

УДК 622.013.364:553.94

ОБЩИЕ ПОДХОДЫ К ГЕОМЕТРИЗАЦИИ МАРОЧНОГО СОСТАВА УГЛЯ НА ОСНОВЕ МЕТОДА КЛАСТЕРНОГО АНАЛИЗА

© 2008 И.В. Филатова

Донецкий национальный технический университет

Решение задачи геометризации марочного состава углей Донбасса рассматривается на основе применения метода кластерного анализа, который представляет математическую процедуру многомерного анализа, позволяющего на основе множества показателей, характеризующих ряд объектов, сгруппировать их в классы (кластеры).

The solution of Donbassian coals grade compositions' geometrization task is examined, basing on cluster analysis method. This kind of analysis is multivariate analyzing mathematical procedure, which gives an opportunity to group a number of objects into classes (clusters), in accordance with a set of activities.

Ключевые слова: геометризация марочного состава угля, метод кластерного анализа.

Геометризация качественных показателей является одной из задач геолого-маркшейдерской службы предприятий, на основании решений которой, становится возможным в наглядной форме охарактеризовать закономерности распределения классификационных показателей марок угля и степень их изменчивости. При этом решается вопрос выделения однородных площадей и построения границ марочного состава углей.

Построение граничных линий распространения углей различных по марочному составу, как правило, производится со значительной степенью условности, без совместного учета всех классификационных

показателей марки. Промышленная маркировка углей основывается на совокупности свойств, которые характеризуют степень метаморфизма и вещественный состав. С 1996 года (ДСТУ 3472-96) в соответствии с единой классификацией углей Украины [1] по степени метаморфизма и технологическим свойствам основными классификационными параметрами марки угля являются средний показатель отражения витринита R_p , выход летучих веществ V^{daf} , толщина пластического слоя u , индекс Рора RI и теплота сгорания Q_s^{daf} . Согласно [1] выделяют девять марок углей: бурые (Б), длиннопламенные (Д), длиннопламенные газовые (ДГ), газовые (Г), жирные (Ж), коксовые

(К), отошенные спекающиеся (ОС), тощие (Т) и антрациты (А).

Решение задачи геометризации марочного состава углей Донбасса рассмотрим на основе применения метода кластерного анализа, который представляет математическую процедуру многомерного анализа, позволяющего на основе множества показателей, характеризующих ряд объектов, сгруппировать их в классы (кластеры). Группирование производится таким образом, что объекты одного класса являются однородными и подобными в сравнении с объектами другого класса. На основании этого совокупность качественных показателей, которые принадлежат к разным маркам угля, разделяются на группы в соответствии с задачей кластеризации. При этом качественные показатели описываются численными образами и вычисляются расстояния между ними, которые выражаются в евклидовой метрике или в других метриках. Достоинство такого анализа заключается в том, что разбивка объектов производится не по одному параметру, а по набору признаков.

Рассмотрим общие подходы геометризации марочного состава угля на основе метода кластерного анализа

На первом этапе выполняется предварительное обнаружение кластеров и оценивание их числа, основанное на исследовании поведения вариационного ряда расстояний между различными точками [1-5].

Формирование основного вариационного ряда (ОВР) и его исследование. Исследуемое множество X представлено скважина-

ми, по которым определены показатели марочного состава:

$$X = \{x_1, x_2, \dots, x_n\},$$

заданных p числовыми признаками (например, R_o, V^{daf}, y, RI и Q_s^{daf})

$$x_j = (x_{j1}, x_{j2}, \dots, x_{jp}), p=3, j=1, 2, \dots, n.$$

Значение каждого показателя нормируется (относительно среднего значения) по парам признаков выходу летучих веществ V^{daf} и толщине пластического слоя y ; выходу летучих веществ V^{daf} и среднему показателю отражения витринита R_o ; толщине пластического слоя y и среднему показателю отражения витринита R_o и так по всем парам признаков.

За меру сходства двух элементов принимается евклидово расстояние

$$r_j = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

Элементами x_i и y_i являются пары нормированных классификационных показателей, характеризующие марочный состав угля (например, выход летучих веществ V^{daf} и толщина пластического слоя y ; выход летучих веществ V^{daf} и средний показатель отражения витринита R_o ; толщина пластического слоя y и средний показатель отражения витринита R_o). По результатам расчетов строится основной вариационный ряд (ОВР):

$$r_{(1)} < r_{(2)} < \dots < r_{(s)}, s = n \cdot (n - 1) / 2.$$

Структура исследуемых множеств отражается в поведении вариационных рядов, для которых строятся гистограммы, по которым оцениваются локальные минимумы

вариационных рядов. И, следовательно, делается вывод о неоднородности или однородности представленного множества и наличии нескольких кластеров данных.

Для установления статистической значимости локального минимума в интервале $[r_{m_o}, r_{m_o+1})$ определяются ближайшие локальные максимумы в интервалах $[r_{i_o}, r_{i_o+1})$ и $[r_{t_o}, r_{t_o+1})$. Выполняется проверка гипотезы H_o о постоянстве плотности вероятности в промежутке $[r_{m_o}, r_{t_o+1}) \supset [r_{m_o}, r_{m_o+1})$. Подсчитывается число наблюдений \tilde{n} , попавших в интервал $[r_{m_o}, r_{t_o+1})$ и число равных промежутков гистограммы \tilde{s} , содержащихся в $[r_{m_o}, r_{t_o+1})$. Проверка гипотезы H_o проводится по критерию согласия $\tilde{\chi}^2$ и сводится к проверке того, что из \tilde{s} интервалов число точек равно \tilde{n} / \tilde{s} , для чего вычисляется статистика:

$$\tilde{\chi}^2 = \sum_j \frac{(v_j - \tilde{n}/\tilde{s})^2}{\tilde{n}/\tilde{s}}$$

где v_j – число наблюдений, попавших в j -промежуток интервала $[r_{m_o}, r_{t_o+1})$;

$$\tilde{n} = \sum_j v_j$$

Фиксируется уровень значимости α и по таблицам распределений χ^2 с $\tilde{s} - 1$ степенью свободы находится значение χ_α^2 . Если $\tilde{\chi} \geq \chi_\alpha^2$, то гипотеза H_o отвергается и рассматриваемый локальный минимум гистограммы считается статистически значимым.

Если $\tilde{\chi} < \chi_\alpha^2$, то гипотеза H_o не отвергается и локальный минимум статистически не значим. Следовательно, множество X или однородно, или состоит из кластеров, пере-

секающихся или близко расположенных друг к другу. Для обнаружения кластеров в этом случае применяются другие методы. Некоторую информацию о структуре множества X дают дополнительные вариационные ряды – минимальных и максимальных расстояний. Для полученных евклидовых расстояний формируются два множества \tilde{R}_{\min} и \tilde{R}_{\max} :

$$\tilde{R}_{\min} = \{r_{\min}^1, r_{\min}^2, \dots, r_{\min}^n\},$$

$$\tilde{R}_{\max} = \{r_{\max}^1, r_{\max}^2, \dots, r_{\max}^n\}.$$

Удалив из полученных множеств \tilde{R}_{\min} и \tilde{R}_{\max} равные элементы и расположив оставшиеся в возрастающем порядке, получают новые множества – вариационные ряды:

$$R_{\min} = \{r_{\min}^{(1)}, r_{\min}^{(2)}, \dots, r_{\min}^{(u)}\}, u \leq n-1,$$

$$R_{\max} = \{r_{\max}^{(1)}, r_{\max}^{(2)}, \dots, r_{\max}^{(l)}\}, l \leq n-1,$$

элементы, которых соответствуют различным парам точек.

Если построенные вариационные ряды R_{\min} имеют локальные минимумы, число которых равно \tilde{m}_1 , то множество X состоит из $k \geq \tilde{m}_1 + 1$ кластеров.

Для вариационных рядов R_{\max} число кластеров равно $k \geq \tilde{m}_2 + 1$, где \tilde{m}_2 – число локальных минимумов [4].

Выполнив предварительное обнаружение кластеров, оценив их число, выполняется непосредственно процедура кластерного анализа. Второй этап – распределение запасов по маркам рассматривается на основе использования программ математической статистики с применением методов иерархического кластерного анализа.

Смысл иерархической процедуры объединения в кластеры заключается в следующем: перед началом кластеризации все объекты считаются отдельными кластерами, т.е. имеется $p = n$ кластеров, каждый из которых включает по одному элементу. На первом шаге алгоритма определяются два наиболее близких или сходных объекта, которые объединяются в один кластер, общее количество которых сокращается на 1 ($p \rightarrow p - 1$). Итеративный процесс повторяется, пока на последнем ($p - 1$)-м шаге все классы не объединятся. На каждом последующем шаге процедуры объединения рассчитываются расстояния от образованного кластера до каждого из оставшихся кластеров.

Для определения, какое количество кластеров следует считать оптимальным, служит показатель, выводимый под заголовком «коэффициент». Под коэффициентом подразумевается расстояние между двумя кластерами, определенное на основании выбранной дистанционной меры (евклидово расстояние). На этапе, где мера расстояния между двумя кластерами увеличивается скачкообразно, процесс объединения в новые кластеры необходимо остановить, так как в противном случае были бы объединены уже кластеры, находящиеся на относительно большом расстоянии друг от друга.

Результатом кластерного анализа является получение зоны перекрытия кластеров (марок угля), которая характеризует часть площади пласта, в которой размещаются перекрытия границ, полученные в результате анализа данных опробования статистическими методами. Данная зона пред-

ставляет своеобразный доверительный интервал между марками, в котором располагается граница марочного состава.

Достоинством метода кластерного анализа является то, что позволяет разделить запасы по маркам на основе комплекса показателей. Но известно, что количество классификационных показателей, характеризующих марочный состав углей по пластам различно, что накладывает определенные ограничения на использование метода кластерного анализа для определения марочного состава по совокупности классификационных показателей марки угля. Поэтому, если использовать при кластерном анализе всю совокупность показателей, то разделение по маркам будет проведено по минимальному количеству данных. Этот фактор относится к недостаткам применения кластерных методов и ограничению их применения.

Практическая реализация геометризации марочного состава углей Донбасса на основе метода кластерного анализа выполнена для двадцати угольных пластов угольных шахт Донбасса.

Вывод: метод кластерного анализа применим для разделения запасов угля по маркам. Достоинством является возможность учитывать совокупность классификационных показателей марок угля. В качестве меры сходства двух элементов принимается евклидово расстояние, элементами которого являются пары классификационных показателей, характеризующие марочный состав угля.

Литература

1. ДСТУ 3472-96. Вугілля буре, кам'яне та антрацит. Класифікація: Держстандарт України. – Київ, 1997. – 5 с.
2. *Апраушева Н.Н.* Три алгоритма естественной кластеризации объектов. – М.: ВЦ АН СССР, 1986. – 22 с.
3. *Апраушева Н.Н.* Предварительное обнаружение идеальных кластеров и оценивание их числа. – М.: ВЦ АН СССР, 1987. – 20 с.
4. *Апраушева Н.Н.* Некоторые методы обнаружения кластеров. – М.: ВЦ АН СССР, 1988. – 22 с.
5. *Апраушева Н.Н., Ражабов Б.Ж.* Экспериментальные исследования по обнаружению кластеров. – М.: ВЦ АН СССР, 1989. – 26 с.
6. *Апраушева Н.Н., Гридина Е. Д.* Дополнительные исследования по обнаружению кластеров. – М.: ВЦ АН СССР, 1991. – 20 с.
7. *Мирний В.В., Філатова І.В.* Застосування методу кластерного аналізу для розподілу запасів вугілля за марками // Вісник Житомирського державного технологічного університету: Технічні науки. Житомир: ЖДТУ. – 2007. – Випуск II (41). – С. 164-172.

Філатова І.В.

ассистент кафедры «Маркшейдерское дело» Государственного высшего учебного заведения «Донецкий национальный технический университет».
