

УДК 622.723: 004.275

М.П. ТИХАНСКИЙ, канд. техн. наук, доц., С.Л. ЦВИРКУН, преподаватель
Криворожский национальный университет

ОПРЕДЕЛЕНИЕ ОПТИМАЛЬНОГО КОЛИЧЕСТВА КЛАСТЕРОВ ПРИ ИСПОЛЬЗОВАНИИ МЕТОДА ГУСТАФСОНА-КЕССЕЛЯ ДЛЯ ОПРЕДЕЛЕНИЯ ТЕХНОЛОГИЧЕСКИХ РАЗНОВИДНОСТЕЙ ЖЕЛЕЗНОЙ РУДЫ

Рассмотрена задача кластерного анализа в условиях неопределенности для формирования автоматического управления процессом сортировки крупнокускового железорудного сырья с оперативным распознаванием его минералого-технологических разновидностей отдельных кусков в потоке на конвейерной ленте учитывая физико-механические свойства. Кластер-анализ получил широкое применение в различных отраслях, где имеются множества объектов произвольной природы, которые необходимо автоматически разбить на кластера. Все алгоритмы кластеризации можно подразделить на две группы: алгоритмы с заданным количеством кластеров и алгоритмы, которые сами устанавливают оптимальное количество кластеров. Наиболее предпочтительными алгоритмами кластер-анализа являются алгоритмы, которые сами устанавливают количество кластеров или алгоритмы самоорганизации, одним из которых является алгоритм Густавсона-Кесселя. Оценка качества кластеризации была проведена с использованием скалярных мер достоверности. Проведены экспериментальные исследования показателей качества кластеризации от количества кластеров для определения минералого-технологических разновидностей железной руды. Было установлено по результатам проведенных исследований, что для формирования автоматического управления сортировкой кусковой руды с оперативным распознаванием минералого-технологических разновидностей отдельных кусков в потоке, с помощью нечеткой кластеризации её характеристик, оптимальным количеством кластеров, является 7, при этом необходимо 29 итераций.

Ключевые слова: нечеткая кластеризация, кластер, метод Густафсона-Кесселя, рудоподготовка, сортировка руды.

Проблема и ее связь с практическими задачами. Необходимость снижения себестоимости и повышения технико-экономических показателей обогащения магнетитовой руды ставит задачи совершенствования технологии и методов управления обогащением в целом и рудоподготовкой в частности. На данный момент отечественные горнообогатительные комбинаты перерабатывают в среднем 5-8 технологических разновидностей руды со значительным различием значений характеристик в каждой [1,2]. В таких условиях одним из наиболее перспективных технологических и экономических производственных резервов является снижение затрат на транспортировку горной массы и энергетических затрат на переработку отдельных технологических разновидностей руды, в частности относящихся к «доменной» руде, которая после добычи может быть направлена в металлургический передел без переработки на обогатительной фабрике. Несмотря на значительное количество выполненных работ, задача формирования достаточно эффективного автоматического управления процессом предобогащения железных руд в условиях подземной добычи и обогатительных фабрик не получила достаточно полного решения. Следовательно, вопросы автоматизации процессов управления сортировкой кускового рудного материала с учетом физико-механических свойств технологических разновидностей железных руд являются важными, актуальными и требуют проведения дальнейших исследований.

Анализ исследований и публикаций. Для формирования автоматического управления процессом сортировки крупнокускового железорудного сырья с оперативным распознаванием его технологических разновидностей отдельных кусков в потоке целесообразно использовать операцию кластеризации. Кластер-анализ, или автоматическая классификация получили широкое применение в различных отраслях, всюду где имеются множества объектов произвольной природы, описываемых в виде векторов $x = \{x_1, x_2, \dots, x_N\}$, которые необходимо автоматически разбить на группы “схожих” объектов, называемых кластерами.

Большинство алгоритмов кластеризации не опираются на традиционные для статистических методов допущения. Они могут использоваться в условиях почти полного отсутствия информации о законах распределения данных. Кластеризацию проводят для объектов с количественными, качественными или смешанными признаками.

Существует множество методов кластеризации, которые можно классифицировать на четкие и нечеткие. Четкие методы кластеризации разбивают исходное множество объектов X на несколько непересекающихся подмножеств. При этом любой объект из X принадлежит только одному кластеру.

Нечеткие методы кластеризации позволяют одному и тому же объекту принадлежать одновременно нескольким (или даже всем) кластерам, но с различной степенью.

Нечеткая кластеризация во многих ситуациях более "естественна", чем четкая, например, для объектов, расположенных на границе кластеров [3,4].

Все известные алгоритмы кластеризации можно подразделить на две основные группы: алгоритмы с априорно заданным количеством кластеров и алгоритмы, сами устанавливающие оптимальное количество кластеров.

Если количество кластеров неизвестно, то наиболее предпочтительными алгоритмами являются алгоритмы, сами устанавливающие количество кластеров или алгоритмы самоорганизации, одним из которых является алгоритм Густавссона-Кесселя [5].

Цель исследования. При распознавании технологических разновидностей железорудного сырья были проведены исследования с целью определения оптимального количества кластеров при использовании метода Густафсона-Кесселя.

Изложение материала и результаты. Кластеризация производилась по следующему алгоритму [6,7]. В процессе оптимизации были приняты следующие значения параметров: весовой показатель $m=2$, значение допуска для останова вычислений $\varepsilon=0,001$, определитель для каждого кластера $\rho=1$.

После инициализации матрицы принадлежности случайными значениями $U^{(0)} \in M_{fc}$ на каждом шаге $l=1,2,\dots$ выполняем следующие шаги.

Вычисляем центры кластеров

$$v_i^{(l)} = \frac{\sum_{k=1}^N (\mu_{ik}^{(l-1)})^{m_{xk}}}{\sum_{k=1}^N (\mu_{ik}^{(l-1)})^m}, 1 \leq i \leq c. \quad (1)$$

Вычисляем ковариационные матрицы кластеров

$$F_i^{(l)} = \frac{\sum_{k=1}^N (\mu_{ik}^{(l-1)})^m (x_k - v_i^{(l)})(x_k - v_i^{(l)})^T}{\sum_{k=1}^N (\mu_{ik}^{(l-1)})^m}, 1 \leq i \leq c. \quad (2)$$

Добавляем масштабированную единичную матрицу

$$F_i := (1-\gamma)F_i + \gamma(F_0)^{1/n} I. \quad (3)$$

Определяем собственные значения λ_{ij} и собственные вектора ϕ_{ij} , определяем $\lambda_{i,\max} = \max_j \lambda_{ij}$. Восстанавливаем значение F_i

$$F_i = [\phi_{i,1} \dots \phi_{i,n}] \text{diag}(\lambda_{i,1} \dots \lambda_{i,n}) [\phi_{i,1} \dots \phi_{i,n}]^{-1}. \quad (4)$$

Вычисляем расстояния

$$D_{ikA}^2(x_k, y_i) = (x_k - v_i^{(l)})^T \left[(\rho_i \det(F_i))^{1/n} F_i^{-1} \right] (x_k - v_i^{(l)}). \quad (5)$$

Рассчитываем значения элементов матрицы принадлежности по формуле

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^c (D_{ikA}(x_k, v_i) D_{jk}(x_k, v_j))^{2/(m-1)}}, 1 \leq i \leq c, 1 \leq k \leq N. \quad (6)$$

До тех пор пока не будет выполнено условие останова расчетов $\|U^{(l)} - U^{(l-1)}\| < \varepsilon$.

В процессе оптимизации были приняты следующие значения параметров: $m=2$, $\varepsilon=0,001$, $\rho=1$ для каждого кластера. В процессе исследования количество кластеров с варьировалось от 2 до 14.

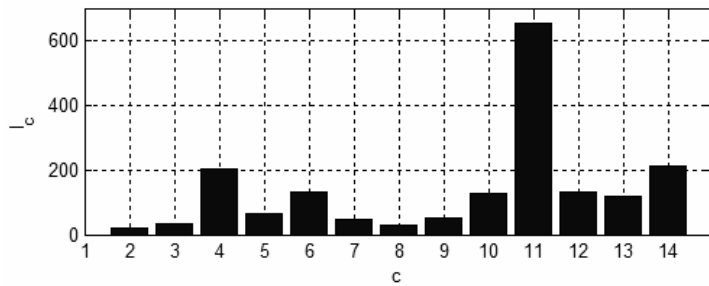


Рис. 1. Количество итераций при различном количестве кластеров

Наименьшее количество итераций потребовалось при разделении характеристик проб железорудного сырья на 2 кластера - 20, наибольшее - 654 при разделении на 11 (рис. 1).

Результаты кластеризации ха-

рактеристик руды кусковой руды приведены на рис. 2.

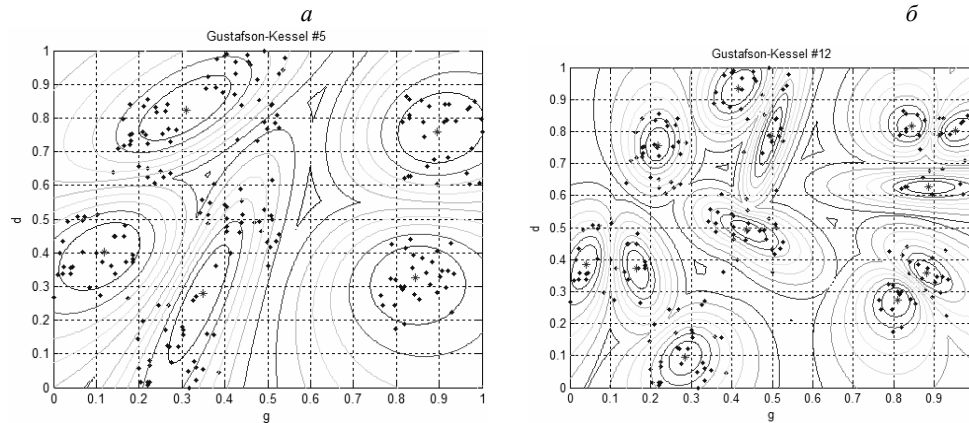


Рис. 2. Результаты кластеризации характеристик кусков руды при различном количестве кластеров: *a* - 5 кластеров; *б* - 12 кластеров

Оценка качества кластеризации была проведена с использованием скалярных мер достоверности [7].

Графическое представление зависимостей показателей качества кластеризации от количества кластеров по алгоритму Густафсона-Кесселя приведено на рис. 3-9.

Коэффициент распределения РС, зависимость которого от количества кластеров показана на рис. 3, является недостаточно информативным [4].

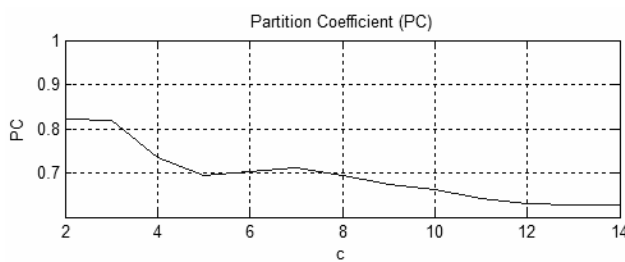


Рис. 3. Коэффициент распределения (PC)

При этом, следует отметить локальный всплеск при количестве кластеров $c=7$.

Слабую зависимость от анализируемых данных имеет показатель энтропии классификации CE: с увеличением количества кластеров значение функции монотонно возрастает (рис. 4).

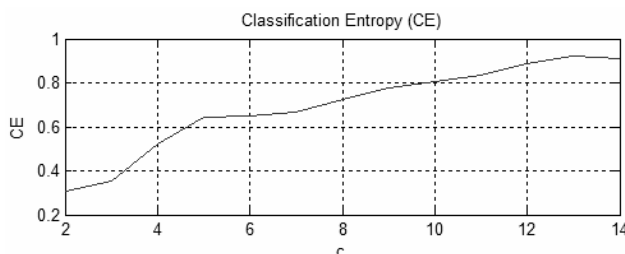


Рис. 4. Энтропия классификации (CE)

Показатель распределения SC сравнительно быстро убывает при увеличении количества кластеров с 2 до 7 (рис. 5), после чего убывание существенно замедляется [8].

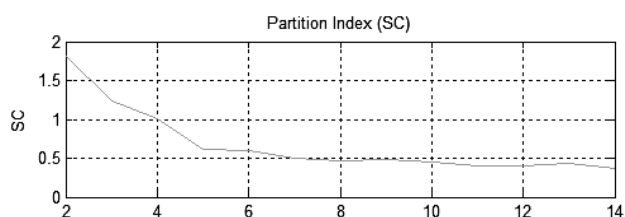


Рис. 5. Показатель распределения (SC)

Зависимость показателя разделения S от количества кластеров приведена на рис. 6 [8].

Здесь наблюдаются существенные колебания показателя при увеличении количе-

ства кластеров с 2 до 7 и сравнительно медленный спад при дальнейшем увеличении.

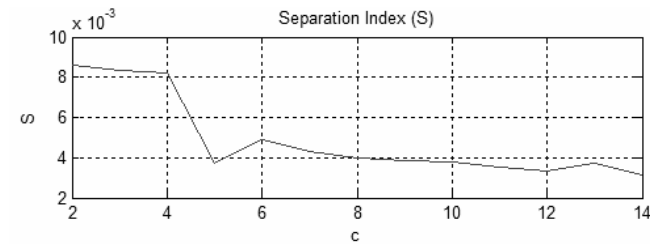


Рис. 6. Показатель разделения (S)

Показатель Кси-Бени XB изменяется достаточно сильно (рис. 7), что не позволяет однозначно определить оптимальное количество кластеров [9].

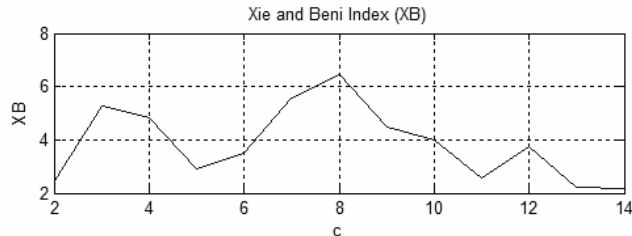


Рис. 7. Показатель Кси-Бени

Показатель Данна DI (рис. 8) указывает на то, что оптимальным количеством кластеров является $c=7$ [6].

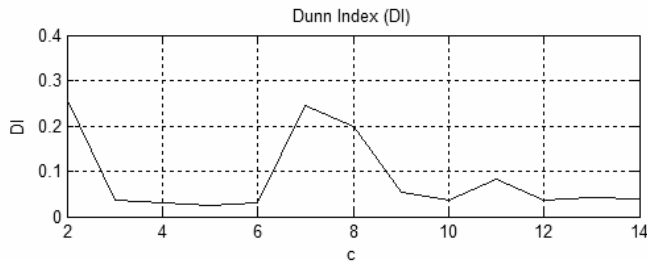


Рис. 8. Показатель Данна

Значение альтернативного показателя. Данная ADI, зависимость которого от количества кластеров приведена на рис. 9, достаточно сильно убывает с увеличением количества кластеров до 7, а с последующим увеличением практически не изменяется.

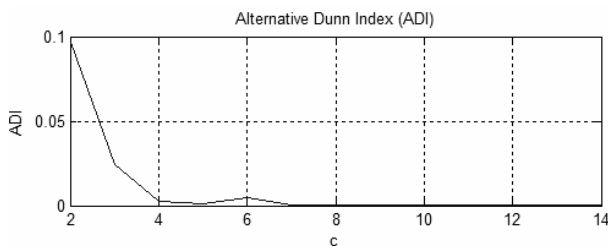


Рис. 9. Альтернативный показатель Данна

Результаты проведенных исследований показали, что оптимальным количеством кластеров при определении технологических разновидностей железорудного сырья является $c=7$, при этом необходимо 29 итераций.

Результаты кластеризации с помощью алгоритма Густафсона-Кесселя при количестве кластеров от пяти до девяти приведены на рис. 10.

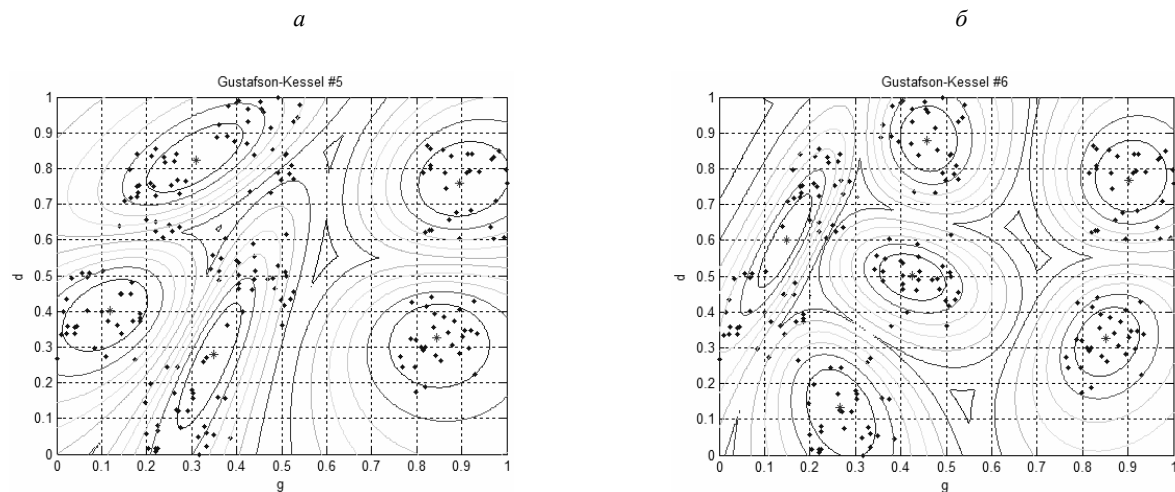


Рис. 10. Результаты кластеризации характеристик кусков руды при различном количестве кластеров

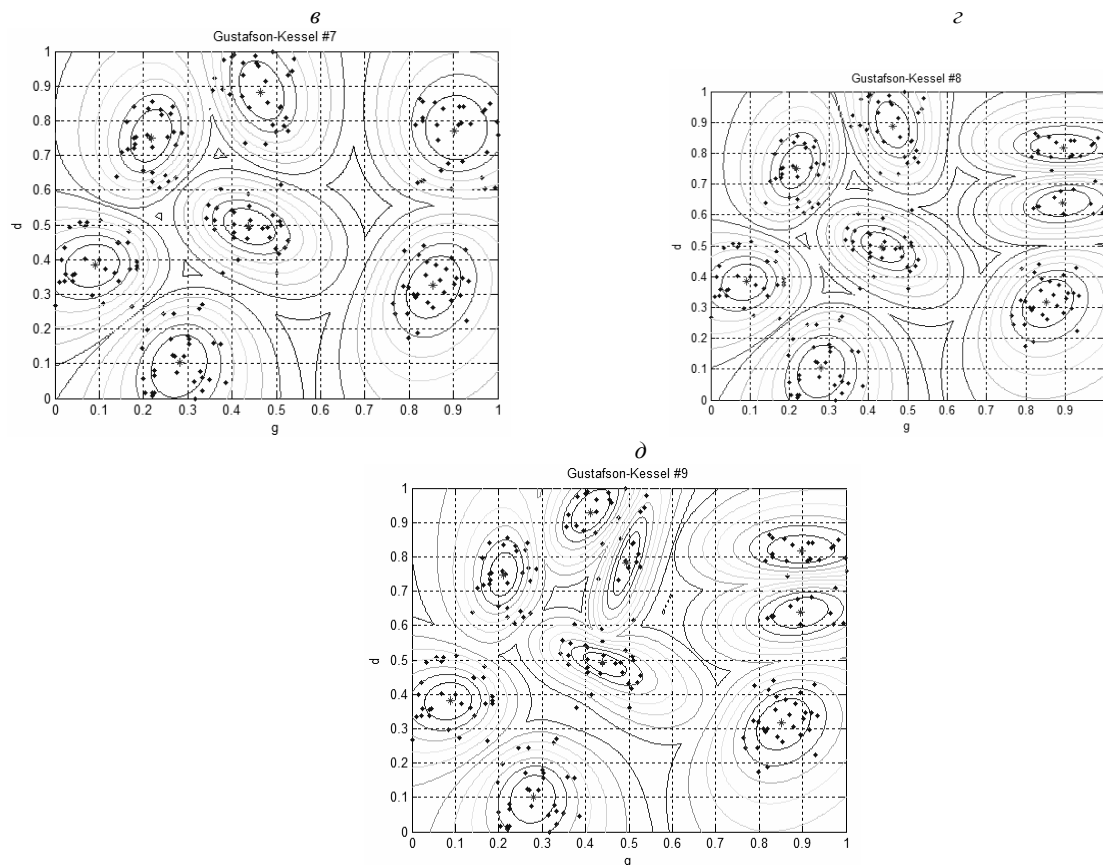


Рис. 10. Результати кластеризації характеристик кусків руди при різному кількості кластерів (Продолжение рис. 10)

Выводы. Таким образом, формирование автоматического управления сортировкой кусковой руды с оперативным распознаванием минералого-технологических разновидностей отдельных кусков в потоке, осуществлялось с помощью нечеткой кластеризации её характеристик с использованием алгоритма Густафсона-Кесселя.

Результаты проведенных исследований показали, что оптимальным количеством кластеров при определении технологических разновидностей железорудного сырья, гематитовых разновидностей руд месторождений Кривбасса, является $c=7$, при этом необходимо 29 итераций вычислений с заданной точностью.

Список литературы

1. Губіна В. Г. Проблема залізвмісних відходів гірничо-металургійного комплексу України - системний підхід / В. Г. Губіна, Б. О. Горлицький // Зб. наук. праць Інституту геохімії навколишнього середовища. - 2009. - Вип. 17. - С. 79-92.
2. Губін Г. Г. Гірничо-металургійний комплекс України між кризами / Г. Г. Губін, А. Г. Губіна // Вісник КТУ. - 2010. - Вип. 25. - С. 218-224.
3. Штовба С.Д. Введение в теорию нечетких множеств и нечеткую логику / С.Д. Штовба. - Режим доступа: <http://matlab/exponenta.ru/fuzzylogic/book1>.
4. Bezdek J. C. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, 1981.
5. Gustafson D.E., Kessel W.C. Fuzzy clustering with a fuzzy covariance matrix // Proc. IEEE CDC, San Diego, CA. - 1979. - P. 761-766. - Vol. 7. - P. 773-781.
6. Balasko B. Fuzzy Clustering and Data Analysis Toolbox / Balasko B., Abonyi J., Feil B. - 74 p.
7. R. Babuka, P.J. van der Veen, and U. Kaymak. Improved covariance estimation for Gustafson-Kessel clustering. In Fuzzy Systems, 2002. FUZZ-IEEE'02. Proceedings of the 2002 IEEE International Conference on, volume 2, pages 1081 – 1085, 2002.
8. Bensaïd 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 image segmentation. *IEEE Transactions on Fuzzy Systems*, 4:112-123, 1996.
9. Xie X. L. and Beni G. A. Validity measure for fuzzy clustering. *IEEE Trans. PAMI*, 3(8):841 {846, 1991.
10. Morkun V., Tevirkun S. Investigation of methods of fuzzy clustering for determining ore types. Metallurgical and Mining Industry, 2014, No.5, p.p. 12-15. <http://www.metaljournal.com.ua/assets/Journal/3-MorkunTs.pdf>.

Рукопись поступила в редакцию 28.03.15