УДК 519.21 : 539.261

Е. Д. КОНСОН, М. С. НАХМАНСОН
(Ленинград)

БАЙЕСОВСКИЙ ПОДХОД К ЗАДАЧЕ КАЧЕСТВЕННОГО РЕНТГЕНОВСКОГО АНАЛИЗА

Одной из основных задач, решаемых с использованием баз рентгенографических данных, является задача качественного анализа состава соединений [1, 2]. В существующих системах качественного анализа [3—5] идентификация фаз проводится на основе анализа разнообразных невязок между рентгенограммой исследуемого образца и эталонными рентгенограммами отдельных фаз. Поэтому наиболее эффективно такие алгоритмы могут работать при распознавании однофазных образцов, а также на начальном этапе анализа, когда выделяется значительное число фаз, заведомо не входящих в образец, в результате чего сужается область окончательного поиска. Определяющей особенностью предлагаемого подхода является сопоставление рентгенограммы образца с комбинациями рентгенограмм эталонных фаз.

Данный подход не содержит существенных ограничений, связанных со спецификой рентгеновских исследований, и может быть использован при проведении анализа по спектральным данным иного происхождения.

1. Постановка задачи. В работах [4, 5] в основу алгоритмов положено векторное представление рентгенограмм, при котором рассматриваемый диапазон межплоскостных расстояний от $d_{\min}$ до $d_{\max}$ разбивается на n промежутков длиной $\delta = (d_{\max} - d_{\min})/n$. Рентгенограмма интерпретируется как вектор-столбец, j-й элемент которого задается значениями рентгенограммы в промежутке $[d_{\min} + (j-1)\delta,\ d_{\min} + j\delta]$. Например [4], если в промежутке имеется линия, то j-й элемент полагается равным относительной интенсивности этой линии, в противном случае — нулю. Значение n выбирается таким, чтобы ни в один из промежутков не попало более одной линии. В [5] на основании физических предположений применяется иная зависимость для определения значений элементов вектора.

Пользуясь векторным представлением, изложим формальную постановку следующим образом. Пусть заданы в R_n векторы a_j, $j = 1, 2, \ldots, N$. Элементы каждого из векторов a_j удовлетворяют условию $0 \leqslant a_{ij} \leqslant 100$, $i = 1, 2, \ldots, n$. Векторы получены на основе рентгенограмм эталонных фаз, содержащихся в специализированной библиотеке. Рентгенограмма представляемого для качественного анализа образца тоже интерпретируется как вектор $y \in R_n$, причем $0 \leqslant y_i \leqslant 100$, $i = 1, 2, \ldots, n$.

Считается, что рентгенограммы эталонов и образца заданы с аддитивными ошибками Δa_j и Δy, которые переносятся в векторное представление.

Анализируемая рентгенограмма есть взвешенная сумма рентгенограмм всех фаз, входящих в образец. Поэтому для точных значений векторов, т. е. для $a_j - \Delta a_j$ и $y - \Delta y$, можно записать соотношение

$$y - \Delta y = \sum_{j=1}^{N} x_j (a_j - \Delta a_j) \qquad (1)$$

(x_j — скалярные неотрицательные величины).

Преобразуем (1) следующим образом:

$$y + \Delta = \sum_{j=1}^{N} x_j a_j, \tag{2}$$

где $\Delta = \sum_{j=1}^{N} x_j \Delta a_j - \Delta y$.

Переходя к матричным обозначениям $A = (a_1, \ldots, a_N)$ и $x = (x_1, \ldots, x_N)^T$, имеем

$$y + \Delta = Ax, \quad x \geqslant 0. \tag{3}$$

Будем полагать, что вектор приведенных к выходу ошибок ограничен следующим образом: $-\varepsilon_1 \leqslant \Delta \leqslant \varepsilon_2$. Тогда последнее соотношение образует систему неравенств для возможных значений x после наблюдения y:

$$y - \varepsilon_1 \leqslant Ax \leqslant y + \varepsilon_2. \tag{4}$$

В задаче качественного фазового анализа по наблюдаемому y требуется высказать суждение о составе образца, указать входящие в него фазы. Полученная система неравенств (4) показывает, что однозначного суждения для каждой фазы в общем случае дать нельзя, так как существует множество допустимых комбинаций фаз. Под допустимой комбинацией понимается группа векторов a_j, $j \in J$, для которой существуют положительные коэффициенты x_j такие, что сумма $\sum_{j \in J} a_j x_j$ удовлетворяет системе неравенств (4).

Подчеркнем, что суммы с различными положительными коэффициентами x_j, но образованные на одной группе векторов a_j не различаются и считаются одной допустимой комбинацией.

Обычно предполагается, что образец не содержит фаз, которые не включены в исходное множество A. Это означает, что всегда имеются допустимые комбинации, по крайней мере одна.

Покажем, как, используя вероятностную модель для описания образца y, можно проранжировать фазы и комбинации фаз, удовлетворяющие системе неравенств (4).

2. **Вероятностная модель.** Будем считать, что фазы входят в образец независимо в вероятностном смысле одна от другой, причем априорная вероятность появления каждой фазы a_j в поступающем на анализ образце одинакова и равна p_0. Это предположение согласуется с процедурой создания специализированной библиотеки [6]. В нее исследователем включаются лишь те фазы, которые потенциально могут появляться в образцах в данных конкретных условиях анализа.

Следствием сделанного предположения является априорное биномиальное распределение комбинаций фаз. Вероятность появления образца, содержащего k конкретных фаз, есть $p_0^k (1 - p_0)^{N-k}$. Данная система событий полная, однако она содержит «пустой» образец, состоящий из нуля фаз. Это событие не имеет физического смысла, поэтому исключим его из системы, а вероятности оставшихся событий для сохранения условия полноты пронормируем, умножив на $L = [1 - (1 - p_0)^N]^{-1}$.

Таким образом, данная модель не позволяет априорно высказать суждение о предпочтительном появлении каких-либо конкретных фаз и комбинаций в анализируемых образцах.

3. **Апостериорное ранжирование.** Пусть для анализа представлена рентгенограмма и построен соответствующий вектор y, задающий систему неравенств (4). Пусть по этой системе определено множество всех допустимых комбинаций D:

$$D = \{D_1, D_2, \ldots, D_T\}, \quad T \leqslant 2^N - 1. \tag{5}$$

Нетрудно определить элементы соответствующего множества априорных

вероятностей:

$$p(D_j) = Lp_0^{k_{D_j}}(1-p_0)^{N-k_{D_j}}, \quad j = 1, 2, \ldots, T. \tag{6}$$

Здесь и в дальнейшем k обозначает число фаз, входящих в комбинацию, которую указывает нижний индекс при k. Пользуясь формулой Байеса, для апостериорных вероятностей получаем выражение

$$p(D_j|y) = p(D_j)\Big/\sum_{D_j \in D} p(D_j), \tag{7}$$

т. е. апостериорные вероятности отличаются только нормирующим множителем.

Не нарушая общности, будем считать, что $p(D_1|y) \geqslant p(D_2|y) \geqslant \ldots \ldots \geqslant p(D_T|y)$. Тогда результатом качественного фазового анализа можно назвать комбинацию D_1, обладающую максимальной апостериорной вероятностью. Другими словами, содержащаяся в y экспериментальная информация позволяет выдвинуть гипотезу: образец состоит из фаз комбинации D_1. При этом максимальна вероятность того, что гипотеза справедлива.

Если несколько комбинаций имеют одинаковую наибольшую апостериорную вероятность, то можно разрешить альтернативу, пользуясь дополнительной информацией, или провести ранжирование на уровне фаз.

Пусть $D(a_j) = \{D_1(a_j), D_2(a_j), \ldots, D_{t_j}(a_j)\}$ есть множество допустимых комбинаций, в которые входит фаза a_j. Тогда апостериорная вероятность j-й фазы в образце определяется как

$$p(a_j|y) = \sum_{D_s(a_j) \in D(a_j)} p(D_s(a_j))\Big/\sum_{D_i \in D} p(D_i). \tag{8}$$

Аналогично можно считать, что $p(a_1|y) \geqslant p(a_2|y) \geqslant \ldots \geqslant p(a_N|y)$, и указывать как результат анализа несколько первых фаз, связывая с этим вероятностную аргументацию, которая приводилась выше.

4. Отыскание допустимых комбинаций и их свойства. Введем в систему неравенств (4) дополнительные векторные переменные z_1 и z_2 и перейдем к следующей системе:

$$Ax + z_1 = y + \varepsilon_1, \quad Ax - z_2 = y - \varepsilon_2,$$
$$z_1 \geqslant 0, \quad z_2 \geqslant 0, \quad x \geqslant 0. \tag{9}$$

Данная система накладывает эквивалентные ограничения на вектор x. Используя новые обозначения

$$A_P = \left(\begin{array}{c|c|c} A & E & 0 \\ \hline A & 0 & -E \end{array}\right), \quad x_P^{\mathrm{T}} = (x, z_1, z_2)^{\mathrm{T}}, \quad y_P^{\mathrm{T}} = (y + \varepsilon_1, y - \varepsilon_2)^{\mathrm{T}}, \tag{10}$$

ее можно записать следующим образом:

$$A_P x_P = y_P, \quad x_P \geqslant 0. \tag{11}$$

Полученные условия являются ограничениями задачи линейного программирования, записанной в канонической форме [7]. Условия выделяют в пространстве R_{N+2n}, где определен вектор x_P, выпуклый многогранник. Нетрудно видеть, что базис каждой угловой точки многогранника однозначно задает допустимую комбинацию в данном выше определении.

Вектор $\overset{\circ}{x}_P$, задающий угловую точку, содержит не более $2n$ строго положительных компонент, остальные равны нулю. Те компоненты вектора x, которые оказались положительными в $\overset{\circ}{x}_P$, определяют вхо-

дящие в допустимую комбинацию векторы a_j. Будем называть эти комбинации базовыми. Они образуют множество $B = \{B_1, B_2, \ldots, B_Q\}$.

Можно показать, что имеют место следующие свойства базовых и допустимых комбинаций.

4.1. Число фаз в любой базовой комбинации заключено в пределах от 1 до n.

4.2. Если при заданном y минимальное число фаз, образующих базовую комбинацию, равно m, то не существует допустимых комбинаций из $m_1 < m$ фаз.

4.3. Любая допустимая комбинация содержит базовые.

4.4. Допустимая комбинация с числом фаз m является базовой.

4.5. Для любой фазы из допустимой комбинации найдется базовая комбинация, содержащая эту фазу.

4.6. Обозначим через $D(B_j) = \{D_1(B_j), \ldots, D_{s_j}(B_j)\}$ множество допустимых комбинаций, содержащих B_j. Для суммы априорных вероятностей допустимых комбинаций, содержащих B_j, справедлива оценка

$$Lp_0^{k_{B_j}}(1-p_0)^{N-k_{B_j}} \leqslant \sum_{D_i(B_j)\in D(B_j)} p(D_i(B_j)) \leqslant Lp_0^{k_{B_j}}. \qquad (12)$$

5. **Уточнение вероятностной модели.** Поскольку число фаз, содержащихся в любом анализируемом образце, как правило, существенно меньше числа фаз, имеющихся в специализированной библиотеке, можно считать, что

$$P\{0 < k \leqslant l\} = 1 - \alpha, \qquad (13)$$

где k — случайное число фаз в образце, l — целое положительное число, $l \ll N$, α — положительное близкое к нулю число.

Пользуясь выражением для вероятности появления образца из k фаз, раскроем условие (13):

$$\frac{1}{1-(1-p_0)^N} \sum_{k=1}^{l} C_N^k p_0^k (1-p_0)^{N-k} = 1 - \alpha. \qquad (14)$$

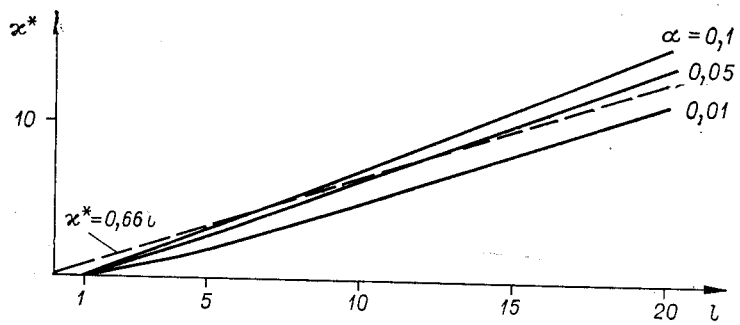
При заданных N, l, α последнее соотношение является уравнением относительно p_0. Если $N \to \infty$, а параметры l и α фиксированы, то биномиальные члены в сумме стремятся к соответствующим членам распределения Пуассона с параметром $\varkappa = p_0 N$ [8], а (14) переходит в асимптотическое уравнение для $\varkappa$.

На графике приведены зависимости корня этого уравнения $\varkappa^*$ от l при различных α. В диапазоне $\alpha = 0{,}01 - 0{,}1$ для p_0 допустимо использование приближенной формулы $p_0 \approx 0{,}66\, l N^{-1}$. При возрастании N вероятность p_0 имеет порядок N^{-1}. Число эталонов в специализированных библиотеках колеблется в пределах 10^2—10^4, а параметр l имеет порядок 10^1. Поэтому вероятность p_0 можно считать весьма малой: $\sim 10^{-3}$—10^{-1}.

6. **Оценка апостериорных вероятностей.** Прежде всего укажем ограничения для значений знаменателя в формулах (7) и (8). Так как согласно п. 4.3 любая допустимая комбинация содержит базовую, запишем:

$$\sum_{B_i\in B} p(B_i) \leqslant \sum_{D_i\in D} p(D_i) \leqslant \sum_{B_i\in B} \sum_{D_j(B_i)\in D(B_i)} p(D_j(B_i)). \qquad (15)$$

Оценка снизу содержит вероятности появления базовых комбинаций. Оценка сверху включает все допустимые комбинации, некоторые из них могут повторяться, так как допустимая комбинация может содержать более одной базовой.



С учетом неравенства (12) соотношение (15) можно преобразовать:

$$L \sum_{B_i \in B} p_0^{k_{B_i}} (1-p_0)^{N-k_{B_i}} \leqslant \sum_{D_i \in D} p(D_i) \leqslant L \sum_{B_i \in B} p_0^{k_{B_i}}. \qquad (16)$$

Отсюда следует, что для произвольной допустимой комбинации D_i выполняются неравенства

$$\frac{p_0^{k_{D_i}} (1-p_0)^{N-k_{D_i}}}{\sum\limits_{B_i \in B} p_0^{k_{B_i}}} \leqslant p(D_i | y) \leqslant \frac{p_0^{k_{D_i}} (1-p_0)^{N-k_{D_i}}}{\sum\limits_{B_i \in B} p_0^{k_{B_i}} (1-p_0)^{N-k_{B_i}}}. \qquad (17)$$

Построим оценки апостериорных вероятностей отдельных фаз a_j. В силу результатов п. 4.5 запишем следующие неравенства для значений числителя в формуле (8):

$$\sum_{B_i(a_j) \in B(a_j)} p(B_i(a_j)) \leqslant \sum_{D_i(a_j) \in D(a_j)} p(D_i(a_j)) \leqslant$$
$$\leqslant \sum_{B_i(a_j) \in B(a_j)} \sum_{D_s(B_i(a_j)) \in D(B_i(a_j))} p(D_s(B_i(a_j))). \qquad (18)$$

Оценка сверху включает все допустимые комбинации, содержащие базовые, в которые входит фаза a_j. Пользуясь соотношением (12), оценки (18) можно преобразовать к виду, аналогичному (16), а затем, подставив их в формулу (8), получить для апостериорных вероятностей a_j неравенства

$$\frac{\sum\limits_{B_i(a_j) \in B(a_j)} p_0^{k_{B_i(a_j)}} (1-p_0)^{N-k_{B_i(a_j)}}}{\sum\limits_{B_i \in B} p_0^{k_{B_i}}} \leqslant p(a_j | y) \leqslant \frac{\sum\limits_{B_i(a_j) \in B(a_j)} p_0^{k_{B_i(a_j)}}}{\sum\limits_{B_i \in B} p_0^{k_{B_i}} (1-p_0)^{N-k_{B_i}}}. \qquad (19)$$

Нетрудно убедиться, что если $p_0 < 0{,}5$, то справедливо неравенство

$$p_0^k (1-p_0)^{N-k} > p_0^{k+1} (1-p_0)^{N-k-1} \qquad (20)$$

при любых N и k. С учетом пп 4.2, 4 это означает, что максимальной апостериорной вероятностью будут обладать базовые комбинации из m фаз, т. е. они явятся решением задачи качественного фазового анализа. Не нарушая общности, можно полагать, что этими комбинациями будут $B_1, B_2, \ldots, B_r$. В п. 5 показано, что практически вероятность p_0 считается достаточно малой. Поэтому имеют смысл асимптотические значения апостериорных вероятностей при $p_0 \to 0$. Для базовых комбинаций с учетом (17) имеем

$$\lim_{p_0 \to 0} p(B_i | y) = \begin{cases} 1/r & \text{при } i = 1, 2, \ldots, r; \\ 0 & \text{при } i = r+1, \ldots, Q. \end{cases} \quad (21)$$

Асимптотические значения вероятностей $p(D_j|y)$ тоже равны нулю, если D_j не совпадает с какой-либо комбинацией из группы B_1, B_2, ..., B_r. Из (19) для отдельных фаз получаем

$$\lim_{p_0 \to 0} p(a_j | y) = \pi_i / r, \quad j = 1, 2, \ldots, N, \quad (22)$$

где π_i — число комбинаций из группы B_1, B_2, ..., B_r, в которые попадает фаза a_j.

При $r = 1$ комбинация B_1 является решением задачи с вероятностью, близкой к единице. Если $r > 1$, то можно проранжировать входящие в группу фазы, пользуясь формулой (22).

Таким образом, в работе показано, что на основе байесовского подхода к задаче качественного фазового анализа могут быть получены вероятностные выводы о составе анализируемого образца. Отыскание вероятностных характеристик связано с обходом угловых точек выпуклого многогранника. Практически этот обход реализуется с помощью вычислительных процедур симплексного метода линейного программирования.

ЛИТЕРАТУРА

1. **Евграфов А. А. и др.** Автоматический рентгенофазовый анализ с использованием машинного банка эталонных данных.— В кн.: 50 лет отечественного рентгеновского приборостроения. Л.: Машиностроение, 1978.
2. **Powder** Diffraction File. Inorganic Section JCPDS.— Pennsylvania: Swarthmore, 1944—1974.
3. **Johnson G. G., Vand V.** Computerised Multiphase *X*-ray Powder Diffraction Identification System.— Advanses in *X*-ray Analysis, 1968, N 11, p. 376—384.
4. **Fiala J.** Optimisation of Powder Diffraction Identification.— J. Appl. Cryst., 1976, N 9, p. 429—432.
5. **Бурова Е. М. и др.** Алгоритмизация процесса обработки данных рентгеновского фазового анализа.— ДАН, 1977, т. 232, № 5, с. 1066—1068.
6. **Евграфов А. А., Нахмансон М. С., Черный Ю. А.** Проблема идентификации фаз при качественном анализе поликристаллических смесей.— В кн.: Аппаратура и методы рентгеновского анализа, 1977, вып. 22.
7. **Карманов В. Г.** Математическое программирование. М.: Наука, 1975.
8. **Кендалл М. Дж., Стьюарт А.** Теория распределений. М.: Наука, 1966.

Поступила в редакцию 28 мая 1979 г.;
окончательный вариант — 17 декабря 1980 г.

УДК 681.327.8

В. Г. ЧЕРЕПАНОВ

(Красноярск)

ОЦЕНКА ПРОИЗВОДИТЕЛЬНОСТИ И ИНФОРМАТИВНОСТИ ИНТЕРФЕЙСОВ

При разработке систем и подсистем сбора и обработки данных возникают вопросы выбора интерфейсов, удовлетворяющих требованиям заданной производительности. В настоящее время разработано значительное количество интерфейсов различного назначения [1—23]. Рекомендации по выбору интерфейсов для конкретного применения носят, как правило, качественный характер [24—28], что не исключает субъективности оценки. В настоящей статье сделана попытка количественной оценки интерфейсов вычислительных устройств и систем.