Т. Д. Гутман, Ш. А. Мустафин

НЕКОТОРЫЕ ПРИЕМЫ ЧИСЛЕННОГО АНАЛИЗА АЭРОКОСМИЧЕСКИХ СНИМКОВ НА ЭВМ

Один из основных типов «рисунков» на аэрокосмических снимках (АС) — совокупность одномерных кривых (SK) на прямоугольнике. В зависимости от особенностей исследуемых территорий, возможны различные числовые классификационные признаки (КП), определяющие особенности классов отображаемых на АС геологических объектов.

1. Представляется целесообразной следующая процедура классификации кривых на АС с применением ЭВМ:

Определение 1. Пусть имеем прямые параллелепипеды Π и Π_1 в пространстве R, $\Pi \subset \Pi_1$ и $M \subset \Pi$, $1 < |M| < \infty$, $0 < p < 1$. Будем говорить, что концентрация точек из M на Π_1 p-неслучайна, если в предположении, что точки из M распределены на Π случайным образом по равномерному закону, имеем: вероятность $P(\Pi_1,R)$ того, что не менее чем $|M \cap \Pi_1|$ точек из M попадут в Π_1, не превышает p.

(1) Выдвигаем гипотезы о классификационных признаках $КП_i$, $i \in 1,k$, $k > 0$; то есть определяем характеристики формы кривых на АС, служащие разделяющими критериями для искомых классов. (2) Выделяем отрезки на диапазонах значений $КП_i$ ($ДКП_i$), на которых группируются значения, характерные для искомых классов (р-отрезки). Обозначим совокупность р-отрезков МР: $MP = \{mp_{ij} | i \in 1,k, j \in 1,m_i\}$. (3) Выделяем параллелепипеды b_i, $i \in 1,r$, вида $mp_{1j(1)} \times \ldots \times mp_{kj(k)}$, где $j(i) \in 1,m_i$, для которых выполняется: концентрация на них точек, изображающих кривые из СК, P_gr-неслучайна, где P_gr — параметр задачи. Совокупность кривых из СК, изображения которых попали в b_i, назовем классом K_i, $i \in 1,r$. Проверку правомерности такого выбора реализуем по следующей процедуре D:

Для каждого класса V_i и каждого КП L находим вектор оценок для $P(V_i,L)$, см. определение 1. Оценки 2 и 5 — величины, пропорциональные вероятности реализации заданной плотности точек на произвольных диапазонах с длинами, определяющими выделенные в (1) классы. Они основаны на соотношениях из [2] для вероятностей выпадения повышенных точечных плотностей на локальных участках. Оценки 1, 3, 4 — верхние границы для $P(V_i,L)$, использующие различные уточнения и следствия для неравенства Чебышева; неравенства для вероятностей больших уклонений; оценки длин доверительных интервалов для средних арифметических выборок заданных объемов [1]. Классы, проекции которых на какой-либо КП L оказались P_gr-неслучайными, считаем достоверными.

2. Чтобы исследовать локальные проявления КП, разбиваем прямоугольник П, описанный около АС, равномерной прямоугольной сетью R, длина звена которой выбирается из содержательных соображений, и применяем к каждой тройке вида $<V_i,K,L>$ (где V_i — один из классов, L — одно из КП, выделенных на этапе (1) из п. 1) алгоритм этапа (3) — процедуру D. Обозначим полученный набор чисел $\{P(V_i \cap K,L) | V_i$ — один из выделенных классов, K — клетка сети R, L — КП$\}$ через PP. Если относительная доля чисел из PP, не меньших P_gr, больше P_gr1 (P_gr1 — параметр задачи), считаем, что КП L имеет локальный R-порядок на классе Vi.

3. Числовые массивы, полученные путем применения метода к пяти АС, использовались для проверки гипотезы о неслучайности кольцевых структур на АС и выделения АС для геологически однородных территорий). Результаты расчетов подтверждаются тем, что они попутно выявили заведомо «неинформативные» признаки и признаки, породившие проверяемую классификацию.

Литература: 1. ***Ширяев А. Н.*** Вероятность. М.: Наука. 1989. 638 с. **2.** ***Гутман Т. Д., Бакиров Н. К.*** Алгоритмы разбиения конечного множества в пространстве R^n на квазиплоские сегменты / Ин-т матем. с ВЦ УНЦ РАН. Уфа. 1999. 50 с. рус. деп. 28.04.99, № 1336–В99.

Т. Д. Гутман

АЛГОРИТМ ВЫДЕЛЕНИЯ ТЕКТОНИЧЕСКИХ НАРУШЕНИЙ НА СЕЙСМОГРАММАХ «ВРЕМЯ–ПЛОТНОСТЬ»

Определение 1. Пусть $V \subset X^2$, $\varepsilon > 0$. Тогда $\Pi(V)$ — прямоугольник, описанный около V, со сторонами, параллельными координатным осям; $R(V,\Delta x,\Delta y)$ — вписанная в $\Pi(V)$ прямоугольная сеть со звеньями Δx и Δy по горизонтали и вертикали соответственно; $O_\varepsilon(V)$ — множество точек X^2, удаленных от V не более, чем на ε.

Обозначение 1. Пусть П — прямоугольник, dx, dy, $d > 0$, $R0 = R(M,dx,dy)$. Тогда назовем Ro любую область на П, высекаемую из П двумя параллельными ломаными, состоящими из звеньев сети R0 и пересекающими ровно по одному разу каждую вертикаль сети R0. Пусть A_1 — произвольная Ro. Тогда любую область A_2 такую, что $A_1 \subset O_d(A_2)$ и $A_2 \subset O_d(A_1)$, назовем Rh(A,d). z — функция, заданная на клетках сети R0; Z(M) — среднее арифметическое значений z на клетках произвольного множества M, состоящего из клеток сети R0; $Z_0 = Z(\Pi)$.

Определение 2. Пусть $W \subset \Pi$ — прямоугольник, состоящий из клеток сети R0, $\Delta x = n_1 \times dx$, $\Delta y = n_2 \times dy$, где n_1, n_2 — натуральные числа, $Z_0 < g < \max_{\kappa\in j}\{z(K)\}$, J — множество клеток сети R0, лежащих в W. Назовем g_+-клетками те клетки сети $R_1 = R(W,\Delta x,\Delta y)$, среднее значение z на которых больше g. Пусть $C(W,z,\Delta x,\Delta y)$ — число всех g_+-клеток на W, $l(W,z,\Delta x,\Delta y)$ — сумма количеств компонент связности g_+-клеток на каждой строке сети R_1; $c_1 > 0$; $P = P(W,\Delta x,\Delta y,g,c_1,l_1)$ — вероятность, что при случайном распределении c_1 g_+-клеток на сети W по равномерному закону выполнится: $l(W,z,\Delta x,\Delta y) < l_1$.

Лемма. Пусть O — отрезок, разбитый на M равных частей, m из которых — черные, а остальные — белые. Предположим, что M достаточно велико, а белые клетки распределены между M позициями по равномерному закону. Обозначим через $P(M,m,n_1)$ вероятность, что число n_1 «белых» промежутков на отрезке O не больше $n > 0$. Тогда имеем: $P(M,m,n_1) < n/N$, где $N = M/(\gamma+n)$, $\gamma = (1\text{-}m/M)\times\sum_{i=1}^{M-1}[(m/M)^i]$, $n = m/M \sum_{i=1}^{M-1} [(1\text{-}m/M)^i]$. Утверждение леммы следует из оценки средних длин m и n «черных» и «белых» соответственно промежутков; среднего числа N белых промежутков (без учета краевого эффекта) и применения неравенства Чебышева к случайной величине 1/N и числу $1/n_1$.

Замечание. Подсчитав $c_1 = C(W,z,\Delta x,\Delta y)$ и $l_1 = l(W,z,\Delta x,\Delta y)$, можем оценить с помощью леммы $P(W,\Delta x,\Delta y,g,c_1,l_1)$; то есть вероятность выпадения l_1 при отсутствии для функции z на сети R_1 структурного порядка (СП) $(\Delta x,\Delta y,g)$. Назовем P критерием K достоверности СП.

Задача 1. При некотором $d > 0$ имеется набор пар вида $N = \{(A_i,B_i)|A_i\text{-Ro};\ B_i = Rh(A_i,d);\ Z(B_i) > Z_0\}_{i\in 1,k}$. Назовем дислокацией (Д) на произвольной $Rh(A_i,d)$ область, на которой терпят скачки функция z или ординаты ветвей границы Д. Известно, что (1) Д локализуются на линиях тектонических разрывов (ЛТР), причем ЛТР делят П на области относительно постоянных порядков структурности по z; (см. определение 2 и замечание); (2) углы наклона звеньев ЛТР изменяются в определенных диапазонах в заданном порядке. Требуется выделить ЛТР.

Алгоритм А выделения СП на строке S сети R0. Начав с крайней слева клетки строки S, наращиваем на каждом шаге текущую подстроку Z строки S по одной клетке, выбирая из СП на Z, «неслучайных» по критерию K, СП с наибольшим Δx. Процесс наращивания подстроки Z завершаем, как только значение критерия K начнет увеличиваться. Запомнив выделенную подстроку Z, вычитаем ее из S и повторяем процесс, пока S не окажется пренебрежимо малой. Назовем выделенные подстроки р-строками.

Алгоритм В корреляции р-подстрок соседних строк сети R0 C1 и C2. Перебираем пары соседних р-подстрок строк C1 и C2 слева направо, занося пары «с общим СП» в список S(C1). Обозначим процесс применения алгоритма А к j-ой снизу строке сети R0 через A(j); процесс применения алгоритма В к j-ой и (j-1)-ой снизу строкам сети R0 через B(j).

Алгоритм С выделения СП на сети R0. Основание итерации. A(1); j = 2. Шаг итерации (ШИ). A(j); B(j); по результатам B(j) наращиваем списки р-подстрок с общим СП, составляющих односвязные области; j = j+1. Если j-я снизу строка R0 — ее верхняя строка, завершим С; иначе переходим на начало ШИ.

Алгоритм выделения ЛТР. Выполняем алгоритм С применительно к сети R0 и функции z. Находим множество M граничных клеток выделенных СП. Так как искомые ЛТР лежат на границах СП, ищем их на объединении клеток из M, отслеживая по клеткам из M методом ветвей и границ с учетом ограничений на форму ЛТР и постоянства СП между соседними ЛТР.

Т. Д. Гутман

К ПРОБЛЕМЕ КЛАССИФИКАЦИИ В ПАРАМЕТРИЧЕСКИХ ПРОСТРАНСТВАХ ПО КРИТЕРИЮ БЛИЗОСТИ КЛАССОВ ПЛОСКОСТЯМ ЗАДАННОЙ РАЗМЕРНОСТИ

Определение 1. Пусть $M \subset R^n$, $|M| < \infty$, $T \in M$. Тогда ro(T,M) — наибольшее значение для радиуса l шара O(l,T) с центром в T, при котором шар O(l,T) не содержит точек из M\T.

Определение 2. Пусть $M \subset R^n$, $|M| < \infty$, $e > 0$. Обозначим через sk(M,e) набор точек из M, полученный по итеративному алгоритму A: на каждом шаге A из M вычитается точка S, для которой ro(S,M) минимально. A завершается, как только $ro(S,M) < e$.

Обозначение 1. Пусть $M \subset R^n$, $|M| < \infty$. $T \in M$. Тогда $L = L(T,n,|M|)$ — набор прямых, составляющий равномерную достаточно плотную (в данном рассмотрении) сеть на множестве осей, пересекающих начало координат. Пусть $l \in L$. Тогда d(T,M,l) — среднее арифметическое для модулей разности проекций точки T и точек из M\T на ось l; J(T,M) — значение $l \in L$, для которого d(T,M,l) минимально:

$$d\big(T, M, J(T, M)\big) = \min_{l\in L(T,n,|M|)} d(T, M, l).$$

n(T,M) — среднее арифметическое для модулей разности проекций T и всевозможных точек из M\T на оси из L(T,n,|M|).

$$\mathrm{Disp}(T, M) = \sum_{l\in L(T,|M|,n)} \frac{\big[n(T, M) - d(T, M, l)\big]}{L(T, n, |M|)}.$$

Теорема 1. Пусть M локализовано около некоторой гиперплоскости R размерности $k<n$, $z = \sin(360(n\text{-}1))/|L|$, $D = D(T,M,J(T,M)))\text{-}n(T,M)|/n(T,M)\text{-}[(2/p\text{-}\sin(z))/(2/p)]$. Тогда