Investigation of methods of fuzzy clustering for determining ore types

Vladimir Morkun

*Vice-Rector for research, Doctor of Science, Professor of Computer Science, Automation and Control Systems department
State Higher Educational Institution "Kryvyi Rih National University", Ukraine*

Sergij Tcvirkun

*Automation and Control Systems department
State Higher Educational Institution "Kryvyi Rih National University", Ukraine*

Abstract
The article deals with cluster analysis for determining the types of lump ore on the line of conveyer belt for further separation of lumps rich in iron.
Key words: IRON ORE RAW MATERIALS, CLUSTER, CLASSIFICATION ALGORITHM, FUZZY CLUSTERING.

To provide high technical-and-economic indexes of extraction and mineral processing it is necessary to control the quality on the planning, extracting and ore transportation stages. As the belonging of ore material to certain type may be determined under several characteristics, clusterization is reasonable to use for this operation.

Most of clustering algorithms do not bear on the traditional for statistical methods suppositions; they may be used in conditions of almost full absence of information about data distribution law [1].

Initial data for clustering is measurement matrix of indirect indicator of technological sorts of iron, which consists of *n* lines, each of which contains characteristic values of individual sample:

$$X = \begin{bmatrix} X_{11} & X_{12} & X_{1n} \\ X_{21} & X_{22} & X_{2n} \\ \dots & & \\ X_{M1} & X_{M2} & X_{Mn} \end{bmatrix} \qquad (1)$$

where *n* is the amount of characteristics; *M* – amount of ore samples. In this case the task is to divide samples of ore, which is represented by several technological types, on several clusters, where the similarity of samples allows to distinguish certain technological variety.

For clustering of ore samples characteristics let us consider the methods of fuzzy clustering.

In real situation the division of ore material on the technological types is difficult to present by two membership degrees 0 or 1. More natural is to use partial belonging within the range from 0 to 1, which will allow the ore samples, characteristics of which are on the borders between several clusters, to belong to them with different degree [1].

Fuzzy C-means clustering algorithm, which was used for clustering of characteristics of ore samples, is based on the minimization of C-means functional [1,3].

$$J(X;U,V) = \sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^m \|x_k - v_i\|_A^2 \qquad (2)$$

vector of a moment center of clusters

$$D_{ikA}^2 = \|x_k - v_i\|_A^2 = (x_k - v_i)^T A(x_k - v_i) \qquad (3)$$

The results of clustering of iron samples, fulfilled under Fuzzy C-means algorithm, fig. 1, showed true determination of cluster centers, as compared with pattern; this allows drawing conclusion about use perspectiveness of this method.

Also for classification of characteristics of iron ore raw materials there was used Gustafson-Kessel algorithm, which improves Fuzzy C-means algorithm using adoptive norm of distance [4] for each cluster

$$D_{ikA}^2 = (x_k - v_i)^T A_i (x_k - v_i), \qquad 1 \le i \le c, \quad 1 \le k \le N.$$

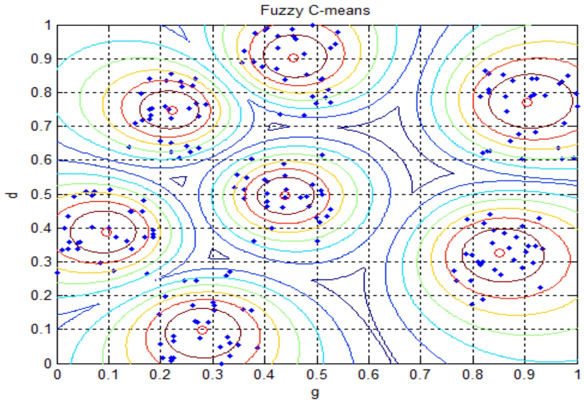

Figure 1. Result of Fuzzy C-means clustering of iron materials samples

Objective function of an algorithm is determined as follows:

$$J(X;U,V,A) = \sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^2 D_{ikA_i}^2 \qquad (4)$$

The result of iron samples clustering for determination the types of ores, fulfilled under Gustafson-Kessel algorithm is shown in the figure 2.

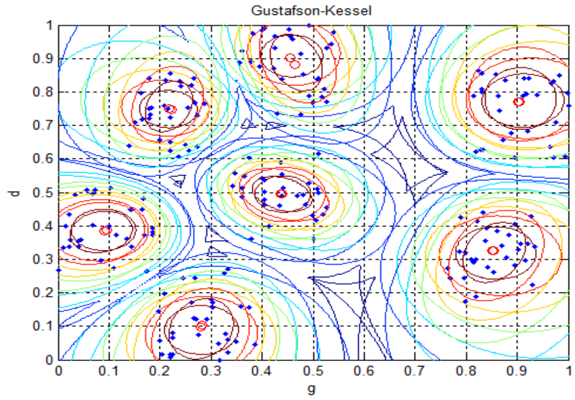

Figure 2. The result of Gustafson-Kessel clustering of iron ore samples.

In furtherance of the work [4] Gath and Geva in the work [5] formed the new distance function on the base of indistinct maximum likelihood estimate

$$D_{ik}(x_k, v_i) = \frac{\sqrt{\det(F_{\omega i})}}{\alpha_i}\exp\left(\frac{1}{2}(x_k - v_i^{(l)})^T F_{\omega i}^{-1}(x_k - v_i^{(l)})\right)$$

(5)

The result of iron ore samples clustering fulfilled under Gath – Geva algorithm is given in the figure 3.

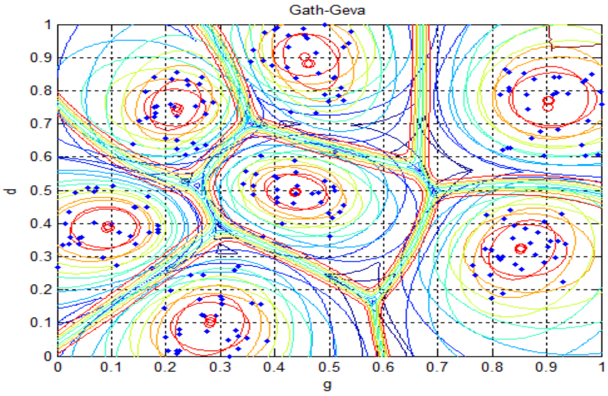

Figure 3. The result of Gath – Geva clustering of iron material samples

Estimation of clustering quality was fulfilled with the usage of the following scalar measures of certainty.

Partition coeficcient (PC): measures the value of “covering” between clusters [3]

$$PC(c) = \frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{N}(\mu_{ij})^2 \qquad (6)$$

where μ_{ij} – membership function of data point j in the cluster i. Optimal amount of clusters corresponds to maximum value of functional.

Classification entropy (CE) is the measure of indistinct distribution

$$CE(c) = -\frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{N}\mu_{ij}\log(\mu_{ij}), \qquad (7)$$

Exponent of distribution (ED) represents the ratio of compaction sum to division of clusters [6].

$$ED(c) = \sum_{i=1}^{c}\frac{\sum_{j=1}^{N}(\mu_{ij})^m\left\|x_j - \upsilon_i\right\|^2}{N_i\sum_{k=1}^{c}(\mu_{ij})^m\left\|\upsilon_k - \upsilon_i\right\|^2} \qquad (8)$$

The smaller value ED corresponds to the better result of clustering.

Selectivity index (S) [6].

$$S(c) = \frac{\sum_{i=1}^{c}\sum_{j=1}^{N}(\mu_{ij})^2\left\|x_j - \upsilon_i\right\|^2}{N\min_{i,k}\left\|\upsilon_k - \upsilon_i\right\|^2} \qquad (9)$$

Index Xie-Beni (XB) determines the quantitative evaluation of correlation of total variation in clusters and division of clusters [7]

$$XB(c) = \frac{\sum_{i=1}^{c}\sum_{j=1}^{N}(\mu_{ij})^m\left\|x_j - \upsilon_i\right\|^2}{N\min_{i,j}\left\|x_j - \upsilon_i\right\|^2} \qquad (10)$$

Minimum value of the index corresponds to the optimal amount of clusters for determining ore types.

Dunn's Index (DI) is used to reveal “compact and well-divided clusters” [2]

$$DI(c) = \min_{i\in c}\{\min_{i\in c, i\neq j}\{\frac{\min_{x\in C_i, y\in C_j} d(x,y)}{\max_{k\in c}\{\max_{x,y\in C} d(x,y)\}}\}\} \qquad (11)$$

Alternative Dunn Index (ADI)

$$ADI(c) = \min_{i\in c}\{\min_{j\in c, i\neq j}\{\frac{\min_{x_i\in C_i, x_j\in C_j}\left|d(y,\upsilon_j) - d(x,\upsilon_j)\right|}{\max_{k\in c}\{\max_{x,y\in C} d(x,,y)\}}\}\} \qquad (12)$$

The results of comparison of clustering methods according to the considered above indexes for determining the types of ore are given in the table 1.

Table 1 Results of comparison of clusterization methods

	PC	CE	ED	S	XB	DI	ADI
FCM	0.6984	0.6996	0.5178	0.0044	4.5559	0.2443	0.0003
GK	0.7111	0.6681	0.5034	0.0043	5.5842	0.2443	0,0001
GG	0.9940	0.0118	1.1264	0.0098	1.7113	0.2443	0.0090

Gustafson-Kessel and Gath – Geva methods have the best indexes. However, disadvantage of Gath – Geva algorithm is the necessity in preprocessing of initial data by means of clustering with usage, for example, fuzzy clustering method FCM. In such a way the most effective here is the usage of Gustafson-Kessel fuzzy clustering method.

References

1. Shtovba S.D. *Vvedeniye v teoriyu nechetkikh mnozhestv i nechetkuyu logiku*. [Introduction to fuzzy-set theory and fuzzy

logic]. Available at- HTTP : MATLAB / exponenta.ru / fuzzylogic / book1.
2. Balasko B., Abonyi J., Feil B. Fuzzy Clustering and Data Analysis, Toolbox, 74 p.
3. Bezdek J. C. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, 1981.
4. Gustafson D.E., Kessel W.C. Fuzzy clustering with a fuzzy covariance matrix. Proc. IEEE CDC, San Diego, CA. 1979. P. 761 - 766. No 7., P. 773 - 781.
5. Gath I., Geva A.B. Unsupervised optimal fuzzy clastering. IEEE Trans. Pattern Anal. Machine Intell., 1989. No. 7., P. 773 - 781.
6. Bensaid A.M., Hall L.O., Bezdek J.C., Clarke L.P., Silbiger M.L., Arrington J.A., Murtagh R.F. Validity-guided (Re)Clustering with applications to imige segmentation. *IEEE Transactions on Fuzzy Systems*, 4, 112-123, 1996.
7. Xie X. L. and Beni G. A. Validity measure for fuzzy clustering. *IEEE Trans. PAMI*, No3(8), 841-846, 1991.