© О.С. Смирнова, Ф.А. Голынская, 2016

УДК 552.574

О.С. Смирнова, Ф.А. Голынская

СТАТИСТИЧЕСКИЕ МЕТОДЫ В ПРОГНОЗИРОВАНИИ САМОВОЗГОРАНИЯ УГЛЕЙ

Приведены результаты статистического анализа данных, полученных по результатам геологоразведочных работ, ряда угольных месторождений с целью установления степени опасности самовозгорания углей. Задача исследований состояла в том, чтобы, используя геологические данные об угольном пласте, составить прогноз самовозгорания углей. Авторами была разработана оригинальная методика, в основе которой лежит ранговая модель данных, и идея, состоящая в классификации данных по «близости» к эталонным группам наблюдений. Полученные в каждой точке наблюдений (скважине) данные о степени самовозгораемости углей позволяют показать пространственное распределение разных по степени опасности самовозгорания углей в пределах шахтного поля, отдельного месторождения и пр. путем применения геоинформационных программных продуктов, например, семейства ArcGIS.

Ключевые слова: статистический анализ, самовозгорание углей, геологические факторы, ранговая модель, эталонные группы, нормативы, эталонные точки-концентраторы, степень опасности самовозгорания.

Предотвращение и ликвидация угольных пожаров является частью программы, осуществляемой в соответствии с Киотским протоколом к Рамочной Конвенции ООН об изменении климата, принятым 11 декабря 1997 г. в Киото. В протоколе указано: «Принятое соглашение призвано решить самую важную проблему мировой экологии: в атмосфере постоянно накапливаются газы, вызывающие парниковый эффект. Источником выбросов этих веществ является промышленность, использующая ископаемое топливо» [4]. В Российской Федерации Киотский протокол был ратифицирован и подписан в 2004 г., вступил в силу 16 февраля 2005 г. [3].

Наблюдения в горных выработках показали, что самовозгорание углей может быть вызвано не только горнотехническими условиями разработки угольных пластов (вентиляция, осушение, обрушение кровли и т.д.). Важную роль в этом процессе играют условия залегания угольного пласта и вещественный

состав углей, то есть те характеристики, которые связаны с генетическими условиями формирования угольной залежи. К настоящему времени установлены геолого-генетические факторы, которые, по мнению ученых, могут влиять на самовозгорание углей, которые условно можно разделить на две группы: а) факторы, связанные с условиями залегания угольного пласта – мощность угольного пласта, глубина его залегания и угол наклона, нарушенность, строение угольного пласта, вмещающие породы; б) факторы, связанные с составом и свойствами вещества углей – степень метаморфизма, петрографический состав, зольность, влажность, выход летучих веществ, сернистость, минеральный состав, газоносность [1, 2].

В результате изучения геологического строения угольных месторождений, физико-химических исследований и анализа статистических данных о самовозгорании углей установлены геологические факторы самовозгорания углей и граничные значения их параметров разной степени опасности самовозгорания для ряда бассейнов и месторождений.

Предварительный анализ собранных данных, произведенный с помощью статистического пакета STATISTICA, показал, что распределения рассматриваемых факторов самовозгорания углей, во-первых, не могут быть признаны нормально распределенными на сколько-нибудь удовлетворительном уровне надежности, во-вторых, эти факторы, как правило, не имеют значимых корреляций. Кроме того, данные имеют разную природу (непрерывные и дискретные с малым числом значений), и различный диапазон значений. Перечисленные причины исключают корректное использование большинства классических методов многомерного статистического анализа. Обычно в подобных ситуациях для анализа данных используются методы непараметрической (ранговой) статистики. Нами была разработана оригинальная методика, в основе которой лежит ранговая модель данных, и идея, несколько напоминающая постановку задачи в классическом дискриминантном анализе, состоящая в классификации данных по «близости» к эталонным группам наблюдений.

С использованием нормативов определения уровней опасности факторов (всего 13) осуществляется переход от исходных данных к ранговой шкале (таблица). Следует заметить, что исходные данные по конкретному месторождению могут содержать не все факторы. При переходе к ранговой шкале пропущенные данные мы заменяем нулями, которые в соответствии

Уровни опасности геологических факторов в ранговой шкале

Регион	Мощность, м	Угол наклона, град	Тектоническая нарушенность, м/км^2	Карст	Влажность, W^a, %	Сера общая, S_t^d, %	Глубина залегания угольного пласта, м	Метаноносность, м^3/т	Выход летучих компонентов, V^{daf}, %	Инертинит, J, %	Строение (количество угольных пачек)	Зольность, A^d, %	Витринит, V_t, %	Уровень опасности самовозгорания
Кузнецкий бассейн	1	2	1	0	1	2	3	1	1	2	1	3	1	М*
	2	1	3	0	1	1	1	1	1	2	1	3	1	Н**
	1	1	3	0	1	1	2	1	1	3	1	3	2	М
	2	1	1	0	1	1	2	1	1	1	2	3	1	Н
	1	1	1	0	1	1	1	1	2	1	1	3	1	Н
	1	2	1	0	1	1	3	1	1	1	1	1	1	Н
	1	2	3	0	1	1	3	1	2	1	2	1	1	М
	1	3	2	0	1	1	1	3	2	2	2	1	2	М
	3	3	3	0	1	1	1	3	2	3	3	1	2	О***

Подмосковный бассейн	3	1	1	2	3	2	2	3	3	2	2	2	3	М
	2	1	1	2	3	3	2	3	3	2	2	2	3	М
	3	1	1	1	3	2	3	3	3	1	2	2	2	М
	3	1	1	1	3	2	2	3	3	1	2	2	3	М
	3	1	1	1	3	3	2	3	3	1	2	2	3	О
	3	1	1	1	3	2	3	3	3	3	2	2	3	О
	3	1	1	2	2	2	1	3	3	1	3	2	2	М
	3	1	1	2	1	2	1	3	3	1	2	2	3	М
	3	1	1	3	3	3	1	3	2	3	3	2	1	О
Вес	6	6	6	6	6	6	5	5	5	5	4	4	4	
Точка 1	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	1	1	1	
Точка 2	1,75	1,75	1,75	1,75	1,75	1,75	1,75	1,75	1,75	1,75	2	2	2	
Точка 3	2,75	2,75	2,75	2,75	2,75	2,75	2,75	2,75	2,75	2,75	3	3	3	

М* – угли, малоопасные по самовозгоранию; Н** – угли неопасные по самовозгоранию; О*** – угли, опасные по самовозгоранию.

с нормативами преобразованы в уровни опасности: 1 – низкий, 2 – средний, 3 – высокий так же, как все прочие значения. Это позволит нам при необходимости классифицировать точки не только внутри месторождения, но и провести глобальную классификацию с учетом полного набора факторов.

Данные в преобразованной ранговой таблице разобьем на группы, факторам каждой группы мы припишем одинаковый вес – коэффициент тем более высокий, чем больше их значимость для формирования общего уровня опасности. Нулевой вес позволяет игнорировать фактор при расчетах (например, в случае отсутствия данных).

Будем теперь классифицировать точки 13-мерного пространства рангов по близости (минимуму расстояния) к трем эталонным точкам-концентраторам, являющихся центрами тяжести эталонных групп, составленных из «типичных представителей» мало-, средне- и высокоопасных наборов рангов факторов самовозгорания углей. Наша задача состоит в том, чтобы выбрать меру близости (например, метрику в пространстве рангов), весовые коэффициенты значимости и 3 точки-концентратора. Таким образом, мы получаем модель с очень большим числом параметров, от подбора которых будет зависеть качество классификации. Некоторые из этих параметров могут быть вычислены (хотя бы приближенно), если сделать некоторые разумные предположения о желательных свойствах алгоритма классификации, остальные определяются в ходе компьютерного эксперимента, моделирования.

В прикладной статистике широко применяется метод DATA-MINING – «добыча данных» (самым известным примером этого метода является как раз дискриминантный анализ), который заключается в том, что малая часть большой выборки может быть использована для «обучения» метода, оптимальной подгонки параметров модели, возможно с привлечением эксперта. Этим методом мы и воспользовались, сформировав «обучающую» выборку малого объема (58 точек из разных месторождений), обладающую следующим свойством репрезентативности: она включает в себя как точки с очевидной классификацией, так и различные примеры пограничных случаев. Процесс подгонки модели проводится в несколько этапов. Уточнения значений параметров и возможные усложнения модели на каждом следующем этапе подгонки должны сохранять класс опасности для тех точек обучающей выборки, классификация которых на предыдущем этапе была признана успешной.

Будем поэтапно определять параметры модели. Прежде всего, определимся с метрикой. Нами были рассмотрены два типа расстояний:

обычное евклидово, $\rho_2(x, y) = \sqrt{\sum_{j=1}^{n} (x_j - y_j)^2}$,

где $x = (x_1, ..., x_n)$, $y = (y_1, ..., y_n)$, n – размерность пространства,

и $\rho_1(x, y) = \sum_{j=1}^{n} \left| x_j - y_j \right|$, часто называемое «манхеттенскими квадратами» – длина пути из точки x в точку y по городу с правильной планировкой. Поскольку мы имеем дело с условными ранговыми данными, более предпочтительным кажется выбор именно этого расстояния. Действительно, как показывают проведенные теоретические исследования и результаты моделирования, алгоритм классификации, построенный на основании этого расстояния, оказывается более чувствительным, зато использование евклидовой метрики позволяет получить простые линейные условия дискриминации в пространстве рангов.

В этом случае мы можем принять низкий уровень опасности для точки $r = (r_1, ..., r_n)$, если

$$\sum_{j=1}^{n} w_j r_j (y_j - x_j) < \sum_{j=1}^{n} w_j (y_j^2 - x_j^2) / 2$$

высокий, если

$$\sum_{j=1}^{n} w_j r_j (z_j - y_j) \geq \sum_{j=1}^{n} w_j (z_j^2 - y_j^2) / 2$$

и средний во всех остальных случаях. Здесь точки $x = (x_1,..., x_n)$, $y = (y_1, ..., y_n)$, $z = (z_1, ..., z_n)$ – концентраторы для низкого, среднего и высокого уровня опасности, w_j – вес, r_j – вычисленный по нормативам ранг для j-го фактора. Эта модель заслуживает дальнейшего изучения.

В наиболее простой модели естественно положить все веса равными единице, то есть считать все факторы равнозначными, а точки-концентраторы равными (3,3,...,3), (2,2,...,2), (1,1,...,1). Отправляясь именно от этой модели, мы перешли в результате компьютерных экспериментам к весам 6, 5, 4 для групп факторов (целый вес w удобно истолковывать как дублирование фактора w раз), после чего начали уточнять координаты концентраторов.

В качестве концентратора в классе наибольшей опасности был выбран центр тяжести облака точек, для которых ранг 3 может быть заменен меньшим не более чем в двух любых позициях. Все координаты этой точки будут равны

$$3 - 3(2n - 1) / (2n^2 + 1),$$

n здесь и далее – число факторов. При варьировании 10–13 координат мы получим значения приблизительно в диапазоне 0,72–0,78. Выберем среднее значение 0,75 в качестве стартового для всех координат точки z. Координаты остальных двух точек могут быть получены аналогично, варьированием координат точек $x = (1,1,...,1)$ и $y = (2,2,...,2)$ (причем, только в меньшую сторону), для чего нам будет удобно рассмотреть дополнительный ранг 0 (игнорирование влияния фактора или полная безопасность). Такой выбор точек объясняется тем, что во всех спорных случаях мы всегда «перестраховываемся» – принимаем решение в пользу большей опасности.

Если концентраторы имеют вид $x = (x_1,..., x_n)$, $y = (y_1,..., y_n)$, $z = (z_1,..., z_n)$, то условия минимизации расстояния в обеих рассмотренных метриках преобразуются в простые линейные соотношения, связывающие n_1, n_2 и n_3, где n_r – число появлений фактора r. Так как $n_2 = n - n_1 - n_3$, то фактически мы переходим к двумерным линейным условиям дискриминации, которые формулируются в терминах числа появления «1» и «3» в данном наборе рангов.

Используя этот факт, мы можем вычислить два оставшихся концентратора с помощью следующих простых условий нормировки: а) при $n_2 = 0$ и $n_1 > \lambda n$, $\lambda > 1/2$, будем признавать точку малоопасной (то есть допускать, что происходит компенсация малого числа опасных факторов большим числом безопасных значений), б) при $n_1 = 0$ считать точку уже высоко-опасной, если $n_3 \geq (1 - \lambda)n$. В наших расчетах мы использовали диапазон значений λ от 3/5 до 2/3 (принцип «парламентского большинства»).

Полагая $z = 3 - \gamma$, $y = 2 - \beta$ и $x = 1 - \alpha$ выпишем условия дискриминации в случае линейной метрики ρ_1 и произвольного набора весов w. Точка мало-опасна при $n > n(1/2 + \alpha/(2(1-\gamma)))$ высоко-опасна при $n_3 \geq \beta n_1/(1-\gamma) + n(1/2 - \beta/(2(1-\gamma)))$ и имеет средний уровень опасности в остальных случаях. Здесь $n_1 \geq 0$, $n_3 \geq 0$ и $n_1 + n_3 \leq w$, где $w = \sum_{j=1}^{n} w_j$ сумма весов всех факторов, $n_r = \sum_{j=1}^{n} w_j I(r_j = r)$. В модели с единичными весами w совпадает с n, числом факторов (13), n_r является числом появлений ранга r. Условия нормировки дают недоопределенную систему линейных уравнений $2\lambda - \beta/(1-\gamma) = 1$, $2\lambda - \alpha/(1-\beta) = 1$. Подставляя в нее полученное ранее значение γ, вычислим координаты оставшихся точек-концентраторов. Например, при $\lambda = 2/3$ и $\gamma = 1/4$ мы получим решение $\alpha = \beta = 1/4$, которое и будем использовать как начальное при подгонке параметров модели, варьируя эти значения на обучающей выборке. В ре-

зультате серии компьютерных экспериментов мы пришли к следующим параметрам: x = (0,75,..., 0,75, 1,1,1), y = (1,75,..., 1,75, 2,2,2), z = (2,75,..., 2,75, 3,3,3), то есть, координаты для малозначимых факторов не варьируются, набор весов по группам остался 6, 5, 4. С этим набором параметров мы произвели вычисления уже для полного объема данных.

Все необходимые преобразования исходных данных и вычисления производятся автоматически программой (макросом) на языке Visual Basic для Excel, реализующей предлагаемую методику. Мы сознательно отказались от работы в среде специализированных статистических пакетов в пользу программ Excel MS Office по причине их общедоступности. Данные, подлежащие обработке, собраны в книгу Excel, которая содержит по отдельному листу для каждого месторождения и еще один лист с нормативной информацией.

На основании произведенных вычислений можно сделать следующие выводы:

1. Для определения степени самовозгораемости угольного пласта по данным геологоразведочных работ, ввиду отсутствия между этими данными в точках наблюдений (скважинах) значимых корреляций, наличие у них как непрерывных, так и дискретных значений, целесообразно применение метода многомерной классификации по эталонным точкам, который состоит в классификации данных по «близости» к эталонным группам наблюдений.

2. Все месторождения можно условно разделить на две группы – те, где присутствует большая доля точек, высокоопасных по самовозгоранию углей (Бельцевское, бурые угли) на общем фоне среднего уровня опасности, и те, где, наоборот, много точек с низким уровнем опасности или встречаются лишь единичные очень опасные точки на общем фоне среднего уровня опасности.

3. Полученные в каждой точке наблюдений (скважине) данные о степени самовозгораемости углей позволяют показать пространственное распределение разных по степени опасности самовозгорания углей в пределах шахтного поля, отдельного месторождения и пр. путем применения геоинформационных программных продуктов, например, семейства ArcGIS.

СПИСОК ЛИТЕРАТУРЫ

1. *Голынская Ф.А.* Самовозгорание углей в пластах и причины его возникновения // Вестник МГУ, серия Геология. – 1998. – № 6. – С. 65–66.

2. *Голынская Ф. А.* Характеристика наиболее действенных факторов самовозгорание углей в пластах (статья) // Горный информационно-аналитический бюллетень. – 2011. – № 2. – С. 19–23.
3. *Федеральный* закон «О ратификации Киотского протокола к Рамочной конвенции Организации Объединенных Наций об изменении климата» от 4 ноября 2004 года № 128-ФЗ.
4. *Kyoto* Protocol to the UN Framework Convention on Climate Change. – http://unfccc.int/2860.php/ – The official website of the UNFCCC. ГИАБ

КОРОТКО ОБ АВТОРАХ

Голынская Фарида Асхатовна – кандидат геолого-минералогических наук, доцент, НИТУ «МИСиС», e-mail: ud@msmu.ru,
Смирнова Ольга Сергеевна – кандидат физико-математических наук, старший научный сотрудник, МГУ им. М.В. Ломоносова.

UDC 552.574

O.S. Smirnova, F.A. Golynskaya

STATISTICAL METHODS IN THE PREDICTION OF SPONTANEOUS COMBUSTION OF COAL

The article presents the results of statistical analysis of data obtained from the results of exploration activities, the number of coal deposits with the aim of establishing the degree of danger of spontaneous combustion of coal. The objective of the research was that, using geological data about coal seam, make a prediction of spontaneous combustion of coal. The authors developed an original technique, which is based on ranking the data model, and the idea consisting in the classification data «close» to the reference groups of observations. Obtained at each point of observation (borehole) data on the degree of samovozgoraemost coals allow to show the spatial distribution of different degree of danger of spontaneous combustion of coal within the mine field, private field, etc. through the use of GIS software, such as ArcGIS family.

Key words: statistical analysis, the spontaneous combustion of coal, geological factors, the ranking model, the reference group, the standards, the reference point, the hub, the danger of spontaneous combustion.

AUTHORS

Golynskaya F.A., Candidate of Geological and Mineralogical Sciences,
Associate Professor, National University of Science and Technology «MISiS»,
119049, Moscow, Russia, e-mail: ud@msmu.ru,
Smirnova O.S., Candidate of Physical and Mathematical Sciences,
Lomonosov Moscow State University, Moscow, Russia.

REFERENCES

1. Golynskaya F. A. *Vestnik MGU, seriya Geologiya*. 1998, no 6, pp. 65–66.
2. Golynskaya F. A. *Gornyy informatsionno-analiticheskiy byulleten'*. 2011, no 2, pp. 19–23.
3. *Federal'nyy zakon «O ratifikatsii Kiotskogo protokola k Ramochnoy konventsii Organizatsii Ob"edinennykh Natsiy ob izmenenii klimata» ot 4 noyabrya 2004 goda no 128-FZ* (Federal law «On ratification of the Kyoto Protocol to the framework Convention of United Nations on climate change» on November 4, 2004 № 128-FZ).
4. *Kyoto Protocol to the UN Framework Convention on Climate Change*. http://unfccc.int/2860.php/ The official website of the UNFCCC.