ISSN 2414-9055
UDC 621.926 https://doi.org/10.31721/2414-9055.2016.2.2.31

CLUSTERING RESULTS DIMENSION REDUCTION OF IRON ORE RAW MATERIALS CHARACTERISTICS IN THE PROCESS CONTROL OF ITS PROCESSING

Morkun V.S., ScD, Professor, Morkun N.V., PhD, Associated Professor,
Tron V.V., PhD, Associated Professor
Kryvyi Rih National University

Abstract. The analysis results of dimension reduction methods for the clustering characteristics results of the iron ore raw materials in the process control of its processing are proposed

Keywords. Automation, control system, mathematical model, dimension reduction.

Introduction. Technological units of concentrating production as a control objects can be represented as operators, which are transform the vectors of the input variables in the vectors of output parameters [1-7]. Thus, the output parameters of one technological process are considered as input for the next. For example, the vectors elements of the grinding aggregates output parameters are their qualitative and quantitative indicators. The main quantitative indicators are: feed rate and performance by finished grain-size class. The quality of the ground product is characterized by: density or solids content in the pulp, solid phase particle size distribution, and mineral content in separate size classes of ground material. Thus, in solving the problem of synthesis control of iron ore raw material beneficiation technological processes it is necessary to operate the high dimension data.

Materials and methods. The presence of several mineralogical and technological ore varieties, each of which has several characteristics significantly complicates the processing of data on the technological process parameters [3-7]. Consequently, at the initial stage of processing it is necessary to analyze the possibilities to reduce the initial data dimension and then pass to the formation of mathematical models based on them.

Let's consider an example of the result of three-dimensional data distribution, which describing the mineralogical and technological characteristics of iron ore raw material varieties, which is processed on concentrating plant technological line, by 9 clusters (Fig. 1).

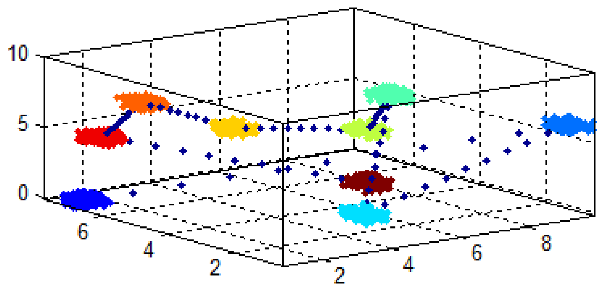


Figure 1. The result of ore varieties characteristics clustering

One of the most common methods of the dimension reducing is a Principal Components Analysis (PCA) [8]. The presentation of data, which formed by it with lower dimension describes the direction of the greatest change in the initial data by finding the lower-dimensional linear basis for the initial multi-dimensional data, in which the dispersion is maximum. The result is a linear transformation of *Θ*, which maximizes the expression

$$\Theta^T \operatorname{cov}_{X-\bar{X}} \Theta \to \max , \quad (1)$$

where $\operatorname{cov}_{X-\bar{X}}$ – is the covariance matrix of data of *X*, which centered relative to the coordinates origin. The result of dimension reduction using PCA is shown in Fig. 2,a.

Nonlinear multidimensional scaling method (MDS) [8-10] provides the dimension reduction while preserving the pairwise distances between the initial data points.

The quality of the conversion is described by a function which evaluates the pairwise distances difference in the initial multi-dimensional presentation and the resulting representation of lower dimension

ISSN 2414-9055

$$F(Y)=\sum_{ij}\left(\left\|x_i-x_j\right\|-\left\|y_i-y_j\right\|\right)^2 \qquad (2)$$

where $\left\|x_i-x_j\right\|$ – is the Euclidean distance between the data points of high dimension; $\left\|y_i-y_j\right\|$ – is the Euclidean distance between the data points of low dimension.

The alternative to the above considered function can be the Sammon cost function, which puts more emphasis on preserving originally short distances

$$F(Y)=\frac{1}{\sum_{ij}\left\|x_i-x_j\right\|}\sum_{ij}\frac{\left(\left\|x_i-x_j\right\|-\left\|y_i-y_j\right\|\right)^2}{\left\|x_i-x_j\right\|} \qquad (3)$$

Minimization of the stress function is performed using different methods of conjugate gradients [9]. The result of the dimension reduction using the MDS method shown in Fig. 2,b.

The drawback of the multidimensional scaling method is that it does not allow to consider the distribution of adjacent points because it is based on the Euclidean distances.

For example, in the case where the multi-dimensional data is on the curvilinear manifold, the distance between them can be significantly larger than the Euclidean. In this case it is advisable to use the Isomap method [8, 11], which considers the curvilinear distance between data points on a given manifold. In Isomap [11], the geodesic distance between x_i data points are calculated by plotting a graph, in which each point of x_i associated with it k nearest neighbors x_{ij} in a data set of X. The shortest path between two points of the graph is the estimate of the curvilinear distance between these two points, which are determined using Dijkstra's algorithm [12, 13].

Results. The result of dimension reduction using the Isomap method shown in Fig. 2,c. Diffusion maps method [8, 14, 15] assumes the formation of the data graph. The edges weights are calculated using the Gaussian kernel functions, which leading to the formation of the matrix of Ψ. The elements of the matrix of Ψ are calculated by the formula

$$\psi_{ij}=\exp\left(-\left\|x_i-x_j\right\|^2/2\sigma^2\right) \qquad (4)$$

where σ – is the dispersion. Then, based on the matrix of Ψ the normalized matrix of Ψ^* is calculated, the elements of which are determined from the expression

$$\psi^*_{ij}=\frac{\psi_{ij}}{\sum_k \psi_{ik}} \qquad (5)$$

The obtained matrix of Ψ^* is considered as a stochastic matrix, which determines the forward transition probability matrix of the dynamic process. Consequently, the matrix Ψ^* represents the transition probability from one data point to another data point per unit time. On the basis of transition probabilities $\psi^*(t)_{ij}$ the diffusion distance is determined [14, 15]

$$D^{(t)}\left(x_i,x_j\right)=\sum_k\frac{\left(\psi(t)_{ik}-\psi(t)_{jk}\right)^2}{\alpha\left(x_k\right)^{(0)}} \qquad (6)$$

where $\alpha(x_i)^{(0)}=m_i/\Sigma_j m_j$ – is the coefficient, which gives more weight to the elements of the graph with higher density; $m_i=\Sigma_j\psi_{ij}$ – is the degree of the node. From this equation we can see that the pair of points with a higher transition probability has a lower diffusion distance.

The idea of the diffusion distance is that it is based on many graph paths, which provides greater noise immunity than the geodesic distance. The representation of lower dimension of Y, which allows to keep the diffusion distance using a spectral theory, which is formed from d non-zero principal eigenvectors, which are found from the expression [14, 15]

$$\Psi^*(t)Y = \lambda Y \qquad (7)$$

As the graph is fully connected, the largest eigenvalue is a null, i.e $\lambda_1=1$, his own vector v_1 is not considered. In the representation of lower dimension the eigenvectors are normalized by the corresponding eigenvalues

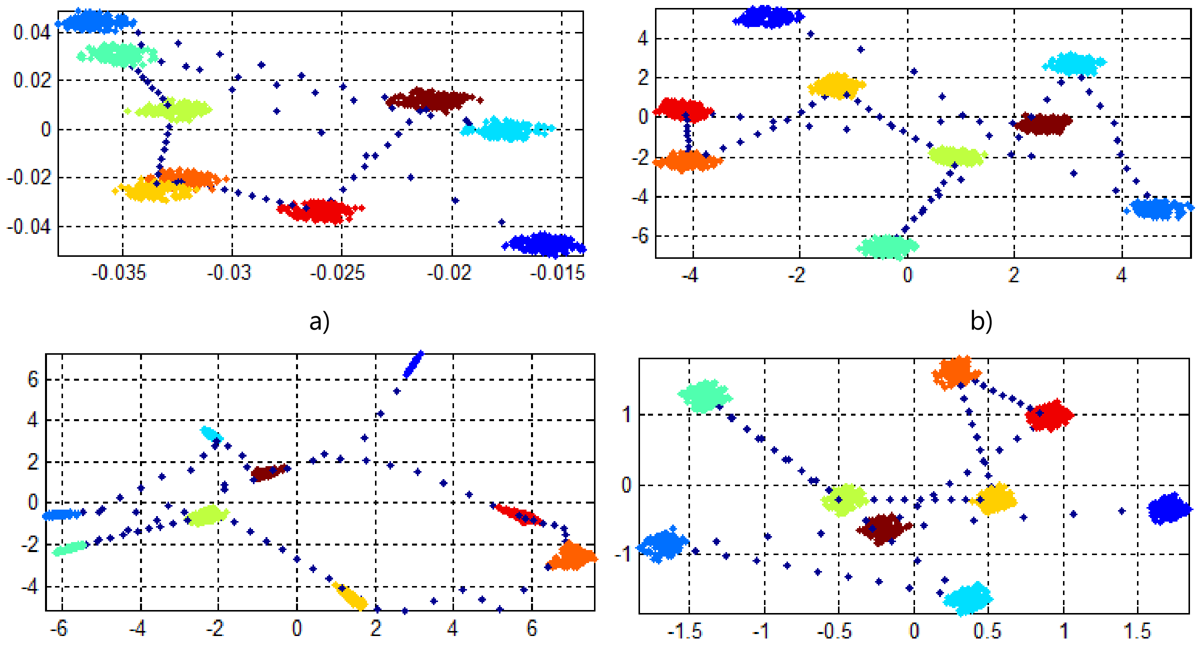
$$Y = \{\lambda_2 v_2, \lambda_3 v_3, \ldots, \lambda_{d+1} v_{d+1}\} \qquad (8)$$

The result of dimension reduction using a Diffusion Map method is shown in Fig. 2,d.

The most correct display of the clustering results of (9 clusters, 1150 points), mineralogical and technological characteristics of ore varieties in the space of lower dimension was obtained by the methods of MDS, Diffusion Map, and PCA. Thus, the least time was spent using the methods of PCA –

ISSN 2414-9055

0.11 s, and Diffusion Map – 0.77 s. The greatest – 47, 67 s using Isomap method.



a) b) c) d)

Figure 2. The result of dimension reduction using the methods of:a) PCA; b) MDS; c) Isomap; d) Diffusion Map

Conclusions. Thus, among the considered methods of the dimension reducing the best results were obtained using PCA, MDS and Diffusion Map methods.

References

1. Tikhonov O.N. Zakonomernosti effektivnogo razdeleniya mineralov v protsessakh obogashcheniya poleznykh iskopayemykh [*Laws of effective separation of minerals during mineral processing*]. Moscow: Nedra, 1984.

2. Lynch A.J. Tsikly drobleniya i izmelcheniya [*Cycles of crushing and grinding*], Moscow: Nedra, 1981.

3. Morkun V. S., Morkun N. V., Pikilnyak A.V. (2014). Iron ore flotation process control and optimization using high-energy ultrasound, Metallurgical and Mining Industry, **2**: 36-42.

4. Morkun V., Tron V., Goncharov S. (2015) Automation of the ore varieties recognition process in the technological process streams based on the dynamic effects of high-energy ultrasound, Metallurgical and Mining Industry, **2**: 31 34.

5. Morkun V., Morkun N., Pikilnyak A. (2014) Ultrasonic facilities for the ground materials characteristics control, Metallurgical and Mining Industry, **2**: 31-35..

6. Morkun V., Morkun N., Pikilnyak A. (2015). The study of volume ultrasonic waves propagation in the gas-containing iron ore pulp, Ultrasonics, **56C**: 340-343.

7. Morkun V., Morkun N., Pikilnyak A. (2014). Simulation of the Lamb waves propagation on the plate which contacts with gas containing iron ore pulp in Waveform Revealer toolbox. Metallurgical and Mining Industry, **5**: 16-19.

8. L.J.P. van der Maaten. An Introduction to Dimensionality Reduction Using Matlab, Report MICC 07-07, MICC/IKAT, Maastricht: Universiteit Maastricht, 2007.

9. Cox T., Cox M. Multidimensional scaling. London: Chapman & Hall, 1994.

10. Kruskal J.B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, **29**: 1–27.

11. Tenenbaum J.B. (1998). Mapping a manifold of perceptual observations. In Advances in Neural Information Processing Systems, *The MIT Press*, **10**: 682–688.

12. Dijkstra E.W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, **1**: 269–271.

13. Floyd R.W. (1962). Algorithm 97: Shortest path. *Communications of the ACM*, **5**(6): 345,.

14. Lafon S., Lee A.B. (2006). Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **28**(9): 1393–1403.

15. Nadler B., Lafon S., Coifman R.R., Kevrekidis I.G. (2006). Diffusion maps, spectral clustering and the reaction coordinates of dynamical systems. *Applied and Computational Harmonic Analysis: Special Issue on Diffusion Maps and Wavelets*, **21**: 113–127.