f(\boldsymbol{w}_i^t \boldsymbol{x}). \tag{1}$$

which is the output signal of a neuron i.e. a linear weighted adder (Fig. 1). The set of joint stimulations of a neuron is the domain of the linear activation function $f(\boldsymbol{w}_i^t \boldsymbol{x})$. The increase of the weight vector $\Delta \boldsymbol{w}_i$ for the linear model of a neuron consistent with Hebb's rule is described by the dependence

$$\Delta \boldsymbol{w}_i = cf(\boldsymbol{w}_i^t \boldsymbol{x})\boldsymbol{x}, \tag{2}$$

where the learning constant c is a positive number. By changing formula (2) a change of the weight of a neuron can be expressed as follows

$$\Delta w_{ij} = c(y_i - y_0)(x_j - x_0), \tag{3}$$

where x_0 and y_0 are certain constants. On the assumption that the learning coefficient $c = 1$

$$\Delta w_{ij} = y_i x_j - y_i x_0 - y_0 x_j + y_0 x_0. \tag{4}$$

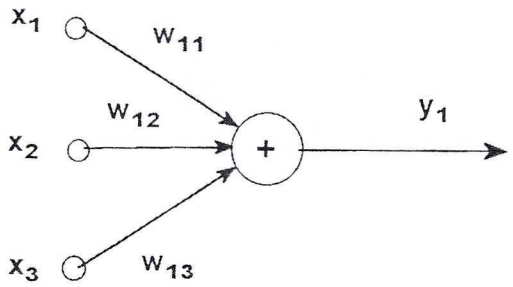

Fig. 1. Adaptation linear weighted adder

By adopting a system consisting of a single linear neuron and the number of learning standards $N = 3$ (the number of the coordinates of the input vector), the increase of weights expressed in the scalar form is

$$\begin{aligned} \Delta w_{1j} = {} & w_{11}x_1x_1 + w_{12}x_2x_1 + w_{13}x_3x_1 + \\ & w_{11}x_1x_2 + w_{12}x_2x_2 + w_{13}x_3x_2 + \\ & w_{11}x_1x_3 + w_{12}x_2x_3 + w_{13}x_3x_3 - \\ & w_{11}x_1x_{01} + w_{12}x_2x_{01} + w_{13}x_3x_{01} - \\ & w_{11}x_{01}x_1 + w_{12}x_{01}x_2 + w_{13}x_{01}x_3 + \\ & x_0, y_0 \end{aligned} \tag{5}$$

Therefore, for any number N of learning standards and k neurons, a change of weights of the network in time t, can be expressed in the form of the equation [2]

$$\frac{\mathrm{d}w_{ij}}{\mathrm{d}t} = \sum_{k=1}^{N} w_{ik} A_{kj} + \sum_{k=1}^{N} c_{ik} w_{ik} + d_k, \tag{6}$$

where A_{kj} is the covariance matrix of the activity of neurons k and j i.e.

$$A_{kj} = \sum_{i=1}^{p} (x_k^{(i)} - \bar{x}_k)(x_k^{(i)} - \bar{x}_j), \tag{7}$$

where: $\bar{x}_k$ – the average value of input vector components, p – the number of input vectors.

If it is assumed that the change of weights is effected numerically by means of the method of the greatest decrease, the minimization of the energetic function Q (the specified characteristic value dependent on the condition of the network) progresses in the direction opposite to the gradient vector, i.e.

$$\frac{dQ}{dw_{ij}} = -\frac{dw_{ij}}{dt} = -\sum_{k=1}^{N} w_{ik} A_{kj} - \sum_{k=1}^{N} c_{ik} w_{ik} - d_k. \tag{8}$$

The form of an energetic function will be obtained after the above differential equation has been solved. Therefore it is possible to say:

$$Q = -\frac{1}{2}\sum_{k=1}^{N}\sum_{j=1}^{N} w_{ik} A_{kj} w_{ji} - \frac{1}{2}\sum_{k=1}^{N}\sum_{j=1}^{N} c_{ik} w_{ik} w_{ji} - \sum_{j=1}^{N} d_i w_{ji}. \tag{9}$$

Two components can be separated from the form of the function Q (formula 9)

– the first

$$Q_1 = \frac{1}{2}\sum_{k=1}^{N}\sum_{j=1}^{N} w_{ik} A_{kj} w_{ji}, \tag{10}$$

– the second

$$Q_2 = -\frac{1}{2}\sum_{k=1}^{N}\sum_{j=1}^{N} c_{ik} w_{ik} w_{ji} - \sum_{j=1}^{N} d_i w_{ji}. \tag{11}$$

The first component denotes the variance σ_i^2 of the activity i – of that neuron. The proof of that statement will be presented in the following part of the paper. The second component of the energetic function can be associated with an objective function in the theory of optimisation. It is so, because in search of optimum weight values, the gradient vector ∇ of the function Q_2 (w) will be equated with zero, i.e.

$$\nabla = \boldsymbol{c}\,\boldsymbol{w} - \boldsymbol{d} = 0 \tag{12}$$

After equation (12) has been solved the vector will be obtained

$$\boldsymbol{w}^* = \boldsymbol{c}^{-1}\boldsymbol{d} \tag{13}$$

2. *The PCA network*

The PCA network is a one-layer network with linear neuron activation functions (Fig.2), and the analysis of the principal components by means of the PCA network without changing the energy of the signals [2] should be treated as the linear transformation method $\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x}$, which transforms a stationary stochastic process in the form of the vector $\boldsymbol{x} \in R^N$ into the vector $\boldsymbol{y} \in R^{K \times N}$.

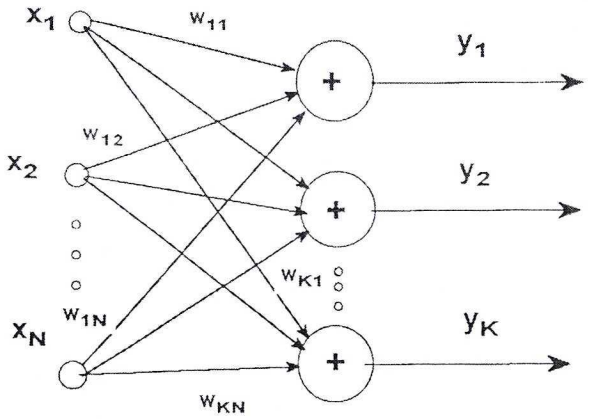

Fig. 2. A unidirectional neuron network for data compression

The transformation process is carried out by means of the non-singular matrix $\boldsymbol{W} \in R^{M \times N}$ for $M \leq N$ with the preservation of basic information concerning the progress of the process. In other words, the PCA transformation for $M \leq N$ is the method of loss compression (the Karhunena-Loève transformation), which in effect decreases the amount of information in the input data into a set of statistically independent components with respect to their significance.

A general geometric interpretation of the linear transformation of the vector $\boldsymbol{x}$ into the vector $\boldsymbol{y}$ with the participation of a specified matrix of the transformation $\boldsymbol{W}$ is shown in Fig. 3.

When a single random observation vector with the average value zero is marked as $\boldsymbol{x} = [x_1, x_2, ..., x_N]^T$, then the autocorrelation matrix for p observation vectors is

$$\mathbf{A} = \sum_{k=1}^{p} x_k x_k^T = XX^T. \tag{14}$$

The matrix $\mathbf{A}$ is a symmetric, real, non-negatively definite matrix, which can be assigned its eigenvalues $\lambda_i (i = 1,2,...,N)$ and orthonormal vectors of eigenvalues $\boldsymbol{w}_i = [\boldsymbol{w}_{i1}, \boldsymbol{w}_{i2}, ..., \boldsymbol{w}_{iN}]$. The matrix $\mathbf{A}$ and the matrix $\boldsymbol{\Lambda} = (\lambda_1, \lambda_2, ..., \lambda_N)$ are congruent matrixes [1], because

$$\mathbf{W}^T \mathbf{A} \mathbf{W} = \boldsymbol{\Lambda}. \tag{15}$$

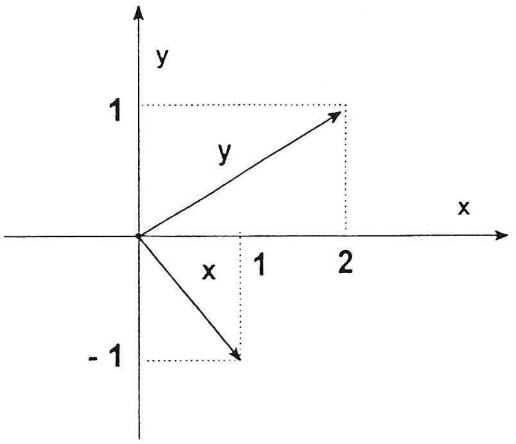

Fig. 3. The linear transformation of the vector $\boldsymbol{x}$ into the vector $\boldsymbol{y}$

If the eigenvalues are put in a series according to the rule

$$\lambda_1 > \lambda_2 > ... > \lambda_N \geq 0 \tag{16}$$

and also the eigenvectors w_i corresponding to the eigenvalues λ_i, and if our interest is limited only to M first eigenvalues, then the matrix

$$\mathbf{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_M]^T \qquad (\mathbf{W} \in R^{M \times N}) \tag{17}$$

and the vector

$$\boldsymbol{y} = \mathbf{W}\boldsymbol{x} \qquad (\boldsymbol{y} = y_1, y_2, ..., y_M), \tag{18}$$

is the vector of principal components. Next, bearing in mind the series of eigenvalues λ_i ($i = 1,2,...,M$) the modules of the coordinates of the vectors $\boldsymbol{y}_i$ ($i = 1,2,...,p$) (transform $\boldsymbol{y}_i$) will be ordered according to decreasing values. If $\lambda_N = 0$, values y_N are equal for all $i = 1,2,...,p$.

From the above, the vector $\boldsymbol{y}$, being the vector of the principal coordinates PCA, has the greatest influence on the reconstruction of the vector $\boldsymbol{x}$ according to the dependence:

$$\check{\boldsymbol{x}} = \mathbf{W}^{-1}\boldsymbol{y} = \mathbf{W}^T\boldsymbol{y} \qquad (\mathbf{W}^T\mathbf{W} = \mathbf{I}, \mathbf{W}^T = \mathbf{W}^{-1}), \tag{19}$$

which is called the Karhunena-Loève resolution. The first principal component $y_1 = \boldsymbol{w}_1^T\boldsymbol{x}$ is a normalized linear combination of those components of the input vectors, which make it possible to obtain the greatest variance value equal λ_1.

In order to get more information in this matter the covariance matrix of vectors of principal components will be calculated, i.e.

$$\mathbf{S} = \sum_{i=1}^{p} [y_i - \bar{y}][y_i - \bar{y}], \tag{20}$$

where:

$$\bar{y} = \frac{1}{p}\sum_{i=1}^{p} y_i = \frac{1}{p}\sum_{i=1}^{p} \mathbf{W}x_i = \mathbf{W}\bar{x}. \tag{21}$$

The matrix **W** consists of orthonormal eigenvectors of the matrix **A**, so

$$\mathbf{WAW}^T = \begin{bmatrix} w_1^T \\ w_2^T \\ \cdots \\ w_N^T \end{bmatrix} \mathbf{A}\,[w_1, w_2, \ldots, w_N] \tag{22}$$

On the basis of dependence (13) of the vectors and eigenvalues of the matrix **A** it is possible to say

$$\mathbf{S} = \begin{bmatrix} w_1^T \\ w_2^T \\ \cdots \\ w_N^T \end{bmatrix} [\lambda_1 w_1, \lambda_2 w_2, \ldots, \lambda_N w_N] = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N). \tag{23}$$

It results from the above equation that the coordinates of vectors $\boldsymbol{y}_i$ ($i = 1,2,\ldots,p$) are not correlated, because the central product moments of the second rank are equal zero. To sum up, the correlation space matrix $\boldsymbol{y}$ is a diagonal matrix of the eigenvalues of the data correlation matrix **X**.

A particular amount of information, corresponding to the energy of the components of the vector $\boldsymbol{x}$, is accumulated in the principal components $y_1, y_2, \ldots, y_M$ of the vector $\boldsymbol{y}$ that represent the components of the vector $\boldsymbol{x}$. The eigenvalue λ_1 connected to the first and at the same time the greatest principal component in terms of the module y_1 and the eigenvector $\boldsymbol{w}_1$ can be regarded as a geometrical object called a "concentration ellipsoid". The orientation of the ellipsoid and the squares of the length of the principal semi-axes in multidimensional space are determined by the eigenvectors and the eigenvalues of the autocorrelation matrix **A**.

On the basis of this geometrical interpretation it is necessary to say that the component y_1 determines the direction of the greatest variance of the input data, and the smallest variance should be found along the direction perpendicular to the direction y_1. Fig. 4 is a graphical illustration of the problem in question

With respect to the above the reconstruction of the vector $\boldsymbol{x}$ can be achieved on the basis of principal components with a loss to the reconstruction corresponding to the influence of the disregarded least significant components. A reduction of dimensions consists in restricting the dimension of the matrix **W** only to M columns, corresponding to the greatest eigenvalues of the diagonal matrix **S**.

Application. The vertical deviation of two atmospheric coolers was calculated. Forty metres high cylindrical coolers were placed over a certain area in four rows, 8 coolers in each row. The measurements were carried out on 6 levels excluding the lowest

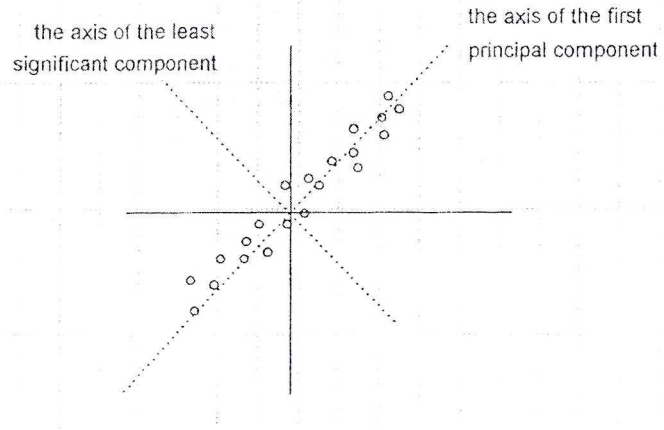

Fig. 4. The direction of the greatest and the smallest variance of input data determined by principal components

available level of the shaft of the coolers. 32 input vectors with six components were obtained from the measurements. After the input vectors had been normalised ($\| \boldsymbol{x} \| = 1$), the correlation matrix **A** was estimated for the input image $\boldsymbol{x}$ as

$$\mathbf{A} = \begin{bmatrix} 10.5597 & 5.4957 & 0.4847 & -3.1897 & -4.6856 & -8.6647 \\ \cdots & 3.8446 & 1.3868 & -1.5818 & -3.5180 & -5.6272 \\ \cdots & \cdots & 1.9678 & 0.0885 & -1.8419 & -2.0858 \\ \cdots & \cdots & \cdots & 1.4725 & 1.5064 & 1.7042 \\ \cdots & \cdots & \cdots & \cdots & 4.0106 & 4.5286 \\ \cdots & \cdots & \cdots & \cdots & \cdots & 10.1450 \end{bmatrix}.$$

The decomposition of the matrix **A** according to eigenvalues by means of the Jacoby method and the eigenvectors associated with them had the following results: $\lambda_1 = 25.9442$, $\lambda_2 = 3.5959$, $\lambda_3 = 2.0778$, $\lambda_4 = 1.2197$, $\lambda_5 = 0.1625$, $\lambda_6 = 0$, and the matrix of the coordinates of the eigenvectors (direction cosines)

$$\mathbf{W}^* = \begin{bmatrix} -0.6037 & -0.3718 & -0.1099 & 0.1638 & 0.3310 & 0.5906 \\ 0.5434 & -0.1408 & -0.6531 & -0.3188 & 0.3001 & 0.2642 \\ 0.0495 & 0.1568 & 0.1695 & -0.3934 & -0.6194 & 0.6371 \\ -0.4045 & 0.7187 & -0.1557 & -0.4316 & 0.3264 & -0.0532 \\ -0.0865 & 0.3664 & -0.5846 & 0.6058 & -0.3708 & 0.0778 \\ 0.4083 & 0.4083 & 0.4083 & 0.4083 & 0.4083 & 0.4083 \end{bmatrix}.$$

The process of the adaptation of network weights for the adopted value $c = 0.07$ was carried out by means of the PCA method with respect to the Sanger rule [3], which normalizes weight vectors ($\| \boldsymbol{x} \|^2 = 1$) according to the formula

$$w_{ij}(k+1) = w_{ij} + cy_i(k)[x_j - \sum_{h=1}^{i} w_{hj}(k) y_h(k)] \tag{24}$$

and the following result was obtained

$$\mathbf{W} = \begin{bmatrix} -0.6259 & -0.3551 & -0.0841 & 0.1623 & 0.3004 & 0.6025 \\ -0.5510 & 0.1378 & 0.6621 & 0.3082 & -0.3105 & -0.2466 \\ 0.0472 & 0.1512 & 0.1736 & -0.3944 & -0.6215 & 0.6438 \\ -0.4038 & 0.7232 & -0.2092 & -0.3896 & 0.2915 & -0.0181 \\ 0.0959 & -0.3584 & 0.5111 & -0.5848 & 0.3988 & -0.0625 \\ 0.1251 & 0.5914 & 0.4305 & 0.0052 & 0.6002 & 0.2986 \end{bmatrix}.$$

where the initial choice of a weight vector was a random choice. The process of adaptation of weights can be regarded as completed if increase of the coordinates of the weight vector is small ($\| \Delta \boldsymbol{w} \| \cong 0$). The result of the numerical solution of the task being one vector of principal components $\boldsymbol{y}$ and correspondingly the input vector $\boldsymbol{x}$ and its reconstruction $\check{\boldsymbol{x}}$ have been shown below, i.e.:

$$\boldsymbol{y} = [-0.9559, 0.1873, -0.1212, 0.0120, -0.1484, 0]^T$$
$$\boldsymbol{x} = [7.6, 8.1, 28.9, 37.3, 62.2, 85.6]^T$$
$$\check{\boldsymbol{x}} = [5.9, 10.1, 31.0, 36.0, 60.7, 85.9]^T$$

Looking at the above data it is possible to notice that a complete consistency between the coordinates of the weight vector **W** determined by means of the Jacoby method and the PCA method was not achieved. In the author's opinion, the reason for the inconsistency is specific character of the method.

The speed of convergence of the algorithm depends on the choice of the learning factor c. A good choice of c can be achieved on the basis of the author's suggestion according to the formula

$$0 < c < 2/\lambda_{\max} \tag{25}$$

where $\lambda_{\max}$ is the greatest eigenvalue of the matrix **A**. The Hebb network in question is a linear network, but the learning algorithms are non-linear. Because of unspecified weight restrictions the factor c should be chosen in such a way that the increase of weights in terms of value stay within the linear model. In this way the Lagrange remainder

$$\frac{1}{2}\lambda_{\min} \| \Delta w \|^2 < R_2(w) < \frac{1}{2}\lambda_{\max} \| \Delta w \|^2. \tag{26}$$

Marking

$$c = \frac{\| \Delta w \|^2}{R_2(w)} \tag{27}$$

and bearing in mind that the eigenvalue $\lambda_{\min}$ of the matrix **A** equals 0, formula (26) is transformed into formula (25).

4. CONCLUSION

The linear transformation $\boldsymbol{y} = \mathbf{W}\boldsymbol{x}$ is an adaptation transformation, which functions on the basis of the generalized Hebb algorithm.

The transformation process progresses in the on-line mode and is passed on the series of two vectors $\boldsymbol{x}$, whose way of acquisition excludes the possibility of an overt form of autocorrelation matrix. Nevertheless, adaptation methods are indispensable while determining eigenvectors in cases when the vector $\boldsymbol{x}$ has large dimensions, because the accuracy of results obtained by means of the *QR* method or the Jacoby method is usually insufficient.

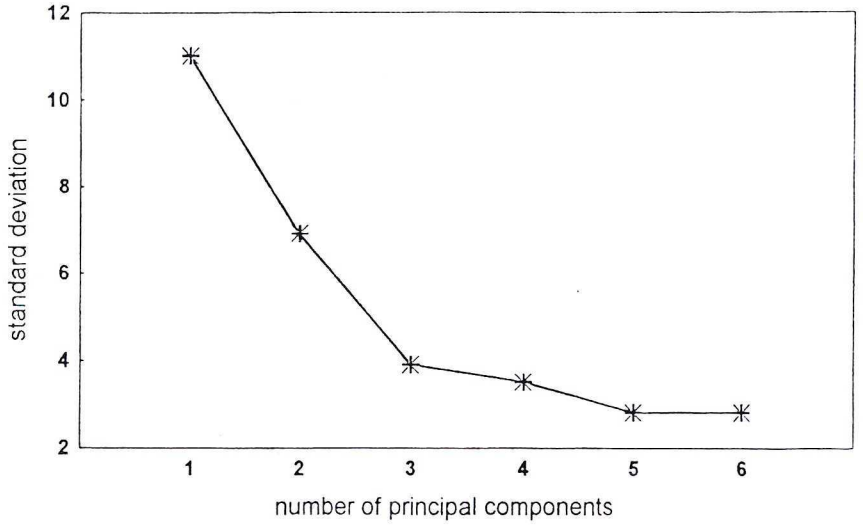

Fig. 5. The characteristics of the accuracy of data reconstruction with respect to the number of principal components

The basic numerical characteristic feature of data of reconstruction error is the variance of the random variable υ, defined as the expected value of the square of the deviation of the random variable from its expected value

$$\upsilon(\boldsymbol{x}) = E[\boldsymbol{x} - E(\boldsymbol{x})]^2 . \tag{28}$$

Figure 5 shows the standard deviation value $\sigma = \sqrt{\upsilon(x)}$ for an increasing number of principal components. The standard deviation σ of data reconstruction is within $11.0 \div 2.8$ mm. The compression quality depends on the number of principal components of the PCA transformation.

REFERENCES

[1] Kwiatkowski W., *Metody automatycznego rozpoznawania danych*, Instytut Automatyki i Robotyki, Wydział Cybernetyki, WAT, Warszawa 2001.

[2] Osowski S., *Sieci neuronowe do przetwarzania informacji*, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2000.
[3] Sanger T. D., *Optimal unsupervised learning in single layer linear feedformed neural network*, vol. 2, Neural Networks, 1989.
[4] Żurada J., Barski M., Jędruch W., *Sztuczne sieci neuronowe*, PWN, Warszawa 1996.

Received April 8, 2003
Accepted August 29, 2003

Józef Gil

Transformacja PCA jako metoda kompresji wyników eksperymentu

Streszczenie

W pracy przedstawiono metodę analizy statystycznej wielowymiarowych danych za pomocą transformacji PCA (Principal Components Analysis), zrealizowaną za pomocą sieci neuronowej. Transformacja PCA, związana z transformacją Karhunena-Loeve jest stosowana w przetwarzaniu sygnałów traktowanych jako procesy stochastyczne. Omawiana w pracy metoda, umożliwia redukcję przestrzeni danych wejściowych na podstawie wyznaczonych niezależnych składników głównych z uwzględnieniem ich znaczenia.

Юзеф Гиль

Трансформация PCA как метод компрессии результатов эксперимента

Резюме

В работе представлен метод статистического анализа многомерных данных с помощью трансформации PCA (Principal Components Analysis), реализованной с помощью нейроновой сети. Трансформация PCA, связанная с трансформацией Karhunena-Loéve, применяется для обработки сигналов, рассматриваемых как стохастические процессы. Представленный в работе метод даёт возможность редукции пространства исходных данных на основе определённых независимых главных компонентов с учётом их значения.