

# 判定潜在震源的模式识别方法的改进

马尔曼

(国家地震局兰州地震研究所)

## 摘 要

本文给出了在非均匀网格划分的情况下,利用改进的连续Hamming方法判定潜在震源的方法和计算过程。实际计算结果表明,用此方法计算,避免了由于潜在震源形状和空间参数分配所带来的一些问题,更适用于各种形状的构造和各种形状的网格。

## 一、引 言

1981年,А.Д.Гвиццани等人<sup>[1]</sup>开始将Hamming模式识别法用于对易发强震的地点的判定。以后,吕宏伯等<sup>[2]</sup>将这种方法作了改进,使之适用于连续数据,并将其应用于地震危险性区划中。王碧泉等<sup>[3]</sup>进一步改进了连续的Hamming方法,并将其应用于对潜在震源的判定中。本文进一步作了改进,对样本的划分采用非均匀网格,使这一方法更适用于一般的情况,并用此方法对玉门及邻近地区的潜在震源进行了判定。

## 二、对传统方法的改进

### ——非均匀网格划分

用模式识别方法进行潜在震源判定时,首先要把工作区划分成样品网格。划分时,尽量与地质构造方向靠近,网格的大小和形状完全一样,即属均匀网格。这样的网格划分存在着许多缺点:

(1)不能充分地利用地质、地球物理和地震的资料。如在地震地质工作中,我们已经知道某些区域断层分布明显且活动强烈,有历史强震记录,这样的区域一般来说是潜在震源区。我们所划分的样品网格,应基本上与这些资料相吻合,而使用均匀网格,要做到这一步是有一定困难的。

(2)使得确定潜在震源形状的误差增大。由于我们工作的区域内不可能完全是同一方向的地质构造,如有一潜在震源区可穿过几个网格,但其面积仅相当于均匀网格的0.5—1倍,这样就有可能将这个区漏判,或者使其面积人为地增大或缩小。

(3)在危险性分析计算中,空间不均匀值 $v$ 难以分配。一个潜在震源区穿过几个网格

或一个网格被几个潜在震源区所共有, 这样,  $v$ 值的分配有困难。

鉴于以上原因, 我们采用了非均匀网格划分。划分的原则如下:

按活断层的空间分布、走向、长度、宽度及活动速率, 历史地震的震中分布、活动频率及强弱程度和震级大小, 构造应力场等资料, 来确定网格的形状、走向及大小。这样, 能充分利用比较直观的资料, 避免了盲目性、形状不可改变及资料浪费的缺点。因此, 采用非均匀网格, 可以灵活、准确地表现这些已知信息的特点, 便于以后的计算。

在非均匀网格的情况下, 由于每个样品的大小不同, 对测量后的特征参数要做进一步处理, 即把与面积有关的参数(如连续数据)要化为单位面积上的数据, 而与面积无关的参数(如二值数据)不做任何处理, 使得参与运算的特征参数有统一的标准。

对于空间分布函数  $f_{i,M_j}$ , 在非均匀网格的情况下, 由于每个样品的面积不一样, 应该用面积比来加权。如第 I 类区有  $n$  个潜在震源, 那么权  $W_i$  为

$$W_i = \frac{S_i}{\sum_{j=1}^n S_j}, \quad (1)$$

则  $F_{i,M_j} = W_i |T - d_i|$ 。 (2)

式中  $S_i$  为第  $i$  个潜在震源的面积,  $T$  为阈值,  $d_i$  为第  $i$  个潜在震源的 Hamming 距离。

将  $F_{i,M_j}$  归一化后就为  $f_{i,M_j}$ , 这样处理后, 说明对于同一级潜在震源, 如果  $|T - d_i|$  相等, 则面积大者危险性大, 这样更切合实际。

### 三、用改进的连续 Hamming 方法对玉门地区潜在震源的判定

#### 1. 特征参数的选择

在玉门潜在震源的判定中, 选择了如下特征<sup>1)2)</sup>:

判定  $M \geq 6$  级地震的潜在震源的标志:

(1) 地质构造特征 (特征总数  $m = 7$ )

- $x_1$ : 是否有现代主要活断层;
- $x_2$ : 是否有全新世活断层;
- $x_3$ : 区域性活断层数目;
- $x_4$ : 区域性活断层长度;
- $x_5$ : 断层活动速率是否较高;
- $x_6$ : 是否有剪切走滑并兼逆冲的断层;
- $x_7$ : 是否处于现代活断层与北北西向隆起相交的部位。

(2) 地震活动特征 (特征总数  $m = 5$ )

- $x_1$ : 是否处于地震活动区中;
- $x_2$ : 是否处于 6 级地震活动的条带中;
- $x_3$ : 是否处于 6 级地震活动条带中的 6 级地震空段上;

1) 王碧泉等, 玉门市地震活动性分析及其模式识别方法划分潜在震源区, 1988.

2) 陈丙午等, 玉门油田抗震防灾规划基础研究报告, 1988.

- $x_4$ : 中强地震 ( $M=4.0\sim 5.9$ ) 的频度;  
 $x_5$ : 中强地震 ( $M=4.0\sim 5.9$ ) 的能量和 (尔格)。

(3) 重力场特征 ( $m=5$ )

- $x_1$ : 重力异常最小值 (毫伽);  
 $x_2$ : 最大重力异常梯度值 (毫伽/km);  
 $x_3$ : 最大莫氏面梯度值 (km/10km);  
 $x_4$ : 是否处于莫氏面等值线转折部位和斜坡带;  
 $x_5$ : 深部断裂状况 (无断裂为 0; 0 至 50km 内的断裂取值为 1; 50 至 100km 断裂取值为 2; 100km 以上的断裂取值为 3; 遇有岩石圈断裂则为 2)。

(4) 磁场特征 ( $m=3$ )

- $x_1$ : 最大航磁梯度值 (nT/km);  
 $x_2$ : 构造交汇点数;  
 $x_3$ : 是否处于磁异常等值线轴向发生突变的部位上。

判定  $M \geq 7$  级地震的潜在震源的标志:

(1) 地震活动性特征 ( $m=3$ )

- $x_1$ : 是否处于 7 级地震活动的条带中;  
 $x_2$ : 是否处于下次 7 级地震的迁移区中;  
 $x_3$ : 是否有 7 级地震的减震作用。

(2) 地质构造特征 ( $m=3$ )

- $x_1$ : 是否有长度大于 60km 的全新世断层活动段;  
 $x_2$ : 全新世以来断层活动年滑动速率是否大于等于 2 mm;  
 $x_3$ : 是否处于剪切走滑并兼逆冲的断裂带上。

## 2. 模式识别方法

本文所采用的模式识别方法为王碧泉等<sup>[8]</sup>所提出的改进的连续 Hamming 方法, 其基本内容为:

取全部样品的集合  $\Omega$ , 分为 D 类和 N 类, 两类中用于训练 (学习) 的集合为  $D_0$  和  $N_0$ , 两个集合中所含样本个数分别为  $N_1$  和  $N_2$ 。每一个样本  $X_i$  可视为一个  $n$  维向量, 由  $n$  个连续数据所描述。 $x_{ij}$  表示第  $i$  个样本的第  $j$  个特征量。

$$X_i = (x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{in})^T, \quad i=1, 2, \dots, N; j=1, 2, \dots, n. \quad (3)$$

由  $n$  个特征值所描述的  $N$  个样品是一个  $N \times n$  的资料矩阵。

连续数据特征标准化为  $x'_{ij}$ ,

$$x'_{ij} = (x_{ij} - \bar{x}_j) / S_j, \quad i=1, 2, \dots, N; j=1, 2, \dots, n. \quad (4)$$

其中  $\bar{x}_j = \left( \sum_{i=1}^N x_{ij} \right) / N$ ;  $S_j = \sqrt{\sum_{i=1}^N (x_{ij} - \bar{x}_j)^2 / (N-1)}$ 。

分别计算  $D_0$  类和  $N_0$  类每一特征的均值:

$$\begin{cases} \bar{x}_j D_0 = \left( \sum_{X_i \in D_0} X_{ij} \right) / N_1 \\ \bar{x}_j N_0 = \left( \sum_{X_i \in N_0} X_{ij} \right) / N_2 \end{cases} \quad (5)$$

$j = 1, 2, \dots, n$

和每一样品的特征方差之和：

$$\begin{cases} S^2(D_0) = \sum_{i=1}^n S_i^2(D_0) = \sum_{i=1}^n \left[ \sum_{X_i \in D_0} (x_{ij} - \bar{x}_j D_0)^2 \right] / (N_1 - 1) ; \\ S^2(N_0) = \sum_{i=1}^n S_i^2(N_0) = \sum_{i=1}^n \left[ \sum_{X_i \in N_0} (x_{ij} - \bar{x}_j N_0)^2 \right] / (N_2 - 1) . \end{cases} \quad (6)$$

若 $S^2(D_0) \leq S^2(N_0)$ ，则计算 $D_0$ 类样品的Hamming核；若 $S^2(D_0) > S^2(N_0)$ ，则计算 $N_0$ 类样品的Hamming核。

Hamming核  $Z = (z_1, z_2, \dots, z_j, \dots, z_n)^T$  . (7)

其中 
$$z_j = \begin{cases} \bar{x}_j D_0 , & \text{当 } S^2(D_0) \leq S^2(N_0) , \\ \bar{x}_j N_0 , & \text{当 } S^2(D_0) > S^2(N_0) . \end{cases}$$

利用加权的欧氏距离计算各样品到Hamming核的距离

$$d_i(X_i, Z) = \left[ \sum_{j=1}^n W_j (x_{ij} - z_j)^2 \right]^{\frac{1}{2}}, \quad i = 1, 2, \dots, N \quad (8)$$

式中 $W_j$ 为权系数。

$$W_j = |\bar{x}_j D_0 - \bar{x}_j N_0| / (\max |\bar{x}_j D_0 - \bar{x}_j N_0| + 0.5) \quad (9)$$

最后进行分类判别。对取定的阈值 $T$ ，任一样品 $X_i$ 的分类标准如下：

对 $S^2(D_0) \leq S^2(N_0)$ 的情况

$$\begin{cases} X_i \in D , \\ X_i \in N , \end{cases} \quad \text{当} \quad \begin{cases} d_i(X_i, z) \leq T , \\ d_i(X_i, z) > T . \end{cases} \quad (10)$$

对 $S^2(D_0) > S^2(N_0)$ 的情况

$$\begin{cases} X_i \in N , \\ X_i \in D , \end{cases} \quad \text{当} \quad \begin{cases} d_i(X_i, z) \leq T , \\ d_i(X_i, z) > T . \end{cases} \quad (11)$$

### 3. $M \geq 6$ 级地震的潜在震源的判定

#### (1) 样品划分

根据活断层的走向、长宽度、活动速率及地震活动性资料，将工作区划分为99个单元，即 $N = 99$ ，见图1。

样品类别的划分原则是，将位于 $M = 6 \sim 6.9$ 级地震震中周围10km范围内，位于 $M \geq 7$ 级地震震中周围20km范围内的网格样品定为D类，其余为N类。根据这个原则，工作区内共有D类样品22个，N类样品77个。

#### (2) 特征的提取

在每一样品内，按前面所述的 $M \geq 6$ 级地震的潜在震源的标志，逐一提取特征数

$x_{ij}$ , 得到连续数据或二值数据, 对N个样品得到n个特征值所组成的 $N \times n$ 阶资料矩阵, 用以识别。

### (3) 判定结果

根据下面的两条原则, 我们进行分析计算:

未来百年内可能发生 $M \geq 6$ 级地震的潜在震源区(D类样品)具有类似的地质构造、地震活动、重力场及磁场, 即有类比性;

适当减小第一类误差 $\tau_1$ (D类错判为N类), 在 $\tau_1$ 和总误识率

$$\varepsilon = \left( \sum_{i=1}^2 P(\omega_i) \tau_i \right) / N \quad (12)$$

较小的前提下确定阈值T,  $P(\omega_1)$ 和 $P(\omega_2)$ 为D类和N类的先验概率。

计算结果见表1, 潜在震源见图1

表1  $M \geq 6$ 级地震的潜在震源判定结果(阈值 $T = 30.33$ )

区类	编号	Hamming距离 $d_i$	空间分布函数 $f_i; M_i$	
I	25	37.3659	0.2297	新区
I	27	45.6577	0.2194	新区
I	32	42.1902	0.2181	新区
I	33	43.8738	0.3328	
I	34	49.6280	0.0454	
I	35	39.4352	0.0384	
I	39	44.0805	0.0111	
I	40	32.2178	0.0257	
I	41	51.3215	0.0160	
I	42	42.6924	0.0406	
I	44	38.4938	0.0402	
I	47	46.1055	0.0483	
I	48	36.0625	0.0186	
I	49	31.6230	0.0358	新区
I	50	39.2919	0.0293	
I	51	41.2871	0.0180	
I	52	41.9645	0.0260	
I	53	40.2191	0.0179	新区
I	54	48.9156	0.0238	新区
I	57	41.5201	0.0203	
I	58	34.8494	0.0309	新区
I	59	43.3731	0.0247	新区
I	60	34.2235	0.0226	新区

续表 1

区类	编号	Hamming距离 $d_i$	空间分布函数 $f_i; M_i$	
I	61	46.0005	0.0312	新区
I	64	31.2807	0.0140	
I	65	40.4273	0.0278	
I	66	35.7878	0.0107	新区
I	67	52.9244	0.0300	
I	68	37.7138	0.0257	新区
I	69	42.1678	0.0376	新区
I	71	32.4103	0.0108	新区
I	75	31.7783	0.0142	
I	79	37.5846	0.0668	新区
I	82	39.9628	0.0141	新区
I	86	37.0524	0.0190	新区
I	87	34.7128	0.0464	新区
I	88	36.6508	0.0287	
I	92	41.6939	0.0390	新区
I	93	41.8623	0.0167	新区
I	94	65.3319	0.0338	

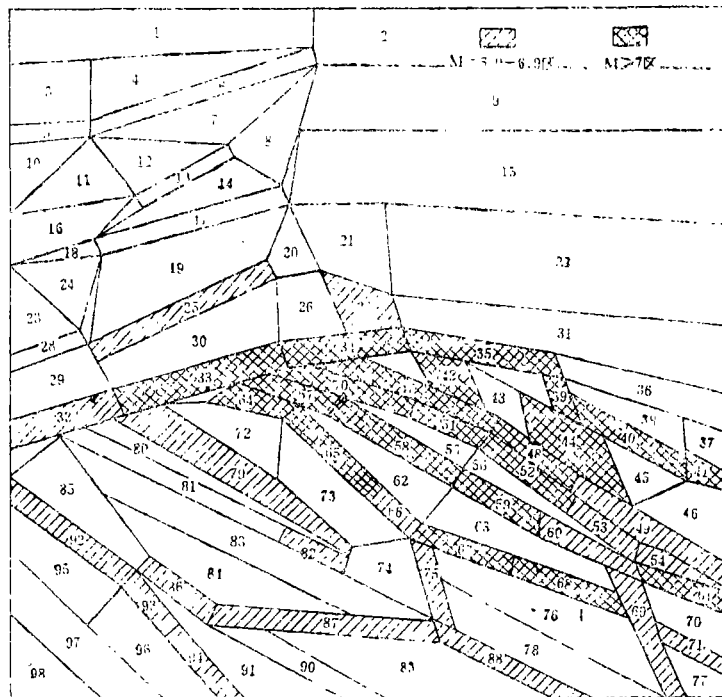


图1 潜在震源判定图  
Fig. 1 Delineating of potential seismic source areas

#### 4. $M \geq 7$ 级地震的潜在震源的判定

$M \geq 7$  级地震的潜在震源的判定步骤和  $M \geq 6$  级地震的完全一样。对样品类别划分原则是, 位于  $M \geq 7$  级地震震中周围 20km 范围内的样品网格为 D 类。其余为 N 类。这样其中 D 类样品 13 个, N 类样品 86 个。特征参数按前面所述  $M \geq 7$  级地震的潜在震源区的标志提取, 判定结果见表 2 和图 1。

为了对比, 这里也给出均匀网格的识别结果, 如图 2 所示<sup>2)</sup>。

表 2  $M \geq 7$  级地震的潜在震源判定结果 ( 阈值  $T = 26.55$  )

区类	编号	Hamming 距离 $d_i$	空间分布函数 $f_{i, M_i}$	
I	33	56.0719	1.0000	
I	34	49.3630	0.0915	
I	35	27.6510	0.0545	新区
I	39	33.5714	0.0172	新区
I	40	33.5174	0.0542	
I	41	52.8729	0.0334	
I	42	38.3657	0.0812	
I	44	38.3657	0.0812	
I	47	30.4598	0.0650	
I	48	30.4598	0.0319	
I	50	56.0719	0.0848	
I	51	32.1611	0.0285	新区
I	52	27.6510	0.0348	
I	57	56.0719	0.0757	
I	58	38.3657	0.0689	新区
I	59	27.6510	0.0319	新区
I	61	27.6510	0.0258	新区
I	64	32.1611	0.0291	新区
I	65	27.6510	0.0385	新区
I	67	27.6510	0.0317	新区
I	68	27.6510	0.0381	新区

## 四、小 结

利用 Hamming 方法进行潜在震源的判定, 避免了人为的主观因素的影响, 使这一定性的事件定量化。因此, Hamming 方法是潜在震源判定的一种重要方法。本文继承了前人的

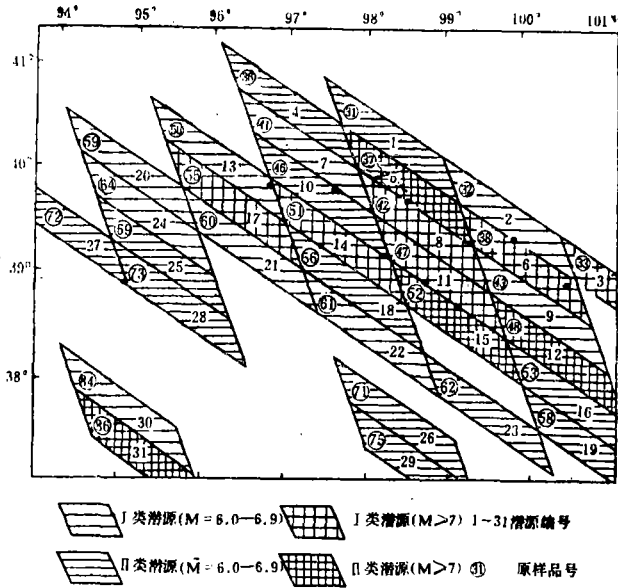


图2 潜在震源区分布图

Fig. 2 Distribution of potential seismic source areas

研究成果，并作了适当的改进，使得样品划分更灵活、更精确，计算时可以充分利用直观而且是最基本的资料，计算结果便于在地震危险性分析中应用。

本文是陈丙午研究员的指导和帮助下完成的。陈丙午研究员还对原稿进行了修改，对计算结果进行了检验、分析，在此表示衷心的感谢。

( 本文1988年12月30日收到 )

参 考 文 献

[1] Гвишиани, А. Д., В. Г. Кособков, К. Обоснованию, Результатам прогноза мест сильных Землетрясений, Полученных Методами Распознавания, Физика Земли, No. 2, 21-35, 1981.  
 [2] 吕宏伯、聂金宗、陈祖荫、马秀芳、王碧泉, Hamming分类方法的改进及其在地震危险区划中的应用, 地震学报, 增刊, 143-153, 1986.  
 [3] 王碧泉、王春珍, 对连续Hamming方法的改进及其在潜在震源判定中的应用, 地震学报, Vol.10, No. 2, 113-123, 1988.



AN IMPROVEMENT ON THE PATTERN RECOGNITION METHOD  
OF DISTINGUISHING POTENTIAL SEISMIC SOURCE AREAS

Ma Erman

(*Earthquake Research Institute of Lanzhou, SSB, Lanzhou, China*)

Abstract

Under the non-uniform grid, method and calculating process distinguishing potential seismic source are proposed using the improved consecutive Hmmaing method. The calculating results show that this method can avoid some problems caused by the shape of potential seismic source area and the distribution of spatial parameters, and it is suited to different structures which have different strikes and different grid shapes.

1989年9月21日肃北4.9级地震\*

根据甘肃区域地震台网和全国地震台网测定,这次地震的基本参数是:

发震时刻:1989年9月21日0时21分58.1秒;

震中位置:39°07'N, 97°16'E, 即肃北蒙古族自治县东部境内;

震级:  $M_s = 4.9$ ;

震源深度:24km。

通过实际考察确定,宏观震中位于微观震中以南10km的野马南山东端,极震区烈度约为V度。在北侧IV度区内的布鲁斯沟和东北侧的甘沟硫磺矿,人们普遍被摇醒,听到门窗和家俱响动及轰隆隆的地声。哈勒乌苏矿点的简易公路边有几处小的陡崖崩塌现象。其III度区东起镜铁山矿区,北经昌马以南(昌马无感),西止于大岔道班,长约200km,区内少数人有感。在距震中东北约80km的玉门市区,由于地处山前地带,加之楼房较多,人们感到摇晃强烈,其烈度可达IV度,为一异常烈度点。

地震前,距震中110km的嘉峪关地震台的地磁,从8月8日起Z分量出现异常,幅度达10伽马。地电阻率东西道从7月24日起出现正异常,幅度达1%。气氛从9月2日起出现负异常。根据上述异常,该台于9月7日和14日曾两次提出“月底前,在距台站300km范围内,可能发生4.5级左右的地震”的预报意见。

(国家地震局兰州地震研究所 高元勤)

THE XIAOBEI EARTHQUAKE (M=4.9) ON SEPT. 21, 1989

Gao Yuanqin

(*Earthquake Research Institute of Lanzhou, SSB, Lanzhou, China*)

\*参加此次地震考察的有高元勤、毛可、赵和云、耿耀明、秦泽群、沈树兴、张俊林、周国民、邵玉安等人。