NEW METHODS AND DEVICES

A BASIS FOR CLASSIFICATION OF ROCKS BY THEIR STRENGTH PROPERTIES

G. A. Bakhtin and A. M. Kabakov

UDC 622.02; 553.076

In this article we give a description and analysis of existing methods of classification of multidimensional objects by means of various statistical criteria. For classifying rocks by their mechanical properties we suggest the method of principal components, which enables us to replace the action of a number of measured variables by a smaller number of unobserved quantities or factors. The noninvariance which arises in this method is removed by means of a special scaling method recommended by the authors. The proposed method of classification differs to advantage from all others in requiring much less calculation and in being much more reliable. It has been used to classify overburden rocks in coal fields.

In investigations of the strength properties of rocks in coal fields the results of tests on N specimens for m properties can be written as a matrix

$$X = (x_{ij}),$$

where $i = 1, 2, 3, \ldots, N$ is the index number of the specimen, and $j = 1, 2, 3, \ldots, m$ is the index number of the rock property.

The data in this matrix can be regarded as values of N random m-dimensional normally distributed quantities.

In classifying rocks by their strength properties by means of matrix (1) we must choose between two possibilities: 1) The matrix is homogeneous and has no internal boundaries; 2) the matrix is inhomogeneous, i.e., it can be subdivided by means of some criterion into at least two classes. To do this we must first choose the criteria of homogeneity.

Let us arbitrarily divide matrix (1) into two submatrices, $X^{(1)}$ and $X^{(2)}$. Let $X^{(1)}$ contain $n_1$ rows and $X^{(2)}$, $n_2$ rows. As a criterion for testing the hypothesis of homogeneity, Rodionov [1] suggests the function

$$V(r^2) = \frac{\sum_{j=1}^{m} (\bar{x}_{j1} - \bar{x}_{j2})^2 n_1 n_2}{\sum_{j=1}^{m} S_{j1}^2 (n_1 + n_2)},$$

where $\bar{x}_{j1}$ and $\bar{x}_{j2}$ are the mean values of the j-th property of the rocks over matrices $X^{(1)}$ and $X^{(2)}$, respectively, $S_{j1}^2$ is an estimate of the dispersion of the j-th property.

If the hypothesis of homogeneity is correct, then the value of $V(r^2)$ will exceed the tabulated value of $\chi^2$ only with very small probability q for a number of degrees of freedom $m$ for all possible variants of subdivision of matrix X. Consequently the hypothesis of homogeneity can be adopted if $V(r^2) \leq \chi_{qm}$ for all $r^2 \in R^2$.

If, however, $V(r^2) > \chi_{qm}^2$ for even one $r^2 \in R^2$, then the hypothesis of homogeneity of X must be rejected.

The shortcoming of the homogeneity criterion (2) is that it needs a very large number of variant calculations. The total number of variants of subdivisions of $R^2$ contains $2^{N-1}$ elements, and this problem is therefore difficult even with a computer. Furthermore, in testing the hypothesis of homogeneity by (2) we establish only...
the possibility or impossibility of subdividing $\mathbf{X}$ into more homogeneous classes. If the hypothesis of homogeneity is rejected, only one thing follows: in $\mathbf{X}$, according to the complex of signs, we can distinguish at least two homogeneous classes. From this criterion we cannot judge the rational number of such classes in $\mathbf{X}$. To eliminate this shortcoming, we must order matrix $\mathbf{X}$ in some way.

Each row and column in $\mathbf{X}$ contains some information on the variability of the strength properties of various rocks in coal deposits. However, these properties are mutually correlated. Therefore in classifying rocks by their properties we must take account not only of their level of variability but also of the mutual correlation. These data are contained in the covariation matrix

$$L = \mathbf{X}^\top \mathbf{X} / (N-1),$$

where $\mathbf{X}^\top$ is the transposed matrix of $\mathbf{X}$.

The diagonal elements of (4) are dispersions, and the off-diagonal elements are covariations.

To order matrix $\mathbf{X}$ with the aid of $\mathbf{L}$ it is convenient to use its first principal component,

$$z = \sum_{j=1}^{n} a_{ij}x_{ij}.$$

An algorithm for calculating the principal components was discussed in detail in [2]. We have devised a very reliable program for realizing this algorithm on a Nairi computer. The first principal component extracts the maximum information from (4), and therefore from it we can also judge the information contained in (1).

A drawback of the method of principal components is its noninvariance with respect to the ratio to the dimensionality of the variables characterizing the properties of the rocks; this cannot be removed even if we translate to standardized variables [4]. In our opinion, the noninvariance is practically eliminated by making the change of variables

$$x_{ij} = (x_{ij} - \bar{x}_{ij})/\sigma_{ij}.$$

Ordering of (1) by the method of principal components is carried out as follows.

1) By Eq. (6), matrix (1) is converted to a matrix of scaled variables $x_{ij}$.

2) From the matrix of scaled variables we determine the covariation matrix (4).

3) From matrix (4) we determine the eigenvector elements corresponding to the greatest eigenvalue of the matrix.

4) From Eq. (5) for each row in (1) we calculate the first principal component.

5) We transpose the rows of matrix (1) in order of increasing principal component.

As a result of the linear ordering of (1), we can no longer arbitrarily transpose the rows, i.e., change the arrangement of the elements of the set of indices $i = 1, 2, 3, \ldots, N$. Here the set of subdivisions of $\mathbb{R}^2$ will contain only $N-1$ variants, and as a result it becomes really possible to subdivide matrix (1) into classes with respect to the variant for which $V(r^2)$ is a maximum.