

38-42

地质数据的 BP 网络分析方法

谢贤平 柴建设 童光照

(北京科技大学 · 100083)

P 628.4

A

数学地质中母函数选取和系数求解是函数逼近的两个关键问题,目前在地质数据的多元统计分析中广泛应用的回归分析或其他映射方法均有局限性。本文采用 ANN 中的 BP 网络模型较好地解决了这两个问题,并对理论和实际的数据进行了比较与验证。结果表明,BP 网络有助于提高地质数据处理的可靠性与预测的精度。

关键词 ANN BP 网络 地质数据处理 回归分析



数学地质

1 问题的提出

数学地质是运用数学方法研究各种地质现象的数量关系和空间形式的科学,其内容相当广泛,地质数据的多元统计分析和地质统计学是其主要内容。多元统计分析的一个重要内容就是函数逼近。在函数逼近中,母函数的选择和系数的求解比较关键。简单函数的逼近,用回归分析等数学方法还能解决,而对复杂函数关系则面临很大的困难^[1]。美国加州 PDO (Parallel Distributed Processing) 小组提出的 Back Propagation 人工神经网络 (Artificial Neural Network, ANN) 算法,简称 BP 网络,可从另一个角度较好地解决这个问题。该算法把 1 组样本的输入输出问题变为一个非线性问题,并采用最优化方法中的沿梯度下降法求解。BP 网络对简单的非线性函数进行复合,从而构成更复杂的函数,这对数学中函数表示法也有重要的启示。本文应用 BP 网络模型对地质数据进行了处理,并与回归分析的结果对比,在此基础上对 BP 网络模型在地质勘探和矿山生产中的应用做一些讨论。

2 BP 网络模型与算法^[2]

ANN 是模拟人脑的思维方式和组织形式而建立起来的数学模型,具有并行处理、联想记忆、抗干扰能力强等特点。目前,已提出了数十种 ANN 模型,比较典型的有 BP 网络、Hopfield 网络等。

BP 网络属于多层状型 ANN,由若干层神经元组成,不仅有输入层、输出层,而且有 1 层或多层的隐层。不同层的神经元的作用也不相同。输入层接受外界的信息;输出层给出输入信息的判别或决策;中间层用来表示或存贮知识。图 1 是一个包含 3 层神经元的 BP 网络的结构图。

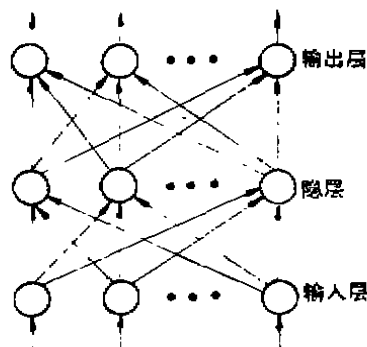


图 1 3 层 BP 网络结构图

本文 1994 年 7 月收到,于纯烈编辑。

如图 1 所示, BP 网络中信号的传递是单向的, 并且同层神经元之间无连接, 相邻层间有连接。连接强度用权值表示, 每层神经元的状态只影响下一层的神经元的状态, 每个神经元的状态对应着一个作用函数 (f) 和阈值 (Q), 输入层神经元阈值为零, 且 $f(x) = x$; 隐层和输出层的作用函数为非线性的 Sigmoid 型函数, 一般采用 $f(x) = 1/(1 + e^{-x})$ 。

BP 网络的学习算法的基本思想是根据网络输出层的实际误差, 从输出层开始, 反过来调整网络的权值和阈值, 最终使得输出的均方误差最小, 此过程称训练或学习, 图 2 是 BP 网络学习算法实现的程序框图, 其中 ERR 为单个干扰最大误差, EMS 为最大场方误差。

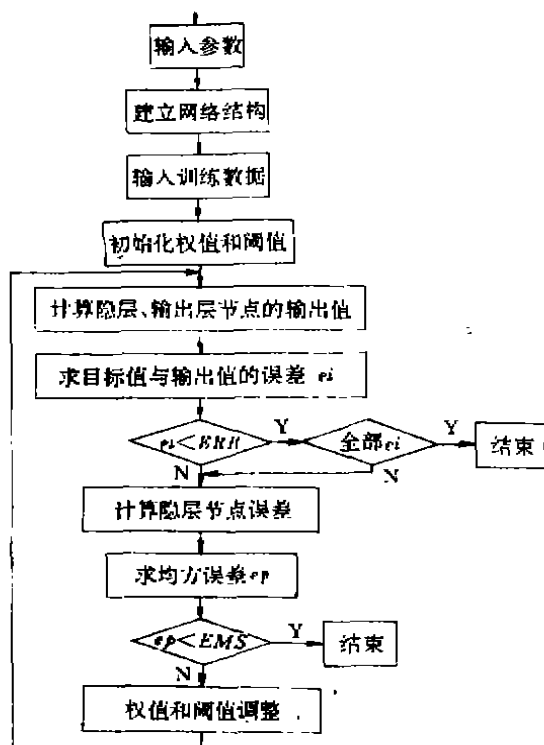


图 2 程序框图

依据图 2, 在微机上用 TurboC++2.0 编写出程序, 该程序允许用户规定输入和输出节点的数目及隐层的层数和每层的节点数, 可以选择相应的网络结构, 允许修改学习速

度、均方误差和干扰误差等参数, 这些操作均在窗口和菜单提示下进行, 用户界面友好^[3]。

3 BP 网络用于回归分析

回归分析是研究变量间相关关系的一种统计方法, 通过建立一个变量和另一个或多个其他变量之间的数学表达式来表示它们之间的定量关系。回归分析在地质工作中有着广泛的用途, 例如, 在多金属矿床中, 研究各元素间的关系, 以便根据其他一些元素的含量预测主要成矿元素的含量, 或者由主要元素的基本分析估计次要元素的含量。回归分析用来确定变量间的定量线性关系时, 一般可以获得满意的结果; 但当变量之间存在非线性关系时, 或者说这种关系还未知时, 用回归分析作预测或分析, 结果并不满意; 而 BP 网络则可很好地解决这类问题。

现有 200 个观测数据, 因变量为 X , 自变量有 A, B, C 和 D 。与这 4 个自变量关系见图 3。取其中的 100 对数据, 采用多元线性回归程度得出如下的回归方程:

$$X = 0.538 + 0.633A - 0.447B + 0.433C - 0.390D$$

总的均方误差为 0.253, 改变自变量 A, B, C 和 D 的值, 用获得的回归方程去计算另外 100 个数据, 观测值与预测值的分布见图 4, 预测的均方误差为 0.198。

采用 BP 网络来处理这批数据, 选用如下网络结构: 输入层用 4 个节点, 表示 4 个自变量, 输出层有 1 个节点即变量 X , 两层有 10 个节点和 4 个节点的中间层。先输入 100 个样本, 对以上 BP 网络进行学习, 均方误差为 0.022。为了验证, 可以再输入 100 个样本, 此时的均方误差为 0.026。观测值与 BP 网络输出值分布见图 5。

从均方误差与图 4 和图 5 的数据分布可看出, BP 网络的计算精度高于回归分析。以上的这批数据由公式 $X = A(1 - B^2) + Ce^{-D}$ 产生, 4 个自变量均为 $[0, 1]$ 间随机数。

4 应用实例

4.1 泥页岩矿物成分与理化指标的关系

在钻探工程中,如要事先知道泥页岩矿物成分与理化指标的关系,则采用人工分析方法不仅成本高,而且工作量大。因此采用了一系列的数理统计方法。例如,逐步回归、岭回归等,取得了一些有意义的结果,但未从根本上加以解决。主要是泥页岩矿物成分与理化指标虽有相关性,但并非全成线性关系。为

此决定选用 BP 网络,用 BP 网络解决该问题的具体步骤是:(1)采用 3 层 BP 网络,分别是输入层、隐层和输出层。输入层:5 个节点,代表泥页岩中的理化指标,包括水分、密度、CEC、CST 和达到平衡时的膨胀率;隐层:节点数据为 b ;输出层:2 个节点,代表泥页岩中的伊利石与蒙皂石的含量。(2)输入 51 个样本,从中随机抽取 4 个样本用于验证,另 47 个样本作为该网络学习用。

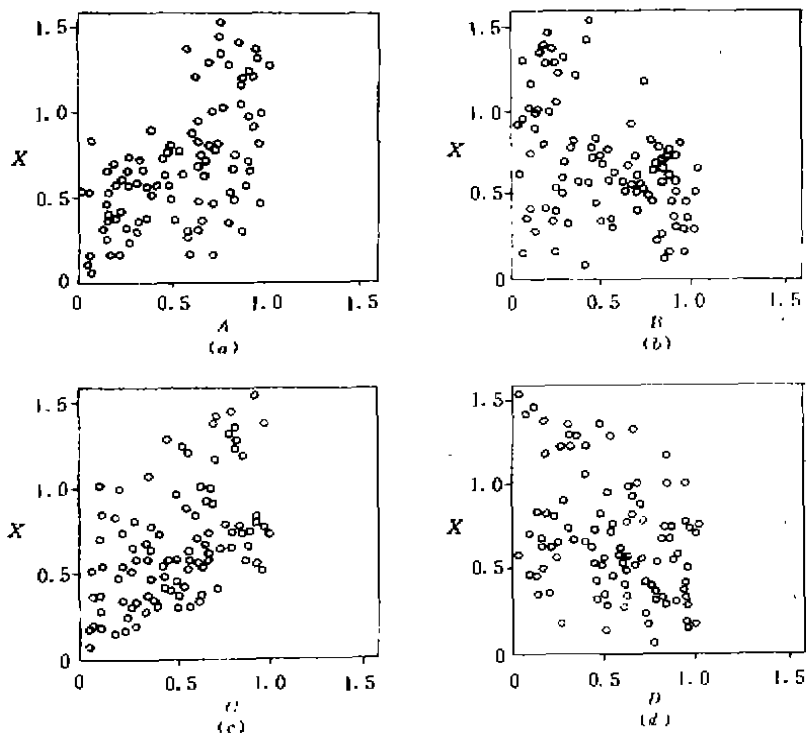


图 3 X 与自变量数据点图

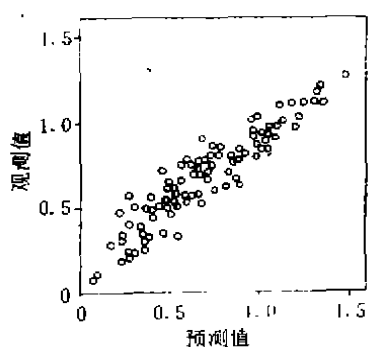


图 4 观测值与回归分析预测值

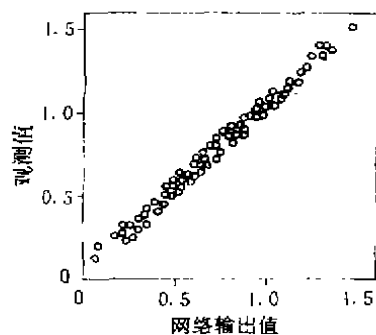


图 5 预测值与网络输出值

BP 网络学习成功后,将 4 个验证样本输入网络,可得到预测值,预测值很接近真实值(表 1),在此网络训练时发现,随着训练时间的增加,网络的均方误差减小,表 1 中的预测值的误差也随之减小。但如果继续增加训练次数,虽然均方误差下降,预测误差却呈现增加趋势,在训练过程中还发现,训练数据略有变化,网络的整体性能不变,这说明 BP 网络有较强的自适应性。

表 1 预测值与真实值(%)

真实值		预测值	
伊利石	蒙皂石	伊利石	蒙皂石
8.70	83.60	9.00	84.81
7.00	86.50	6.51	85.14
5.10	88.90	4.26	89.21
7.60	78.60	7.02	77.49

4.2 BP 网络预测金矿体深部的含矿性

在金矿点评价和金矿床勘探开采过程中,一项重要的工作就是预测矿体深部的含矿性,即预测矿体深部相对于浅部(或地表)矿体的指标(长度、厚度、品位等)的变化和矿体指标的具体值。定量预测多采用统计方法,但数理统计需有大量数据,且数据还应具较佳的分布规律。这一先决条件在矿床普查勘探的初期是难以达到的。而 BP 网络对样本不需求典型的分布规律,我们试图提出应用 BP 网络预测矿床普查勘探初期矿体深部的含矿性。

招掖某金矿床 1 号矿体,据已揭露工程控制得地表(+40m),地下+27m、+12m、-5m 和 -25m 等中段矿体平均厚度、品位和矿体长度(表 2),需预测深部中段的相应数值。在此我们仍采用 3 层 BP 网络,输入层为 1 个节点,代表海拔标高;隐数为 1,节点数为 18;输出层的节点数为 3,代表矿体长度、平均厚度和平均品位。将表 2 中的实测数据作为训练该 BP 网络的学习样本,BP 网络训练成功之后,可得到 -35m 和 -50m 处的含矿性指标预测值,该预测值与用灰色 GM(1,1)模型所得到的预测值^[4]很接近(表 3)。

表 2 某金矿床 1 号矿体部分实测数据

海拔标高(m)	长度(m)	平均厚度(m)	平均品位(g/t)
+40	65	9.23	6.12
+27	175	9.57	6.26
+12	180	9.83	6.90
-6	180	13.41	5.63
-25	215	11.22	6.09

表 3 某金矿床 1 号矿体深部含矿性的 BP 网络预测值

海拔标高(m)	长度(m)	平均厚度(m)	平均品位(g/t)
-35	212(210)	11.75(12.95)	5.85(5.55)
-50	223.5(220.08)	13.46(14.01)	5.65(5.40)

注:括号内数据用 GM(1,1)计算得到^[4]。

蔡煜东等根据矿产资源统计预测的特点,对 BP 网络作了改进^[5],处理矿产资源统计预测问题,得出与数量化理论 I 处理极为相似的结果。认为 BP 网络方法同一般的多因子判别法相比,有如下优点:

(1)容错能力强:神经网络模型中,知识信息采取分布式存储,个别单元损坏不会引起错误,因此用神经网络进行预报识别容错能力强,可靠性高。

(2)判别速度快:训练好的神经网络在对未知样本进行判别时仅需作少量的加法和乘法,因此判别速度快。

有鉴于此,ANN 方法可望成为矿产预测的有效辅助手段^[5]。

5 进一步的讨论

5.1 关于 BP 网络的讨论

(1)在用 BP 网络进行数据处理时,需要提供一定容量的样本供它使用,找出数据间接关系。如果样本过少,网络的训练较难完成,必将影响 BP 网络对各种数据规律的归纳,而达不到解决问题的目的。对于带有误差的样本,需要更多的训练数据,数据的变化范围较大时,需要更多的训练时间。

(2)BP 网络在寻求均方误差最小的过程中,可能存在局部最小而达不到全局最优解以及收敛速度较慢等缺点。可以采用如下改

进方法:①减小学习步长,增加隐层的节点数,或将初始权值重置成许多组不同随机值等,这些对收敛速度的改善不太明显。②改变步长学习算法,最优步长应是随时间变化的。在网络训练过程中,对于误差曲面比较平坦的区域,可以加大步长;遇到谷地区域,误差变化大,此时可放慢搜索速度。

(3)BP 网络中的隐层的数目和每层的节点数全凭经验,目前还缺乏理论依据,因此需要有验证数据。

(4)ANN 不同于传统的程序设计,在处理问题时没有固定的算法,而是用简单的处理单元,强调的是连接方式,从而构成具有并行处理能力、自适应学习能力的网络结构和模拟人脑功能的基本特征。

5.2 地质数据处理中的非线性问题

用 ANN 研究地质现象,也如同用分形理论研究地质现象一样,是非线性科学的一个重要方向。科学的进展是从定性向定量的转变为标志的,作为定量地质学最基本的要求之一是,在解决各类地质问题时给出数量

的准则。以概率论为基础的数学地质工作是定量地质学的主要内容,为地质研究的定量化做出了极大贡献。但是,这一点仍被认为是“巫术统计学”^[6]。虽然,目前许多地质问题尚缺乏定量化的有效手段,但是已有的研究表明,分形理论为复杂地质问题的定量化提供了新途径^[7]。本文的粗浅工作和已有的应用研究表明^[5],ANN 方法也可能为地质数据的非线性问题处理提供了新工具,因为地质科学中的非线性问题比比皆是,地质现象的混沌已为人们所意识到,而神经网络信息处理系统是以非线性为基础的,只是目前神经网络所研究的还是非线性系统的最简单特征。

参考文献

- 1 田世丰. 数学地质浅析,北京,地质出版社,1988,25~42.
- 2 焦李成. 神经网络系统理论,西安,西安电子科技大学出版社,1990,34~36,142~15.
- 3 李英龙等. 化工矿山技术,1993,22(3):19~21,18.
- 4 程小久. 地质与勘探,1990,26(1):28~32.
- 5 蔡煜东等. 矿产与地质,1993,7(3):214~218.
- 6 默克斯 J. W. 国外金属矿山,1993,209(6):4~8.
- 7 孟宪国等. 地球科学,1991,16(2):207~212.

Back Propagation Neural Network and Its Application in Processing Geological Data

Xie Xianping Cai Jianshe Tong Guangxu

The fundamental function selection and coefficient solution are important for function approximation method. The general regress analysis or other cast methods have some disadvantages. The back propagation neural network (BP) can give a good solution. Comparison of theoretical calculation with the test data shows that BP neural network can improve the reliability of data processing and the accuracy of prediction.

Key words: ANN, BP Neural Network, Geological Data Processing, Regress Analysis

(上接第 45 页)

同的接发距、不同发射方式、不同频率等技术手段,排除或减弱外界影响,再综合分析有关资料和工作经验,其结果总的认为比较理想。

4. 仪器对不同的环境(包括土壤结构、管材、管径、附近空间环境等)、不同的操作者来说,都有一定的系统误差(校正系数)。在工作前,应反复试验,细致分析,找准校正系数,对于提高探测工作效率和探测质量都很重要。

5. 对于一个地区、一个工厂的地下管线探测,应遵循先简后繁、从管线稀少地区到管线密集地区的原则。管线密集地段,可按先探

电缆,再探下水管道,最后探金属管道的次序进行。

6. 管线探测的成果是对已有的资料、仪器信号和工作经验三者综合利用的结果。一个好的地下管线探测人员,必须能够灵活应用上述三者,才能取得满意的结果。

7. 利用化探、物探方法探测水污染源和污染方式;对我们来说还是第一次尝试。实践证明,综合方法能快速圈定污染区间,从而缩小探测范围,该方法对今后雨、污分流和地下排水管道中水污染调查都有借鉴意义。