

# 地质灾害的非线性数据处理与建模技术

许 强, 黄润秋

(地质灾害防治与地质环境保护国家专业实验室 四川 成都 610059)

**摘 要:** 本文简略地介绍了几种地质灾害数据处理与建模的非线性方法, 主要包括 GMDH 自组织建模技术、神经网络方法。GMDH 是一种高阶非线性回归建模方法, 它是以简单的二元二次回归方程为基础, 通过“代复一代”的“生产”过程, 客观、自动地求得实际资料的非线性模型。而神经网络则是用工程技术手段模拟生物神经网络的结构特征和功能特征的一类人工系统。与常规统计方法相比, 神经网络最突出的优点为它是通过对网络的学习和训练, 来掌握变量之间的非线性关系。因此, 其处理复杂问题的能力更强大。实例检验效果表明, 这些非线性数据处理与建模技术考虑了地质灾害问题的非线性特性, 其比基于常规统计理论的数据处理方法的精度要高得多。

**关键词:** 数据处理; 数据处理的组合方法; 神经网络; 地质灾害; 非线性科学

中图分类号: P496

文献标识码: A

一个学科的发展水平从某种角度讲可用该学科研究的定量化程度予以评价。地质灾害问题一般都是具有高度复杂性的定性或半定量问题, 几十年来, 有关地质灾害研究的工作者一直致力于地质灾害问题的定量化研究, 并且取得了显著的进展。特别是近年来随着计算机技术的发展, 各种数理统计方法(如回归分析、聚类分析、模式识别、概率统计、灰色系统理论、模糊数学等)都已相继被引入地质灾害问题的研究中, 用它们处理和解决了许多生产实际问题, 且取得了显著的社会经济效益。然而, 随着人们对地质灾害问题认识的逐渐加深和定量化研究程度的不断深入, 越来越感觉到传统的定量化评价方法在地质学这门复杂的学科面前已逐渐暴露出缺点和不足, 研究者们已发现许多地质灾害问题的定量化评价结果与实际相差甚大, 有时甚至会产生错误的结论。产生上述情况的主要原因在于地质灾害问题本身的非线性, 此非线性主要来自于两个方面, 即地质灾害体变形、演化规律的非线性和内外因素相互作用的非线性。这种非线性使得我们不能再仅仅使用一些简单的线性代数方程来真实地描述地质灾害体的发展演化规律和各变量间的相互关系, 必须采用非线性方程和发展新的数据处理方法。作为代表, 本文拟简要地介绍两种非线性科学中用于处理数据间非线性关系的数据处理方法, 即数据处理的组合方法和神经网络方法。

## 1 地质灾害数据处理的 GMDH 自组织建模技术

传统的统计建模都有这样一个特点: 建模之前必须事先猜测或假定模式里的变量之间的函数关系  $f$ 。统计建模工作的实质是根据已知数据或资料去确定该函数  $f$  里的参数或系数。显然, 统计建模方法有很大的缺陷, 其主要表现在以下几方面: (1) 统计建模一般仅当因变量和自变量之间呈线性关系(或通过适当变换可化为线性关系, 如对数关系、指数关系等)或一些简单的函数关系(如低次多项式)时才适用。但对于复杂系统, 因变量与自变量之间在很多情况下并不符合这种简单的关系, 而是往往存在着高

收稿日期: 1999-05-14 改回日期: 1999-08-10

基金项目: 本研究获国家自然科学基金(项目编号: 49702038)、国家杰出青年科学基金(项目编号: 49525204)联合资助

作者简介: 许强(1968-), 男(汉族), 四川南江人, 博士, 副教授, 地质工程专业, 出版专著 3 部(与人合著), 发表论文近

度的非线性关系,此时已不能用传统的统计建模方法来建模,或者建模者事先根本无法再通过经验判断或观察散点图等方式来发现或猜想出变量之间所应满足或近似满足的函数关系。(2)虽然在建模之前已经可以确定函数关系,但赖以建模的数据或资料太少而不具有统计意义。同时,因为建模工作是建立在对函数  $f$  进行根据不足的猜测的基础上,利用少数几个已知样本所建立起的模型可能并不具有代表性,根本无法进行外推(预测)。

为了克服传统建模方法的弊端, A. G. Ivakhnenko 在七十年代初提出了一个新的建模思想:不应该带有任何人为偏见,而应该只根据数据来客观地建立模型。1974 年, Ivakhnenko 为建立多(几)个变量之间的统计模式,提出了数据处理的组方法(Group Method of Data Handling, 简称 GMDH)。这种适用于多变量之间的统计建模的数学基础本质上仍是回归方法,即找出模式的因变量  $Y$  与自变量  $X_i(i=1, 2 \cdots)$  之间的函数关系。但传统的回归建模方法所不同的是:在建模过程中不再需要人为假设函数关系,而是根据所给样本的具体情况,让计算机自动寻找出数据间的函数关系(即具“自组织”特性),并且所得结果往往是高阶非线性函数,因此能最大限度地拟合出地质灾害问题的非线性关系。

1.1 GMDH 方法概述

设想要构造的复杂系统高阶回归模型为

$$y = f(x_1, x_2, \cdots, x_m) \tag{1}$$

式中  $y$  为输出变量,  $x_1, x_2, \cdots, x_m$  为输入变量。

GMDH 方法仅利用输入、输出变量的所有观测数据,而不事先设置任何参数和模型的形式。它的基本思想是以生物有机体演化的方法构造数学模型。

首先从系统的输入变量  $x_1, x_2, \cdots, x_m$  出发,对输入变量中的每一对  $x_i$  和  $x_j$  及输出变量  $y$  计算如下回归方程

$$y = A + Bx_i + Cx_j + Dx_i^2 + Ex_j^2 + Fx_ix_j \tag{2}$$

这里将产生出  $m(m-1)/2$  个较高阶变量,用它们替代最初的  $m$  个变  $x_1, x_2, \cdots, x_m$ ,对输出变量  $y$  进行估计。

在从一组输入、输出观测中找到这些回归方程后,用一个判据来对每一个回归方程进行评估,选出其中最佳者保留下来。于是可得到一组(假定为  $m_1$  个)对  $y$  进行最佳估计的二次回归方程(每个估计都只依赖于两个自变量)。然后,利用每一个刚刚得到的回归方程生成第二代输入变量,并以此代替原始的  $x_1, x_2, \cdots, x_m$ 。

按上述作法,计算  $y$  对这些新的输入变量的二次回归方程,这样将得到更新的一组  $m_1(m_1-1)/2$  个用新变量估计  $y$  的回归方程,此时所得到的是一组四个变量的四次多项式。再选择这些新变量中的最优者,用所挑选出的方程生成第三代输入变量来代替第二代,并用第三代输入变量逐对组合构造二次回归方程。继续这一过程,直到回归方程开始比前一代的回归方程的估计能力有所下降为止。在逐代构造回归方程的过程停止后,挑选出最后一代的二次多项式中最好的一个,最终所得到的对  $y$  的估计是两个变量的二次多项式,这两个变量各是上一代另外两个变量的二次多项式,而另外两个变量又各是另外更上一代两个变量的二次多项式,……,第二代变量则是原始变量的二次多项式。换句话说,如果做出所有的代数替换,将得到下面形式的非常复杂的多项式

$$y = a + \sum_{i=1}^m b_i x_i + \sum_{i=1}^m \sum_{j=1}^m c_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m d_{ijk} x_i x_j x_k + \cdots \tag{3}$$

上式被称为 Ivakhnenko 多项式(也称 Kolmogorov-Gabor 多项式)。

从总体上讲, GMDH 实际上是一种具“自组织”特性的高阶非线性回归建模方法,通过该方法我们可以只依靠少量数据便能构造出能较客观地描述象地质灾害系统演化这样复杂过程的回归方程。与传统的回归分析方法相比, GMDH 具有以下的特点: 1. 它通过简单的二元二次回归原理构造出下“一代”较为复杂的次级回归方程,并通过“适者生存”原理来淘汰掉次级回归方程中那些不理想的项,只保留

“下一代”构造矩阵中  $r < L$  的列, 并以此构成次级矩阵, 这其中又利用了“优选法”原理; 2. 通过“代复一代”的“生产”过程, 使得我们可较轻易地获得能较客观地描述复杂非线性系统(过程)的非线性模式, 并避免求解在标准高阶非线性回归中所要求解的“病态”高阶回归方程。例如若建模时进行到 8“代”叠代, 则模式的非线性幂次已高达 256 阶, 而要直接计算如此高阶的标准非线性回归方程, 则是一件相当困难的、甚至是不可能的事; 3. 最终的回归模式的幂次是根据资料和有关的判据(如方均根判据)客观、自动地给出的, 而不象标准的回归方法那样是事先人为地假设回归方程的幂次或函数形式。

因此, GMDH 自组织建模方法无论是从建模思想, 还是从解决问题的能力来看, 都比传统的回归分析方法要优越得多。为了检验 GMDH 建模方法的效果, 下面以卧龙寺新滑坡的变形时间序列为例建立 GMDH 模型, 并利用所建模型进行变形预测。

1.3 实例分析

卧龙寺新滑坡是一个塬边黄土滑坡, 1971 年发现裂缝后于该年 3 月 11 时起开始对其变形进行观测, 5 月 5 日凌晨 3 时 15 分产生剧滑, 其 5 # 裂缝的变形时序(表 1)。

表 1 卧龙寺新滑坡位移监测资料  
Table 1 Displacement monitoring data of the latest Wolongsi landslide

年、月	1971 年 3 月												
日	15	16	17	18	19	20	21	22	23	24	25	26	27
位移(mm)	1.0	1.5	1.7	2.5	3.2	4.0	4.4	5.1	5.9	6.3	7.0	7.3	7.8
年、月	1971 年 3 月				1971 年 4 月								
日	28	29	30	31	1	2	3	4	5	6	7	8	9
位移(mm)	8.2	8.4	8.7	9.0	9.2	9.4	10.0	10.1	10.3	10.4	10.5	10.8	11.1
年、月	1971 年 4 月												
日	10	11	12	13	14	15	16	17	18	19	20	21	22
位移(mm)	12.0	13.0	13.5	14.0	15.0	16.1	16.4	17.2	17.6	18.2	19.0	19.2	20.0
年、月	1971 年 4 月								1971 年 5 月				
日	23	24	25	26	27	28	29	30	1	2	3	4	5
位移(mm)	23.0	24.0	25.2	26.0	27.0	28.2	30.0	31.0	31.0	32.0	33.0	42.0	47.0

利用表 1 的数据, 建立 GMDH 模型, 取独立变量数  $M=6$ , 预测步数为 3。利用后 42 个数据作为训练集, 经过两代生成后代迭代满足要求, 得到最后一代最优 Ivaknenko 多项式为:

$$y = a + bu + cv + du^2 + ev^2 + fuv \tag{4}$$

式中  $a=1.122$ ;  $b=-0.3984$ ;  $c=1.053$ ;  $d=-0.9516$ ;  $e=1.240$ ;  $f=2.220$ 。GMDH 模型的预测结果(图 1)。图 1 表明, GMDH 模型能充分利用观测数据, 通过“一代传一代”的自组织“生产”过程, 较客观地掌握了系统变量间复杂的关系(不带任何人为因素), 具有较高的预测精度。

为了更能说明问题, 利用传统的回归分析方法对表 1 的数据进行回归拟合。结果表明, 用指数曲线拟合原始数据的拟合度最高, 表达式为  $y=2.73556e^{0.0546367x}$  (图 2)。将图 1 与图 2 对比可知, GMDH 预测方法能更准确地掌握斜坡的演化趋势, 故其预测结果与原始序列间的拟合程度要比传统的回归分析方法高得多。

2 基于神经网络理论的地质灾害数据处理与定量化分析方法

从本文第 1 节的分析可以看出, GMDH 法从本质上仍属于回归分析, 都是用一确定的模型(函数)来表达变量间的关系。事实上, 地质体的结构是相当复杂的, 它涉及的工程地质条件及岩体特性参数大多是不完全定量的, 甚至是随机的、模糊的, 并且这些变量之间还存在着复杂的交叉效应和动态效应, 难以用确定性的数学模型加以描述。换句话说, 影响岩土体系统特性的各要素之间可能存在着非常复杂的非线性关系, 这种关系有时甚至复杂到不能用一个(一组)简单的代数方程来描述。针对这种高度非线性和不确定性问题, 用神经

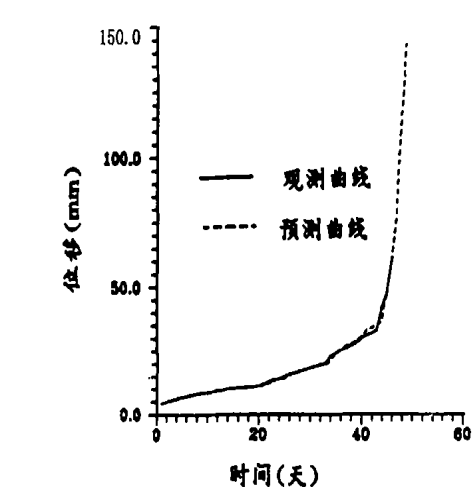


图 1 卧龙寺新滑坡 GMDH 拟合和预测结果  
Fig. 1 fitting and predicting results of the latest Wolongsi landslide using GMDH

表 2 某水电站坝区高边坡稳定性分级

Table 2 Stability grades of high slopes for a hydroelectric power station

原型号	编号	输入变量					期望输出	稳定性等级
		$\varphi$	$\alpha$	$h$	$H$	$\beta$		
III	1	0	0	0	0	1	100	低
I-1	2	0	0	0	1	1	100	
I-2	3	0	0	0	1	1	100	
II-1	4	1	0	0	1	0	010	中
I-1	5	0	1	0	1	1	010	
III	6	1	0	1	0	1	010	
II-1	7	1	0	1	1	0	001	高
II-2	8	1	0	1	0	0	001	
I-1	9	1	1	1	1	1	001	

网络理论来分析和处理往往更为有效。

神经网络(Artificial Neural Network)是与传统统计分析和 GMDH 方法截然不同的数据处理方法。它是模拟人脑的思维方式,把大量的神经元连接成一个复杂的网络,利用已知样本通过对该网络进行训练(类似于人类的学习过程),让网络存储变量间的非线性关系(类似于人脑的记忆过程),然后用所存储的网络信息对未知样本进行分类或预测(类似于人脑的联想过程)。因此,神经网络是一种比前两种方法都更优越的智能化的数据处理方法。它处理数据的功能,特别是处理具非线性关系的数据的能力,是目前其它方法远远不能比拟的。

作为实例,现以一个斜坡空间稳定性分析问题来说明神经网络的学习和预测过程及数据处理能力。根据黄河某水电站坝区高边坡岩体稳定性的蒙特-卡洛破坏概率模拟结果,将该区斜坡危险程度分为三级(表 2)。其各输入变量的取值原则见表 3。

根据实际需要,设置了三层神经网络(图 3)。输入层为 5 个节点,分别对应滑面强度参数( $\varphi$ )、滑面倾角(或组合交线角)( $\alpha$ )、坡高( $H$ )、坡角( $\beta$ )及空隙水压力水头( $h$ )等五个变量。输出层取为 3 个节点,分别对应

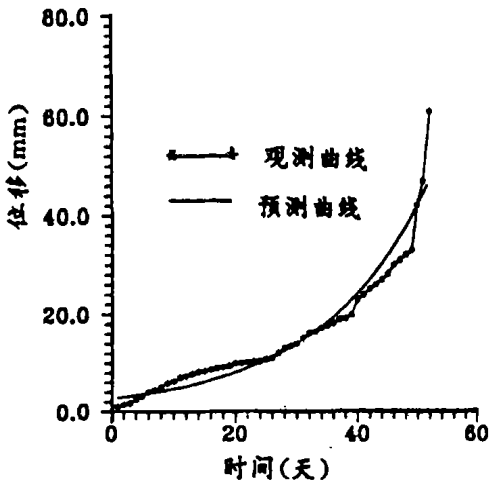


图 2 卧龙寺新滑坡传统回归分析拟合结果  
Fig. 2 Fitting results of the latest Wolongsi landslide using regression analysis

表 3 输入变量的取值原则

Table 3 Definition of variables

变量	门坎值( $c_r$ )	取值原则
$\varphi$	21.8	$\geq \varphi_{c_r}$ 取 0 $< \varphi_{c_r}$ 取 1
$\alpha$	18.0	$\geq \alpha_{c_r}$ 取 1 $< \alpha_{c_r}$ 取 0
$h$	完全充水 无水	$\geq h_{c_r}$ 取 1 $< h_{c_r}$ 取 0
$H$	200.0	$\geq H_{c_r}$ 取 1 $< H_{c_r}$ 取 0
$\beta$	45.0	$\geq \beta_{c_r}$ 取 1 $< \beta_{c_r}$ 取 0

于低等、中等、高等稳定性。中间层取 5 个节点, 用于提高网络的联想(或映射)能力。输出层的取值原则是稳定性最靠近的等级取 1, 其余取 0 值。

运用上述网络模型, 用表 2 给定的数据对网络进行训练, 迭代 5000 次后, 误差满足精度要求, 说明网络的训练(学习)过程完毕。为了检验网络的学习效果, 用学习样本作为假定的未知预测样本对学习样本进行回判输出。结果表明, 训练 5000 次后网络的实际输出与期望输出已非常接近(误差已经非常小), 表明网络经过学习已较好地掌握了输入与输出间的映射关系, 因此可用此已训练好的网络对该区其它稳定性未知的斜坡地段进行预测, 确定出其稳定性级别, 进而可作出该区的斜坡稳定性分区图。限于篇幅, 具体过程和结果请参看文献 [1]。

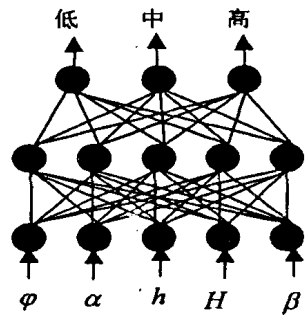


图 3 神经网络模型  
Fig.3 The model of Artificial Neural Network

3 结 论

由于地质体是一个开放体系, 致使地质灾害问题往往具有高度的非线性特性。因此, 在分析和处理地质灾害问题的数据时, 最好采用非线性分析方法。从本文所介绍的 GMDH 法和神经网络理论的基本原理和应用效果可以看出, 与常规统计方法相比, 数据处理的组合方法(GMDH)和神经网络法在分析和处理与地质灾害问题有关的数据时, 具有方法先进, 精度高, 可靠性好, 更符合实际的优点, 值得推广应用。

参考文献:

[1] 黄润秋, 许强著. 工程地质广义系统科学分析原理及应用 [M]. 北京: 地质出版社, 1997  
[2] 秦四清等著. 非线性工程地质学导引 [M]. 成都: 西南交通大学出版社, 1993  
[3] 林振山著. 长期预报的相空间理论和模式 [M]. 北京: 高等教育出版社, 1989  
[4] 靳蕃、范俊波等. 神经网络与神经计算机: 原理·应用 [M]. 成都: 西南交通大学出版社, 1991

NONLINEAR DATA HANDLING AND MODEL ESTABLISHING  
OF GEOLOGICAL HAZARDS

XU Qiang, HUANG Runqiu

(National Laboratory of Geological Hazard Prevention, Chengdu 610059)

**Abstract:** Two kinds of nonlinear methods of data handling and model establishing are introduced in this paper. These are group method of data handling (abbr. GMDH) and artificial neural network (abbr. ANN). GMDH is a higher order and nonlinear regression method. Based on the simple regression function of two variables and two orders, GMDH can get the nonlinear model by means of continual producing process automatically and objectively. Artificial neural network is one of artificial system, which simulates structural and functional characteristics of biological neural network resorted to engineering technological means. Compared with statistical and other mathematical methods, its remarkable advantage lies in that it can grasp the complicated nonlinear relationship between (Turn to page 82)

# STUDY ON THE ENGINEERING GEOLOGICAL PROBLEMS OF TANZITIAN RESERVOIR IN FONGDU COUNTY IN CHONGQING

DENG Rong-gui, FU Xiao-min

(*National Laboratory of Geological Hazard Prevention, Chengdu 610059*)

**Abstract:** The conclusion reduced from study on the engineering geological problems of Tanzitian reservoir dam—site and banks based on the data from field investigation and test, are basis for Tanzitian reservoir design.

**Key words:** Rock-mass stability; reservoir slope; reservoir analysis

---

(Following page 127)input and output variables by learning from given samples, and it has function of associative memory. So ANN can handle data which have very complicated relationship. The results through actual examples testing show that the two methods of GM DH and ANN are superior to routine statistical methods for complex data handling.

**Key Words:** data handling; group method of data handling; artificial neural network; geological hazards; nonlinear science