

# $M \geq 6.0$ 强震后地震序列 类型的模糊聚类分析

曾健 韩渭宾 王虹 席敦礼  
(四川省地震局, 成都 610041)

**摘要** 在充分了解地震序列特征的基础上, 选出 26 项地震序列的原始特征。根据这些原始特征的 Wallen 评分, 精选出 11 项能突出表现地震序列的特征参数作为指标。采用 ISODATA 模糊聚类方法对地震序列进行分类, I 类为  $M \geq 6.0$  强震后三个月内还有震级为  $M \geq 5.0$  的强震发生; II 类为  $M \geq 6.0$  强震后三个月内无  $M \geq 5.0$  的地震。对中国大陆地区 1966 年以来的 22 次  $M \geq 6.0$  强震序列进行聚类分析, 建立预报震后趋势判别模型, 然后以 8 次强震序列为例进行了外推检验。结果表明, 该方法对地震序列类型的判别效果较好。

**关键词:** 地震序列 特征参数 模糊聚类分析 判别模型

## 1 引言

在一次强震发生后, 短期内还有没有较强地震发生, 这是大震现场分析工作中最重要、最迫切的问题之一。震后趋势判断实际上就是在强震发生后的短时间内, 对地震序列的类型作出判断。人们已研究过不少判别地震序列类型的方法, 如  $h$  值、 $b$  值、等待时间及  $k$  值等。应用上述方法大多需要积累一定时段的地震序列资料, 而震后趋势判断又是一项十分迫切的任务, 因此上述方法在一定程度上满足不了实际工作的需要。为了使地震序列类型的划分具有预报价值, 我们曾利用一次中强以上地震 ( $M_0 \geq 4.7$ ) 发生后头三天的地震序列资料, 应用模式识别、模糊数学方法对未来三个月内是否可能发生相当 ( $M = M_0 \pm 0.3$ ) 或更大 ( $M > M_0$ ) 地震作过综合判定<sup>[1],[2],[3]</sup>。本文在此基础上用模糊聚类的 ISODATA 方法<sup>[4]</sup>对地震序列进行分析, 建立判别模型, 试图为大震现场的分析预报提供一项手段。

## 2 资料

### 2.1 样本

设某地发生了一个  $M_s \geq 6.0$  的地震, 记震级为  $M_0$ , 发震时刻为  $T_0$ , 令  $T_1 = [T_0, T_0 + 3 \text{ 天}]$ ,  $T_2 = [T_0 + 3 \text{ 天}, T_0 + 3 \text{ 天} + 3 \text{ 月}]$ , 本文利用  $T_1$  时段内的震级-时间序列资料对  $T_2$  内是否发生  $M \geq 5.0$  的地震作分类预测, 规定  $T_2$  时段内有  $M \geq 5.0$  地震的地震序列为 I 类序列, 否则, 为 II 类序列。选取了自 1966 年以来在中国大陆地区发生的 30 次  $M \geq 6.0$  地震序列资料为样本, 对地震序列进行模糊聚类的分析判别。以其中的 22 个序列(表 1)的 11 项特征参数为指标(表 2)。

### 2.2 指标

根据对地震序列特征的认识, 从不同角度的物理背景选出 26 项地震序列的原始特征。

根据这些原始特征的 Wallen 评分

$$V = \frac{n_1 n_0^0 - n_0 n_1^0}{\sqrt{N_1 N_0 N^1 N^0}} \quad (1)$$

计算出各单项原始特征的 I 类标准及其最大 Wallen 评分, 精选出 Wallen 评分较高又相对独立的 11 项原始特征作为聚类分析的指标, 这 11 项指标的含义及其最大 Wallen 评分值列于表 2。由 22 个已知样本的 11 项指标, 采用迭代法, 以使单项 V 评分达到最大为标准, 求出各指标的 I 类标准(表 2), 以此作为初分类的依据。由于样本统计指标原始数据存在着量纲上的差异和数量上的差别, 还不能直接用于聚类分析, 必须把各指标标准化。即对指标进行归一和加权处理, 把指标变换到 [0, 1] 闭区间里, 并对能突出表现样本基本特征的指标给予较大的权数。

表 1 建模所用的 22 个地震序列

编号	地名	日期	北纬	东经	震级	类别
01	邢台	1966-03-08	37°21'	114°55'	6.8	I
02	通海	1970-01-05	24°06'	102°36'	7.7	I
03	松潘	1976-08-16	32°36'	104°06'	7.2	I
04	宁夏	1976-11-07	27°36'	101°06'	6.7	I
05	乌恰	1983-02-13	40°08'	75°20'	6.8	I
06	澜沧	1988-11-06	22°50'	99°43'	7.6	I
07	巴塘 2	1989-04-25	29°58'	99°24'	6.7	I
08	中甸	1966-09-28	27°30'	100°06'	6.4	I
09	河间	1967-03-27	38°30'	116°30'	6.3	I
10	炉霍	1967-08-30	31°37'	100°20'	6.3	I
11	盐津	1969-07-26	21°43'	111°45'	6.4	I
12	普洱	1970-02-07	23°05'	101°02'	6.2	I
13	珂坪	1972-01-16	40°20'	79°04'	6.2	I
14	松潘	1973-08-11	32°54'	104°06'	6.5	I
15	巴音木仁	1976-09-23	39°50'	106°30'	6.2	I
16	库车	1979-03-29	41°54'	83°27'	6.0	I
17	溧阳	1979-07-09	31°27'	119°15'	6.0	I
18	五原	1979-08-25	41°12'	108°03'	6.0	I
19	道孚	1981-01-24	31°00'	101°10'	6.9	I
20	甘孜	1982-06-16	31°50'	99°45'	6.0	I
21	小金	1989-09-22	31°33'	102°23'	6.8	I
22	古浪	1990-10-20	37°06'	103°30'	6.2	I

### 3 方法

确定样本和指标之后, 采用 ISODATA 模糊聚类方法进行分类判别。由于以往的基于模糊等价关系进行的系统聚类法在使用时, 常常得到待预报样本独立成为一类, 而使得预报意见无所依据<sup>[3],[5],[7]</sup>, 本方法以样本从属函数和各指标的聚类中心为依据, 采用预先给定分类类别数的办法, 使预报样本总要进入某一类, 从而克服了上述弊病。

表2 所用特征及其最大 Wallen 评分值

序号	符号	含 意	I 类标准	Vmax
1	NT <sub>21</sub>	第二天的地震个数/第一天的地震个数	>0.20	0.53
2	NT <sub>31</sub>	第三天的地震个数/第一天的地震个数	>0.13	0.62
3	H	每4小时一个间隔,最小二乘法求h值	>0.03	0.42
4	M <sub>1</sub> -M <sub>3</sub>	最大震级-第三大震级	≤2.50	0.42
5	M <sub>1</sub> -M <sub>4</sub>	最大震级-第四大震级	≤2.70	0.42
6	MD <sub>2</sub>	最大震级-第二天的最大震级	≤3.20	0.54
7	EB <sub>3</sub>	第三天的最大能量/第三天的能量和	≤0.88	0.52
8	E <sub>21</sub>	(第二天的能量和/第一天的能量和)×10	>0.01	0.46
9	E <sub>31</sub>	(第三天的能量和/第一天的能量和)×10	>0.03	0.43
10	B	最小二乘法求b值	>0.38	0.57
11	M <sub>1</sub>	最大震级	>6.60	0.76

### 3.1 软分类

通常情况下所用的分类法被称为硬分类,因为它的分类矩阵的元素(样本的从属度):

$$U_{ik} = \begin{cases} 0 & \text{样本 } X_i \text{ 不属于 } k \text{ 类} \\ 1 & \text{样本 } X_i \text{ 属于第 } k \text{ 类} \end{cases} \quad (2)$$

这样,每一样本只能属于某一类,而且一定属于其中的一类。对于有些问题,常常不应认为样本一定就属于某一类而不属于其它任何类,换句话说,样本从某些特征考虑应属于这一类,而从另外一些特征看又好象属于另一类,即认为样本以某种从属程度从属于这一类,而又以另一种从属程度从属于另一类。这样一来,每一类就都是样本集上的一个模糊子集。这种分类称为软分类。

对于样本集

$$X = \{x_1, x_2, \dots, x_n\} \quad (3)$$

若要将其分为  $c$  类,则它每一个分类结果都对应一个  $c \times n$  矩阵  $U$ ,该模糊分类矩阵  $U$  应满足下述条件:

- (1)  $U_{ik} \in [0, 1]$ ,即矩阵元素(从属度)在 0 与 1 之间取值;
- (2)  $\sum_{i=1}^n U_{ik} = 1$ ,即每列元素和为 1,对一个样本而言,它对各类别的从属度之和为 1;
- (3)  $\sum_{k=1}^c U_{ik} > 0$ ,这一条件保证了每一类不空。

显然,对应样本集  $X$  的任一种  $c$  组软分类,都有一个模糊矩阵  $U$  与之对应;反之,任一满足上述条件的矩阵  $U$  也都对应着样本集  $X$  上的一种  $c$  组软分类。

用  $Mfc$  表示所有矩阵  $U$  的集合,即

$$Mfc = \{U \in V_{c \times n} \mid U_{ik} \in [0, 1], \forall i, k; \sum U_{ik} = 1, \forall j; \sum U_{ik} > 0, \forall i\} \quad (4)$$

我们称  $Mfc$  是  $X$  的  $c$  组软分类空间<sup>[4]</sup>。

为了获得一个合理的软分类,我们必须按某种分类准则,从  $Mfc$  中挑出一个最好的软分类,聚类准则如下。

### 3.2 聚类准则与聚类判据

按泛函理论,分类中采用的聚类判据为求出  $U, V$ , 并使

$$J_m(U, V) = \sum_{k=1}^n \sum_{i=1}^c (U_{ik})^m \|X_k - V_i\|^2 \quad (m \geq 1) \quad (5)$$

所表示的泛函达到最小。式中  $U$  为各样本属于各类别的从属函数矩阵,  $V$  为聚类中心矩阵, 每个元素有  $J$  个分量,  $X_k$  为第  $k$  个样本, 有  $j$  个特征值(指标)。  $n$  为样本数,  $c$  为要分的类别数。

对于(5)式,一般取  $m > 1$ , 当  $m = 1$  时就蜕化为硬分类, 从直观的几何意义上理解为各类样本到相应聚类中心距离的平方和为最小。  $m = 2$  时, 从物理意义上讲, 相当于求解一个最小电阻网络, 即在某个精度条件下, 比较各样本指标后, 矛盾最小的一种分类。参数  $m$  越大则分类越模糊, 它的意义也更不明确<sup>[4],[8],[9]</sup>。一般说来, 本问题的求解相当困难, 不过在  $m > 1, X_i \neq V_k, V_{i,k}$  时可用下面的式子来计算  $U_{ik}$  和  $V_i$ :

$$U_{ik} = \frac{1}{\sum_{j=1}^c \left( \frac{\|X_k - V_i\|}{\|X_k - V_j\|} \right)^{2/(m-1)}} \quad \begin{array}{l} i=1, 2, \dots, C \\ k=1, 2, \dots, N \end{array} \quad (6)$$

$$V_i = \frac{\sum_{k=1}^n (U_{ik})^m X_k}{\sum_{k=1}^n (U_{ik})^m} \quad i=1, 2, \dots, C \quad (7)$$

具体算法如下:

第一步: 任选一  $U \in M_{fc}$  (即人为地先将样本进行一次初分类)。

第二步: 据已知的  $U$  及式(6)计算聚类中心  $V_k (k=1, 2, \dots, c)$ 。

第三步: 据  $V_k$  及式(5)计算出新的分类矩阵  $U^*$ 。

第四步: 若  $\text{MAX}[|U_{ik}^* - U_{ik}|] < \epsilon$ , 则  $U_{ik}^*$  及相应的  $V_{ik}^*$  即为所求, 否则再回到第二步, 进行又一次迭代过程。  $\epsilon$  为预先指定的一个微小正数,  $\epsilon$  越小, 结果越精确, 但计算量也越大。

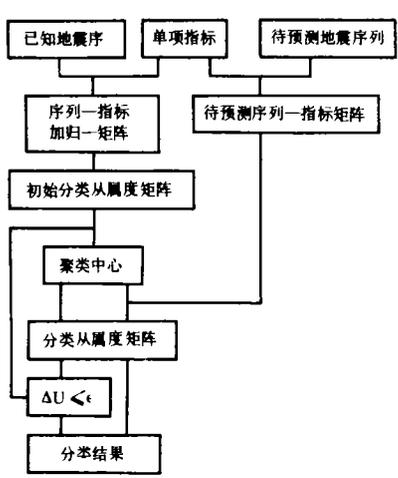
## 4 结果与分析

### 4.1 研究思路

根据前述地震序列资料和 ISODATA 模糊聚类方法, 我们提出了将模糊聚类方法用于震后趋势早期判断的研究思路(图 1)。

### 4.2 建立预报模型

首先对 22 个已知序列地震样本进行聚类分析, 建立预报模型。①输入已知地震序列的样本数和指标数, 原始单项指标, 分类类别数和聚类精度  $\epsilon$ ; ②对第  $i$  个样本的每一个样本指标都用  $\sum r_{ij}$  去除, 即用  $r_{ij}/\sum r_{ij}$  代替原来的  $r_{ij}$ ; 使指标归一化, 并以各指标的  $V_{\max}$  为权数对各指标加权, 从而建立序列-指标加权归一矩阵; ③以各样本中符合 I 类标准的指标占指标总数的百分比作为该样本从属于 I 类的从属度, 建立初始分类的从属度矩阵; ④以(5), (6)两式求得聚类中心矩阵  $VX(K, J)$ ; ⑤对样本进行分类, 给出其分类从属度矩阵  $A(I, K)$ ; ⑥检验分类误差  $\Delta U$  是否小于给定的精度  $\epsilon$ , 若  $\Delta U \geq \epsilon$ , 则返回到第④步进行新的一次迭代过程, 直至  $\Delta U \leq \epsilon$  为止; ⑦输出, 打印样本属于各类别的从属度矩阵, 其中, 初始分类从属度(迭



代初值)为:

$$U_{ik}^0 = \frac{\sum_{j \in k} (V_{\max})_{ij}}{\sum_{j=1}^p (V_{\max})_{ij}} \quad \begin{matrix} i=1, 2, \dots, N \\ k=1, 2, \dots, C \end{matrix} \quad (8)$$

初始聚类中心为:

$$V_{jk}^0 = \frac{\sum_{i=1}^n (U_{ik}^0)^m Y_{ij}}{\sum_{i=1}^n (U_{ik}^0)^m} \quad (9)$$

式中,  $Y_{ij}$  为归一化和加权后的样本指标。取迭代精度  $\epsilon = 0.12$ , 对已知地震序列进行聚类分析, 得到如下分类从属度矩阵:

图1 模糊聚类方法研究震后趋势早期判断的框图

Fig. 1 A program sketch of estimating stronger earthquake probability by using Fuzzy cluster method.

0.68	0.76	0.63	0.56	0.70	0.53	0.57	0.48	0.45	0.45	0.49	I
0.44	0.47	0.42	0.39	0.45	0.48	0.44	0.48	0.46	0.45	0.49	
0.32	0.24	0.37	0.44	0.30	0.47	0.43	0.52	0.55	0.55	0.51	II
0.56	0.53	0.58	0.61	0.55	0.52	0.56	0.52	0.54	0.55	0.51	

根据 22 个样本对 I 类和 II 类的从属度, 得到聚类结果: 第 1, 2, 3, 4, 5, 6, 7 号样本属于 I 类, 其余样本属于 II 类。与表 1 中的实际情况比较, 所有样本均识别正确, 与实际情况相符。记下并固定此时的聚类中心矩阵  $XV(K, J)$  及各指标的权数和归一化参数, 以此作为震型判别预报模型的参数。

### 4.3 外推检验

按图 1 所示研究思路, 以处理已知地震序列样本相似的过程和相同的标准, 对表 3 中所列 8 个待预测地震序列样本分别提取 11 项单项指标, 输入已建立好的预报模型进行归一化和加权处理, 从而建立待预测序列-指标矩阵, 用已建立好的聚类中心对该待预测样本进行外推判别。每次只代入一个样本, 分别求出它们属 I 类和 II 类的从属度:

0.5317	0.5155	0.5790	0.5437	0.4087	0.4704	0.4206	0.4182	I
0.4683	0.4845	0.4210	0.4563	0.5913	0.5296	0.5794	0.5818	

即第 1, 2, 3, 4 号样本判为 I 类, 第 5, 6, 7, 8 号样本为 II 类, 与表 2 中实际情况比较, 发现所有样本均判别正确。

## 5 结语

ISODATA 模糊聚类方法与其它聚类方法一样, 不能使单个差劣的指标变优良, 但从各指标携带的模糊信息中综合出一个比较清楚的定量结果。聚类结果受台网疏密、观测条件的历史变迁、资料 and 指标参量本身精度等各种因素的影响。尤其需要说明的是, 将任何一

表3 外推所用的8个地震序列

编号	地名	日期	北纬	东经	震级	类别
1	东川	1966-02-05	26°06'	103°06'	6.5	I
2	炉霍	1973-02-06	31°18'	100°24'	7.6	I
3	昭通	1974-05-11	28°12'	104°06'	7.1	I
4	巴塘1	1989-04-16	29°58'	99°15'	6.7	I
5	渤海	1969-07-18	38°12'	119°24'	7.4	II
6	思茅	1973-08-16	23°06'	101°12'	6.3	II
7	康定	1975-01-15	29°24'	101°51'	6.2	II
8	禄劝	1985-04-18	25°52'	102°51'	6.3	II

种应用数学方法用于预测研究,使用已知样本建模作内符检验能取得较好的统计检验和评分效果,不等于外推一定有好效果。对于事先留下的少数已知样本有较好的外推效果,不一定在长期的预报实践中一定会有稳定的好效果。因此,本文所研究的方法及所确定的各种参数都必须在今后的分析预报实践中接受检验,不断改进和完善。

### 参 考 文 献

- 1 韩渭宾,王虹,曾健,席敦礼. 中强以上地震的震后趋势早期综合判断方法的研究. 地震学报,1993,15(1).
- 2 曾健,韩渭宾,王虹,席敦礼. 震后趋势的模糊聚类分析判别. 四川地震,1991,(3).
- 3 曾健,韩渭宾,席敦礼. 鲜水河地震带和松潘、龙门山地震带可能发生5级以上地震的模糊聚类分析. 四川地震,1988,(3).
- 4 冯德益,等. 模糊数学方法与应用. 北京:地震出版社,1985.
- 5 冯德益,等. 模糊数学和地震预报. 地震,1982,(2).
- 6 冯德益,等. 模糊聚类分析在以震报震研究中的应用. 地震,1982,(4).
- 7 陈荣华,冯德益. 模糊聚类分析及其在地震研究中的应用. 地震研究,1985,8(5).
- 8 顾谨平,虞雪君,钱家栋. 地震活动性的综合评判. 地震学报,1989,11(2).
- 9 虞雪君,等. 地震形势的模糊综合评判. 地震学报,1990,12(3).

## FUZZY CLUSTERING ANALYSIS ON THE TYPES OF EARTHQUAKE SEQUENCE AFTER A $M \geq 6.0$ EARTHQUAKE

Zeng Jian, Han Weibin, Wang Hong and Xi Dunli

(Seismological Bureau of Sichuan Province, Chengdu, China)

### Abstract

Based on the full understanding of characteristics of earthquake sequences, 26 items of original features of earthquake sequences are selected from distinct views of physics background. According to the Wallen scores of these original features, 11 items of original features which can prominently represent the characteristics of earthquake sequences are carefully chosen as indexes. Using ISODATA method of fuzzy clustering, earthquake sequences are classified into two categories: Type—I sequences are these in which earthquakes with  $M \geq 5.0$  occur in the same area within 3 months after an earthquake with  $M \geq 6.0$ ; Type—II sequences are these in which no  $M \geq 5.0$  earthquakes occur within 3 months. The discriminating model of post-earthquake tendency has been built by using 22 known earthquake sequences occurred in the China continent since 1966. Then, extrapolation tests are taken by using other 8 earthquake sequences occurred in China. The results indicate that this method has good effects on discriminating the type of earthquake sequences.

**Key Words:** Earthquake sequence; Characteristic parameter; Fuzzy clustering; Discriminating model