

# 基于神经网络的强震中短期预测方法<sup>\*</sup>

韩志强 王碧泉

(中国北京 100081 国家地震局地球物理研究所)

**摘要** 神经网络(neural network)是由大量并行处理的类似生物神经元的简单单元构成的复杂系统. 通过调整各个单元之间的连接权值, 神经网络可以被训练来表达一个特定的映射. 这种作用是神经网络应用的基础. 近十年来, 神经网络已从理论研究进入实用, 并且这个趋势还在不断地发展. 本文提出了一种基于神经网络模型的强震预测方法. 神经网络先从存在的地震演化序列或地震前兆学习, 然后对未来的强震作中短期预测. 提出了两个神经网络预测模型: 一个是基于地震演化序列的神经网络预测模型 EE, 并将它用于中国大陆未来一年的最高震级的预测; 另一个是基于地震前兆的神经网络预测模型 EP, 并将它用于华北地区未来(2 年)强震发生时间的预测. 结果表明, 本文提出的这种基于神经网络的预测模型有一定的预测能力, 并且使用方便, 有较好的应用前景.

**主题词** 地震预测 神经网络 建模 地震演化序列

## 引言

预测是许多学科中的一个重要问题, 如果知道了一个系统的过去和现在, 那么如何去预测它的未来? 解决这个问题所采用的方法一般分两步: 第一步是建造一个基于某种理论的模型; 第二步是把系统测量的数据作为模型的输入, 从而得到对未来的预测.

一种建立模型的方法是从系统所满足的基本定律出发构造模型, 并测得初始数据作为输入, 根据初始数据就可以精确地预测上述系统的未来. 不过这种方法并不总是很成功, 因为现实存在的系统多是复杂的非线性系统, 如孕震系统, 至今还没有找到一个描述其行为变化的方程, 去构造一个精确的地震预测模型.

再一种可供选择的方法是直接从观测数据出发构造模型. 其中一种是基于所观测的表示系统演化的时间序列建模的预测方法; 另一种是基于某些与系统演化有关的内在因素和外在触发条件建模的预测方法. 地震预测模型的建立基本采取上述两种直接从观测数据出发的建模方法. 随时间演化的地震事件是不连续点的集合, 本文称之为地震演化序列, 它基本上刻划了地震的时空强演化过程. 在地震预测领域, 与地震发生有关的内在因素与外在触发条件被称为地震前兆.

本文旨在将神经网络用于强震的中短期预测. 我们直接从观测数据出发构造模型, 提出了两个神经网络预测模型: 一个是基于地震演化序列(earthquake evolution sequence)的神

<sup>\*</sup> 国家地震局地球物理研究所论著 97A0061.

1996-03-18 收到初稿, 1996-06-15 收到修改稿, 1996-11-16 决定采用.

神经网络预测模型, 简称为 EE; 另一个是基于地震前兆 (earthquake precursor) 的神经网络预测模型, 简称为 EP. 将它们分别应用于中国大陆及华北地区强震的中短期预测, 并进行了回溯性检验, 结果具有一定的预报效能. 这显示了神经网络应用于地震预测有较好的前景.

1 多层神经网络概述

人工神经网络是由大量类似于生物神经元的简单处理单元——人工神经元分层并行连接而成的复杂网络系统, 通过各单元之间连接强度的调整, 神经网络表现出类似于人脑的学习、归纳能力. 本研究采用了多层前馈神经网络 (图 1), 它由一个输入层、一个或多个隐层及一个输出层组成, 同层之间的神经元互不连接, 某一层中的神经元只与相邻层的神经元间连接. 各层作用如下:

输入层——从外部接收信息, 并将信息传入神经网络;

隐含层——接收输入层的信息, 并进行处理;

输出层——输出神经网络的处理结果.

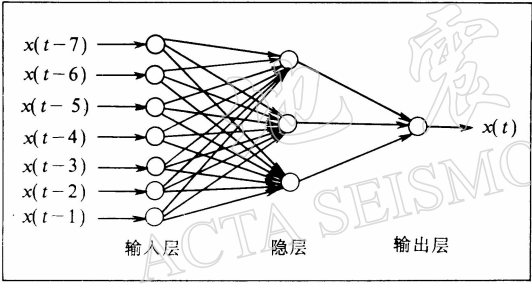


图 1 M7-3-1 神经网络模型

神经网络的应用可分为两个阶段: 第一是训练或学习阶段, 第二是预测或推广阶段. 在学习阶段, 采用 BP 算法 (Rumelhart *et al.*, 1986) 训练网络. 训练过程可以看成是一个最优化问题. 在此过程中, 要找到一组权值  $W_{ij}$  ( $i, j$  为神经元序号), 对于给定的一组训练模式集 (输入-理想输出向量对), 它使得网络输出与理想输出之间的误差最小. 换句话说, 训练网络的目的使网络建立起输入向量空间到输出

向量空间的映射, 这种映射就是所期望的对应于实际问题的输入-输出之间的关系. BP 算法由正向传播和误差反向传播过程组成. 开始训练前要把神经网络的连接权随机地赋初值, 对于给定的训练模式, 正向过程把输入向量输入网络, 从输入层开始逐层计算得到网络的输出; 然后计算网络输出与对应的理想输出之间的误差 (一般用最小平方误差). 这个误差导致了反向传播过程, 即从输出层开始逐层反向计算权重的改变  $\Delta W$  并改变权重. 正向过程和反向过程反复进行, 直到网络输出与理想输出的误差最小, 这样网络就完成了训练或学习过程.

学习过程结束后, 转入预测或推广阶段, 向网络输入新的模式, 观察其能否正确地预测. 如果网络的预测效果好, 就可以把它用于实际工作中, 否则返回训练阶段调整训练模式集, 重新训练网络. 遵循上述步骤, 就可以建立起用于解决实际问题的神经网络模型.

2 基于地震演化序列的神经网络预测模型 (EE)

2.1 理论基础及 EE 模型的构造

近几年发展起来了一系列基于动力学系统理论的短期预测技术 (Farmer, Sidorowich, 1987; Casdagli, 1989), 探索复杂系统时间序列的短期预测. 我们一般用一个非线性映射表

示一个动力学系统. 动力学系统的标准正问题是: 给定一个非线性映射, 描述其迭代的渐进行为, 其中任一迭代分量随时间的演化形成一个复杂的时间序列. 这个时间序列短期预测问题可以看作动力学系统理论中的“反演问题”(Casdagli, 1989). 其提法是: 给定 1 个有限长的时间序列, 它是某个非线性映射的逐次迭代的某一分量, 我们要直接从时间序列出发构造出这个非线性映射, 并用它来预测时间序列的未来发展, 即对于给定的时间序列  $x(n)$ ,  $n=1, 2, \dots, N$ , 求一个光滑函数  $\tilde{f}_N$ , 使

$$\tilde{f}_N[x(n), \dots, x(n+m-1)] = x(n+m) \quad (1)$$

对于  $n=1, 2, 3, \dots, N-m$  成立.  $\tilde{f}_N$  可以看作未知的动力学映射  $f = \tilde{f}_\infty = \lim_{N \rightarrow \infty} \tilde{f}_N$  的逼近, 这就是所谓的动力学系统的“反演”问题. 具体的作法是: 按上式, 把含  $m$  个值的时间延迟序列作为神经网络的输入, 每一个延迟值对应神经网络输入层的一个结点, 神经网络的输出层的结点对应了预测值  $x(n+m)$ . 显然, 这个问题的解是不唯一的, 因为对于  $m$  和所要逼近的非线性映射的形式都是未知的. 也就是说, 神经网络输入层和隐层的结点数是未知的. 为此, 我们把时间序列分成两部分, 一部分用于训练神经网络, 使其表现一个映射; 另一部分用于对经过训练的神经网络进行预测效果的检验, 以此来决定被训练用来预测的神经网络模型的取舍. 最后, 从一系列的神经网络预测模型(代表不同的预测器  $\tilde{f}_n$ ) 中选择一个预测效果较好, 结构最简单的神经网络模型作为实际的预测模型.

虽然上述建模预测理论很容易理解, 但是传统方法常常引出繁琐的数据处理方法, 而且它没有自学习的能力, 易受人为因素的影响. 神经网络是一种自学习系统, 通过对输入-输出向量对(或称输入-输出模式对)的学习, 它可以完成输入空间到输出空间的任意复杂的非线性映射. 神经网络用于描述这种物理过程尚不清楚的从输入空间到输出空间的映射, 并用它进行预测是可行的. 此外神经网络还表现出压制无用噪声和提取输入样本有效信息的能力, 通常在输入样本有噪声或样本不完整时显示出较强的稳定性.

## 2.2 中国大陆强震演化的 EE 模型

### 2.2.1 形成地震演化序列

我们应用上节提出的神经网络预测方法, 把中国大陆作为一个整体来研究其强震震级的演化. 假设中国大陆地震动力学过程可以表述为

$$X(i) = f[X(i-1), \dots, X(i-m)] \quad X = [x_1, x_2, \dots, x_m]$$

我们把 1900~1993 年的年最高震级按地震发生时间先后排成一个时间序列, 见图 2a. 这个序列可以称为中国大陆强震震级演化序列. 下面就直接从这个有限的震级演化序列出发, 用神经网络构造这个未知的非线性映射  $f$ , 以预测中国大陆未来一年的强震的震级. 选取这个序列, 一方面是有许多研究者用线性模型研究过它, 也用非线性模型, 如门限自回归模型(黎令仪, 刘德富, 1985)、GMDH 模型(沈心焯等, 1991)研究过它, 用同一个演化序列便于不同方法之间的比较; 另一方面, 中国大陆地震活动有一些整体规律, 如强震的迁移和成组活动. 对于中国大陆地震活动分区的预测方法的研究将在第 3 节中讨论.

### 2.2.2 EE 模型的构造

#### 2.2.2.1 初始网络的选取

按 2.1 中的讨论, 这个网络的输出层应有一个结点, 其输出值为未来地震震级的预测值. 取这个演化序列的  $m$  个延迟值作为神经网络的输入, 那么神经网络输入层的结点数

为  $m$ ，取网络有一个隐含层. 由于问题的非唯一性，神经网络输入层结点个数  $m$  与隐层结点个数在建模开始时并不能确定. 所以，可根据经验或随机地选取输入层的结点个数和隐层结点个数. 这个过程称为初始网络的选取. 本文选取的初始网络为 M7-3-1 模型，数字分别代表了输入层、隐层和输出层结点的个数(见图 1).

2.2.2.2 神经网络的训练和预测检验样本对的组织

按公式(1)组织学习样本对，取序列前 7 个相连震级值作为神经网络的输入，第 8 个震级值作神经网络的理想输出值，如此向下递推，得到神经网络所需的训练和预测检验样本对. 表示为： $[x(t-7), \cdots, x(t-1)]-[x(t)]$ ，见图 1.

2.2.2.3 预测检验和神经网络结构的确定

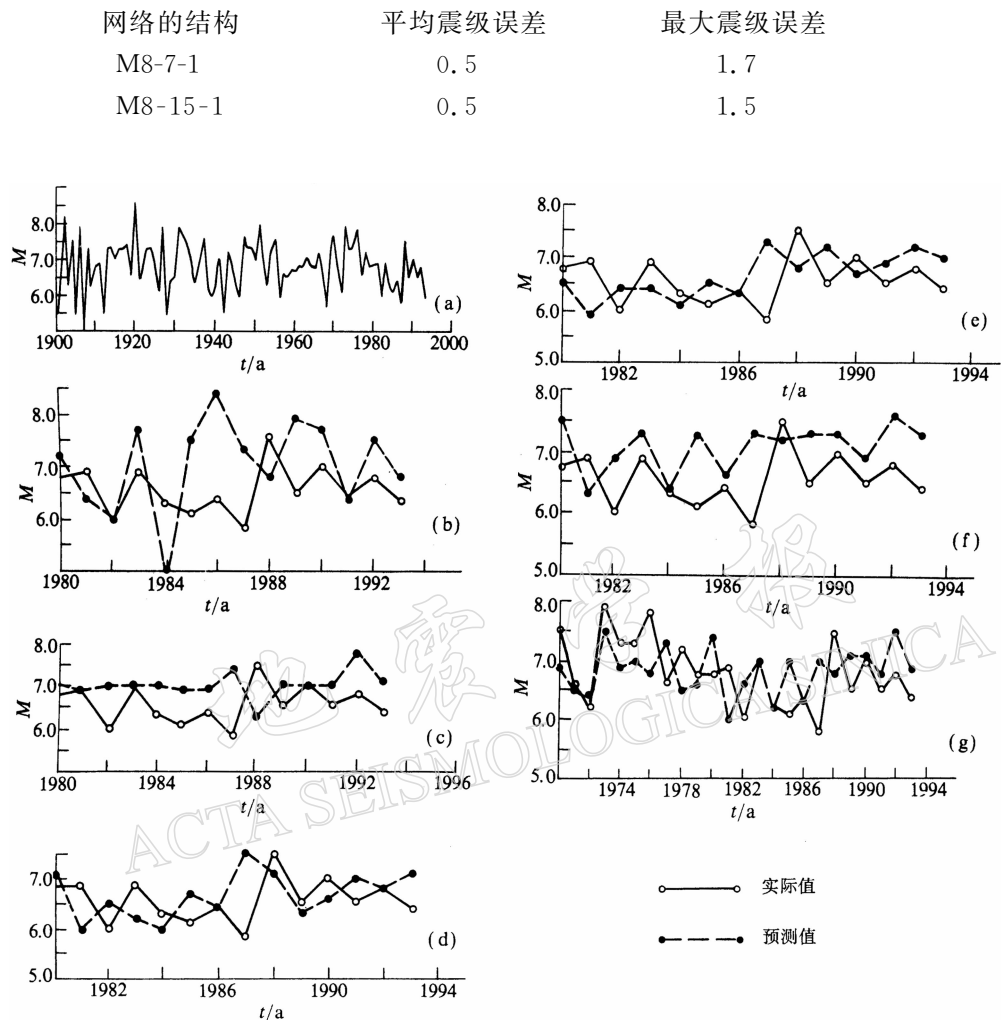
把所得到的震级演化序列分成两部分：前一部分 $[x(0), x(1), \dots, x(79)]$ 共 80 个地震向量用于训练，神经网络通过训练获得序列演化规律的神经网络的内部表示，即一组权系数. 后一部分 $[x(80), x(81), \dots, x(93)]$ 共 14 个震级值用于预测检验，即分析比较预测值和实测值(理想输出)之间的误差大小，预测值与实测值的变化趋势，以确定网络学习的结果是否恰当地反映了这个震级演化序列的内在非线性规律. 增加或减小神经网络输入层结点个数  $m$  与隐层结点个数，比较预测结果，从而选定神经网络输入层的结点数与隐层结点数.

首先进行神经网络输入层结点个数的试验. 从选定的初始网络 M7-3-1 模型开始，把所得到的训练样本对输入神经网络，待网络收敛后，结束学习. 再把检验样本输入网络，网络输出预测值. 然后，计算实际值与预测值的最大震级误差和平均震级误差，作图比较两者的变化趋势. 固定隐层的结点个数，改变序列自变量个数即改变输入层结点个数，重复上述过程. 本试验计算了输入层结点个数取 5, 6, 7, 8, 9, 10, 11, 12, 15 的结果.

当输入层结点数取 5 时，网络学习不能收敛，那么舍弃输入层结点为 5 及小于 5 的网络. 当输入层结点数取 10 时，虽然网络随着结点的增加，收敛速度加快，但网络的预测效果开始变坏，在图上表现为预测值上下剧烈地震荡，预测误差增大，即网络开始变得不稳定. 当输入层结点继续增加，取 11, 12, 15 时，网络变得更不稳定. 经过比较，选取 8 作为本实例网络输入层的结点数. 这里只给出输入层结点数为 6, 8, 15 的神经网络模型的预测结果(见图 2b, c, d).

网络的结构	平均震级误差	最大震级误差
M6-3-1	0.6	1.6
M8-3-1	0.6	1.5
M15-3-1	0.9	2.0

接下来，进行确定隐层结点个数的试验. 固定输入层结点个数为 8，改变隐层的结点数，重复上述预测检验步骤. 当隐层结点数持续增加时，网络预测误差时大时小. 由于隐层结点个数对神经网络的稳定性及计算量的影响较大，在网络预测误差相差不多的情况下，选择一个较小的结点数. 通过试验选隐层的结点为 7. 图 2e, f 给出了隐层结点取 7 和 15 的神经网络模型的预测结果.



2.2.3 EE 模型的预测

现在已经确定了中国大陆地震震级演化的神经网络预测模型为 8-7-1 模型. 只要从这个震级序列连续取 8 年的最高震级值输入网络, 就可以从输出结点得到第 9 年的最高震级预测值, 预测结果瞬间即可得到. 对于神经网络方法的应用, 还存在一个网络结点个数与学习样本对个数关系的问题, 一般学习样本对的数量大而网络结点数相对小为好. 因为在网络结点较多的情况下, 网络在学习时很容易收敛, 但它可能是不稳定的, 其预测效果可能较差. 所以, 在学习样本少的情况下, 单纯地增加网络的结点数以求网络在学习时收敛是没有意义的. 下面改变序列的学习和预测检验段, 使网络重新学习, 观察网络预测效果的变化. 结果如图 2g 所示, 其学习序列的时间段为 1900~1969; 1970~1993 为其预测检

验序列的时间段. 图 2g 的最大震级误差和平均震级误差分别为 1.2 和 0.5, 表明在学习样品数量改变的情况下, 网络的预测性能没有什么变化, 这说明用于训练的样品是足够的, 也说明中国大陆地震震级的演化规律, 在这一时段内没有较大的变化. 如果演化序列的内在规律发生了较大的变化, 需重新训练神经网络以取得更好的预测结果, 这可以根据不同的问题灵活地掌握. 从图 2 的结果看, 预测值强弱变化与实际值相似.

### 3 基于地震前兆的神经网络预测模型(EP)

中国大陆强震不仅具有整体活动的一面, 各个地区也各有自己的特点. 前面提出的 EE 模型一般需要较长的强震演化序列. 对于小的分区很难得到足够长的强震序列, EE 模型的预测可能是不稳定的. 本节讨论基于前兆特征的神经网络预测模型(EP), 研究小的地震活动分区的神经网络地震预测方法. 基于前兆的方法一般可以归为两类: 一类研究单项前兆与某种地震发生的关系; 一类研究多项前兆与某类地震发生之间的关系. 对于后者, 可以用统计模式识别方法建立多项前兆与地震之间的关系; 前者可以看作统计模式识别方法的一个步骤——特征选择. 特征选择是模式识别技术中的一个重要问题, 本文对此不作讨论. 仅在已有的前兆特征研究基础上, 探索神经网络建模预测的方法.

#### 3.1 理论基础及 EP 模型的构造

神经网络的空间映射作用, 使得模式识别和分类成为神经网络的一个重要应用领域. 模式分类器接受一个特征向量, 并产生一个输出类别编码. 特征向量表示特征为空间中的一个点, 类别编码则是这个点所在特征空间某区域的代码. 两个隐层的神经网络可以表示任意复杂的决策区域(decision regions). 实际研究表明, 许多常用的模式识别技术都可以由神经网络实现, 而且在有些应用中, 神经网络表现得更好. 特别对于不完整的数据(如某些特征缺失), 神经网络表现出较好的稳定性(殷勤业等, 1992). 实际的作法是, 把特征向量作为神经网络输入层的输入, 类别的判别由神经网络输出层的输出决定.

#### 3.2 华北地区 6 级以上地震发生时间的 EP 模型

##### 3.2.1 形成样品集合

研究区域选华北地区, 样品的分类采用王碧泉等(1989)提出的方法. 对于 1947~1994 这一时期, 以  $S=2$  年为间隔, 将所研究的时期划分为 24 个时间段, 取  $M_0=6.0$  (实际取 5.9), 对样品进行分类. 若在某时间段内曾发生  $M \geq M_0$  的地震, 则此时间段被划分为 D 类, 反之则被划分为 N 类. 这样就把对发震时间的预测转化为对未来时段的类别的预测. 共得 D 类样品  $N_1=8$  个, N 类样品  $N_2=16$  个. 本文采用地震活动性前兆特征, 按王碧泉等(1988)的研究, 地震活动性特征分九组, 共 94 个特征.

第 1 组为特征 1~11, 它们与地震频度有关; 第 2 组为特征 12~17, 它们与地震频度平静, 活化有关; 第 3 组为特征 18~28, 它们与  $b$  值有关; 第 4 组为特征 29~39, 它们与地震能量有关; 第 5 组为特征 40~50, 它们与地震蠕变有关; 第 6 组为特征 51~61, 它们与震源面积有关; 第 7 组为特征 62~72, 它们与平均能量有关; 第 8 组为特征 73~83, 它们与平均蠕变有关; 第 9 组为特征 84~94, 它们与平均震源面积有关.

杨锦英和王碧泉(1993)在华北地区的研究中进行了特征选择, 认为编号为 8, 76, 87, 12, 31, 47, 53, 69, 19 等特征的有效性较高, 本文就采用这 9 项前兆特征形成输入向量, 共 24 个特征向量-类别对构成了样品集合(见表 1). 所选的 9 个地震活动性特征为

特征内容	特征号
频度短期相对于超长期偏离	8
超长期平均蠕变	76
超长期平均震源面积	87
频度平静	12
长期能量	31
蠕变短期相对超长期偏离	47
长期震源面积	53
平均能量短期相对超长期偏离	69
中期 $b$ 值	19

3.2.2 EP 模型的构造

3.2.2.1 初始网络的选取

选取 1 个 3 层网络作初始网络，网络输入层的结点数应等于输入向量的维数，即所选的前兆特征数，所以网络的输入层结点数为 9。样品被分成了两类，则输出层结点数取 1 个；隐层结点数选为 2。

3.2.2.2 神经网络的训练和预测检验样品的组织

样品输入向量的 9 个分量作为网络的输入，一个分量对应一个网络输入层结点，网络的理想输出为 0.9 和 0.1，分别对应地震样品是 D 类和 N 类的情况。

3.2.2.3 预测检验和网络隐层结点个数的确定

表 1 样品表(8 个 D 类样品，16 个 N 类样品)

样品号	样品时间段	样品原类别	样品号	样品时间段	样品原类别
1	1947-01~1948-12	D	13	1971-01~1972-12	N
2	1949-01~1950-12	N	14	1973-01~1974-12	N
3	1951-01~1952-12	N	15	1975-01~1976-12	D
4	1953-01~1954-12	N	16	1977-01~1978-12	D
5	1955-01~1956-12	N	17	1979-01~1980-12	D
6	1957-01~1958-12	N	18	1981-01~1982-12	N
7	1959-01~1960-12	N	19	1983-01~1984-12	D
8	1961-01~1962-12	N	20	1985-01~1986-12	N
9	1963-01~1964-12	N	21	1987-01~1988-12	N
10	1965-01~1966-12	D	22	1989-01~1989-12	N
11	1967-01~1968-12	D	23	1991-01~1992-12	N
12	1969-01~1970-12	D	24	1993-01~1994-12	N

把此样品集的 24 个样品分成两部分，前一部分用于网络学习判别规则，后一部分用于预测检验。表 1 为样品表。第 1~10 号样品用于学习，11~24 号样品用于检验。网络的预测效果用  $R$  值评价。把学习样品输入 M9-2-1 初始网络，待网络收敛后，停止学习，转入预测检验阶段，把待检验的样品的输入向量输入网络，网络给出其预测类别(见表 2 第 4 列)，最后计算  $R$  值。改变初始网络隐层的结点数，重复上述预测检验过程。当隐层结点数取 1 时，网络不能收敛于给定的误差，所以网络隐层结点数取 1 不合适；当网络隐层结点取 3，4

时,网络的预测结果与初始网络相同;当网络隐层结点取 5 以上时,网络学习时收敛得极快,但网络变得不稳定,不能对 D 类样品作出正确的识别. 因此,我们选择隐层结点个数最少的网络,也就是选 M9-2-1 模型作为前兆预测模型. 检验结果见表 2,其  $R$  值为 0.46,是具有一定预报效能的.

表 2 预测结果检验

样品号	样品原类别	网络输出值	预测的样品类别	预测正确否
11	D	1	D	T
12	D	0.94	D	T
13	N	0.73	D	F
14	N	1	D	F
15	D	0.99	D	T
16	D	0.94	D	T
17	D	0.96	D	T
18	N	0.97	D	F
19	D	0.03	N	F
20	N	0.01	N	T
21	N	0.01	N	T
22	N	0.01	N	T
23	N	0.01	N	T
24	N	0.01	N	T

注:第 1~10 样品学习(略),第 11~24 样品检验,  $R=1-\frac{3}{8}-\frac{1}{6}=0.46$ , T 为正确, F 为错误.

3.2.3 EP 模型的预测

现在已经确定了华北地区强震发生时间前兆神经网络预测模型. 如果要对未来两年内能否发生 6.0 级以上地震进行预测,只要把这一样品输入向量输入预测网络. 如果网络输出为 0.9,则表示未来两年内将发生  $M_s6.0$  以上地震;如果网络输出为 0.1,则表示未来两年内不会发生  $M_s6.0$  以上地震. 如果样品集中的样品数量有了变化,也可以重新划分样品集的学习和预测检验样品,让网络重新学习,用学习后的网络进行预测. 我们改变学习样品集,其结果与上面的基本相同(篇幅所限,结果不一一列出),这表明所选用的神经网络模型比较稳定.

最后,对前兆预测模型的神经网络建模和预测方法中的问题作一些讨论. 对于两类问题,网络输出为 0.9 则判别样品为 D 类;输出为 0.1 则判别样品为 N 类. 如果输出值大于 0.9 或小于 0.1,那么,就把样品肯定地判别为 D 类或 N 类;如果输出在 0.1 到 0.9 之间,本文仍根据其与 0.1 或 0.9 差距的大小把样品判为 D 类或 N 类. 虽然这样,仍存在着错判的危险. 如果这样的错判不能消除,也许意味着把样品按某一震级截然分为两类是不够的;再一个问题就是在计算地震活动性前兆时所采用的按窗口滑动的计算方法,如果 D 类样品和 N 类样品在一个时段内交错出现(如样品 12~18),窗口滑动的步长较小, D 类和 N 类样品的前兆特征值没有较大的改变,所以这样的样品时段的类别很难正确区分,用模式识别方法建模时也存在上述问题. 有的研究者对这个问题采取的作法是,把一次强震后若干年的时段剔除,这样作表面上能提高预报的效果,但实际上并没有解决这个问题. 这个问题也许有待于更进一步的研究来解决.



## 4 结论

中国大陆地震演化序列的神经网络预测模型的预测结果如下：震级预测的平均误差为 0.5；震级误差序列的标准偏差为 0.3. 与黎令仪和刘德富(1985)及沈心焯等(1991)的结果相同. 华北地震前兆神经网络预测模型的预测结果为： $R$  值达 0.46，通过了具有 97.5% 置信水平的  $R$  值检验. 作者认为，基于神经网络建模的地震预测方法是可行的，有一定的应用前景.

与传统建模预测方法比较，神经网络建模预测方法具有以下 3 个优点：第 1，神经网络是一个自适应的系统，它从已存在的数据中学习而避免了人为的干扰；第 2，神经网络具有很强的处理非线性数据的能力. 理论上，一个 3 层前馈神经网络可以表达一个任意复杂的非线性映射；第 3，神经网络可以简化数据的处理和预分析工作.

## 参 考 文 献

- 黎令仪, 刘德富, 1985. 震级序列的门限建模. 地球物理学报, **28**(3): 303~310
- 沈心焯, 朱照宣, 刘式达, 等, 1991. 震级序列的 GMDH 建模. 国际地震动态, (3): 31~34
- 杨锦英, 王碧泉, 1993. 模式识别方法应用于华北地区的检验. 地震预报研究论文集. 北京: 地震出版社, 89~95
- 王碧泉, 杨锦英, 王玉秀, 1988. 自相似地震活动特征的提取. 地震研究, **11**(3): 241~249
- 王碧泉, 王春珍, 杨锦英, 1989. 强震发生时间的中期预测. 八十年代地球物理学进展. 北京: 学术书刊出版社, 100~113
- 殷勤业, 杨宗凯, 谈正, 等, 1992. 模式识别与神经网络. 北京: 机械工业出版社, 323
- Casdagli M, 1989. Nonlinear prediction of chaotic time series. *Physica*, **D**(35): 335~356
- Farmer J D, Sidorowich J J, 1987. Predicting chaotic time series. *Phys Rev Lett*, **59**: 845~848
- Rumelhart D E, Hinton G E, Williams R J, 1986. Learning representations by back propagating errors. *Nature*, **323**(9): 533~536