

基于神经网络的地面坡度与其他地形 定量因子关联性分析

——以黄土高原丘陵沟壑区实验为例

张 婷¹, 汤国安², 王 峰³, 陈 莉³, 王 春¹

(1. 西北大学城市与资源学系, 陕西 西安 710069; 2. 南京师范大学江苏省地理信息科学重点实验室, 江苏 南京 210097;

3. 西北大学计算机科学系, 陕西 西安 710069)

摘 要: 不同的地形因子从不同侧面反映地面的起伏特征或空间变异, 各因子之间所存在的相互关联、相互制约、相互影响的程度与特征, 在很大程度上揭示了地形发育与空间变异的内在本质, 因而是地形学研究的重要内容。他以黄土高原丘陵沟壑区的 16 个样本地区为实验样区, 以高分辨率、高精度的 1: 1 万比例尺 DEM 为基准数据, 应用 BP 神经网络模型, 探讨地面坡度与其他地形因子之间的关联性特征。实验结果表明, 利用神经网络分析方法可以有效评价地形因子对地面平均坡度的关联性。该研究方法为进行地貌多定量指标的的选择和多因子之间关联性的量化提供了一种新的方法。

关键词: 神经网络; 地面坡度; 地形因子; DEM; 关联性

中图分类号: P283. 1

文献标识码: A

当前, 基于 DEM 的数字地形分析在地貌学研究中发挥着越来越重要的作用, 并使之有可能实现对多地形要素相互关系的量化研究。不同地形因子以不同的侧面反映地面起伏或变异特征, 然而, 各因子之间存在着明显的关联性, 对这种关联性的分析, 是地形学研究的重要内容。一个地区的平均坡度是该地区地形起伏特征的重要指标, 利用数字高程模型(DEM) 为信息源, 自动提取地面坡度和平均坡度, 已得到广泛的应用^[1, 2]。然而, 地面平均坡度并不是描述区域地形复杂程度的唯一指标, 平均海拔、地面平均曲率、地面起伏度及地面粗糙度等多个地形因子都从不同方面反映地形的起伏与变异特征。研究不同地形因子间所存在的关系, 特别是建立各因子间的定量模型, 对地形地貌学的研究, 具有重要的意义。

目前, 对多因子之间的关联性研究多采用线性方法予以描述。王秀红以中国土地利用程度和效益分区为例, 将多元统计分析的方法应用于主要地理区划的分区研究^[3]。张元明等应用数量分类和排序方法, 对新疆塔里木河中游地区植被进行了多元统计分析, 并对其规律建立了回归模型^[4]。线性方法在一定程度上能够解释因子间的关联性, 但由于自然界的各因素间往往是一种复杂的非线性关系, 单凭线性描述并不能达到研究目标。神经网络技术方法近年在地学界得到诸多有效的应用, 展示了对复杂对象、模糊信息分析与建模的优势。地面起伏度、平均海拔、平均剖面曲率等地形因子从不同侧面影响着地面平均坡度, 它们之间是一种复杂的非线性关系, 用传统的分析方法难以进行模拟。同时, 地学分析具有海量数据的特征, 工作量大, 信息隐含,

收稿日期(Received date): 2004- 01- 10; 改回日期(Accepted): 2004- 04- 15。

基金项目(Foundation item): 国家自然科学基金资助项目(40271089), 国家高技术研究发展计划课题(2001AA130023); 教育部科研基金重点项目(01111); 南京师范大学高层次人才科研启动基金项目[Foundation item: National Nature Science Foundation of China, No. 40271089; National High Technique Research Development Project, No. 2001AA130023; China Education Ministry Science and Technique Key Research Project, No. 0111; Nanjing Normal University High-level Talent Research Startup Fund Item.]

作者简介(Biography): 张婷(1981-), 女, 地图学与地理信息系统专业研究生, 主要从事地理信息系统方面的研究工作。E-mail: dearb-lipop@sina.com [Zhang Ting(1981-), Female, born in Xi'an, China, Master of Northwest University, Research fields mainly covering

传统的分析方法也难以实现对这些数据间关系的分析。神经网络特别适合于数据量庞大、结构复杂的非线性系统,它用类似于黑箱操作的方法,通过对样本的不断学习,学会这些数据间的模式,从而实现对未知样本的预测。

自1989年G. E. McClellan等人首次将神经网络的方法应用于遥感图像分类^[5]开始,目前神经网络的方法在遥感方面主要活跃于遥感影像的分类、专题信息提取等领域^[6]。Y. B. Dibike, D. Olomatine & M. B. Abbott将神经网络与水动力学模型结合,利用水动力学模型为神经网络提供网络学习样本,再用训练好的神经网络对重要河段航深、二维流场中重要位置的水流运动(包括水位、流速、流向及流量等要素)进行预测,取得了令人鼓舞的结果^[7,8]。彭清娥、雷孝章等根据不同坡度、不同耕作方式情况下水土流失的实测资料,利用神经网络方法,建立了不同坡度下顺坡垄作和横坡垄作的次降雨产沙网络模型,不仅揭示了坡耕地水土流失的内在机理,同时为定量研究顺坡垄作改为横坡垄作的水土保持效益提供了科学依据^[9]。王英、李家彪等人采用BP神经网络,建立控制多金属结核分布的地质因素与多金属结核分布的映射关系模型,探讨地形坡度对多金属结核分布的控制作用^[10]。

本研究在广泛总结前人研究成果的基础上,采用高精度1:1万比例尺5 m分辨率的DEM为提取地形要素的基础数据,以黄土高原丘陵沟壑区的16个样本区域为实验样区,应用BP神经网络模型,研究黄土丘陵沟壑区多地形因子与平均坡度之间的量化关系,该研究是神经网络分析方法在数字地形分析中的一次有益探索。

1 BP神经网络实验方法

神经网络具有多种不同的网络模型。本研究在选择网络模型时,考虑到地形要素输入、输出之间具有复杂的非线性关系,需确定各输入因子间的权值大小的时候,选择BP网络模型,根据网络误差平方和(SSE),通过网络学习过程中对权值的反向调整,能够得到各权值的一组最优解。

BP网络是一种多层前馈神经网络,其神经元的

激活函数是Sigmoid型函数,函数表达式为

$$f(x) = 1 / (1 + e^{-x}) \quad (0 < f(x) < 1)$$

式中 f 为输出因子, x 为输入因子, e 为常数。

因此输出量为0到1之间的连续量,它可以实现从输入到输出的任意的非线性映射,其权值的调整采用反向传播(Back Propagation)的学习算法。对于输入信息,首先向前传播到隐含层的节点上,经过各单元的特性为Sigmoid型的激活函数(又称为作用函数或映射函数)运算后,把隐含节点的输出信息传播到输出节点,最后给出输出结果。网络的学习过程由正向和反向传播两部分组成。在正向传播过程中,每一层神经元的状态只影响下一层神经网络。如果输出层不能得到期望输出,就是实际输出值与期望输出值之间有误差,那么转入反向传播过程,将误差信号沿原来的连接通路返回,通过修改各层神经元的权值,逐次地向输入层传播去进行计算,再经过正向传播过程,这两个过程的反复运用,使得误差信号最小。实际上,误差达到人们所希望的要求时,网络的学习过程就结束^[11-18]。确定BP网络的结构后,利用输入输出样本集对其进行训练,也即对网络的权值和阈值进行学习和调整,以使网络实现给定的输入输出映射关系。经过训练的BP网络,对于不是样本集中的输入也能给出合适的输出,即具有泛化(generalization)功能^[15]。

在选取数据时,应预先留出一小部分数据用于检测。对于BP网络,给出网络的输入量、目标输出,通过对一系列数据的训练,模拟出这些数据的输入与目标输出的关系模型,再用预留数据进行检测,若实际输出与目标输出之间的误差在给定误差范围内,则该网络训练成功,其给出的各个相应权值即为实际权值。

标准的BP网络是由输入层,一个或多个隐层,一个输出层组成。每个点只与邻节点相连接,同一层的节点彼此不相连。其拓扑结构如图1所示。其中, R : 输入数; $S1$: 第1层神经元数; $S2$: 第2层神经元数; $w1$: 输入层与隐含层之间的权向量; $b1$: 输入层与隐含层之间的阈值向量; $w2$: 隐含层与输出层之间的权向量; $b2$: 隐含层与输出层之间的阈值向量; $a1$: 神经元层1(即隐含层)的激活函数; $a2$: 神经元层2(即输出层)的激活函数

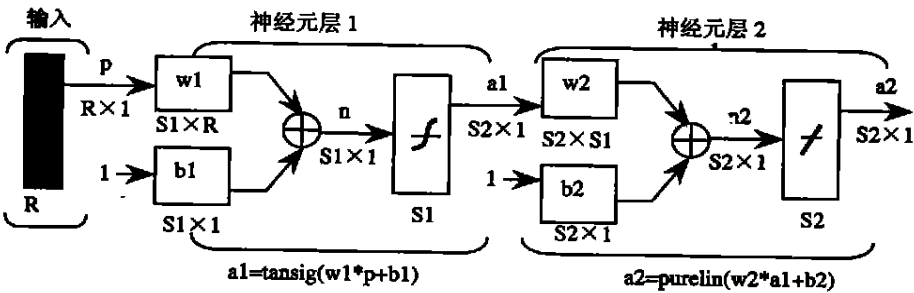


图 1 具有一个隐层的 BP 网络

Fig. 1 BP Network with a latent layer

2 实验方法与结果

2.1 实验数据与预处理

选择陕北延川县黄土丘陵沟壑区的 16 个不同地面复杂度的区域为实验样区, 实验区面积均为 1 km × 1 km。样区的选择体现地貌类型典型、实验基础数据完备、实验区面积大小满足实验要求等条件。

所采用的 DEM 由 1: 1 万地形图等高线数字化, 再经高程内插获得, 栅格分辨率为 5 m。

以 ArcGIS 地理信息系统软件为基本实验软件, 分别提取这 16 个样区地面起伏度、地面粗糙度、平均剖面曲率、海拔标准差与平均坡度等 9 个地形因子, 作为 BP 网络的输入、输出向量。将数据送入网络运算前对输入、输出向量进行极差标准化, 以消除因子量纲影响(表 1)。

表 1 极差标准化后的部分结果

Table 1 Parts of the terrain factors after standardized

实验样区	海拔高差	平均海拔	平均剖面曲率	平均平面曲率	地面粗糙度 (平均)	平均坡长	平均坡度
1	0.949 4	1.000 0	0.683 5	0.306 5	0.315 9	0.168 1	0.445 7
2	1.000 0	0.812 8	0.606 9	0.287 8	0.574 8	0.588 9	0.670 4
3	0.174 0	0.004 4	0.105 2	0.833 7	0.204 5	0.872 1	0.432 9
4	0.358 3	0.196 9	0.017 5	0.867 8	0.102 2	0.772 1	0.094 7
.....
15	0.333 6	0.319 5	0.200 3	0.911 2	0.102 3	0.788 4	0.221 2
16	0.738 9	0.726 6	1.000 0	0.000 0	0.394 9	0.000 0	0.464 9

2.2 各地形因子影响程度的确定及网络的预测

BP 网络模型是由输入层、隐含层与输出层组成。带有隐含层的 BP 网络, 输入层与隐含层有一组对应的权值和阈值, 同时隐含层与输出层之间也有一组对应的权值和阈值。如图 1 所示, 神经元层 1(即隐含层)与神经元层 2(即输出层)之间的权值 w2 与阈值 b2。对整个网络来说, 各输入因子与目标输出间的关系不明确; 但含有隐层的 BP 网络, 对数据间的关系模式学习得较好, 预测精度高。不含隐层的 BP 网络, 只有输入、输出两层, 网络学习的结果能够反映出输入层与输出层间的权值关系, 但

网络学习的效果不如带隐含层的效果好, 预测精度差。

2.2.1 不带隐含层的 BP 网络

不带隐含层的 BP 网络只有输入和输出两层, 训练的结果能够清楚地反映出各个输入因子间的权值关系。

选用不带隐含层的 BP 网络模型后, 首先需确定网络的训练函数、仿真函数(即预测函数)、传递函数及相应的各参数。根据选定的网络参数, 对这 8 个输入和一个输出建立模型, 得出各个输入因子相对于输出的权值关系(即关联度), 见表 2。

表 2 8 个输入因子对平均坡度的关联度
Table 2 Eight input factors with their relevancy degree

输入因子	海拔高差	平均海拔	海拔标准差	平均剖面曲率	平均平面曲率	地面粗糙度 (平均)	地面起伏度 (平均)	平均坡长
权值	- 0.034 946	- 0.094 886 1	0.011 889 343	- 0.412 33	0.331 197 143	1.609 985 714	0.1470 042 86	0.4136 785 71

在比较各个因子的权值大小时, 为判断负值的影响大小, 首先去掉平均剖面曲率这一项, 比较其前后 SSE 和仿真效果。去掉平均剖面曲率后, 网络训练的误差平方和(SSE) 变大, 仿真效果变差。可知, 平均剖面曲率对平均坡度的关联较大, 即可说明权值大小的比较不能以正负来判断, 负值只是表示作用因子为负向因子, 应以绝对值的大小来判断。

为了验证这一点, 去掉海拔标准差来做进一步的研究。去掉海拔标准差后 SSE 和仿真效果变化比去掉平均海拔后的变化小得多, 由此可验证权值大小应以绝对值的大小来判断。
由上述研究可得出, 各因子按权值由大到小依次如表 3 所示。

表 3 8 个输入因子对平均坡度影响大小的排序
Table 3 Arrangement of eight input factors according to the weights

地形因子	地面粗糙度 (平均)	平均坡长	平均剖面 曲率	平均平面 曲率	地面起伏度 (平均)	平均海拔	海拔高差	海拔标准差
权值(绝对值)	1.609 985 714	0.413 678 57	0.412 33	0.331 197 143	0.147 004 286	0.094 886 143	0.034 946	0.010 892
排序结果	1	2	2	4	5	6	7	8

随后, 对于各个输入进行筛选, 对于一些影响较小或相关性较强的因子进行适当选取, 以减少计算的复杂程度和计算量。
由于选择的 8 个地形因子中, 海拔高差、平均海拔和海拔标准差之间具有较强的相关性, 对输出(即平均坡度) 的取值有着较大的影响, 因此需对这 3 个

因子进行筛选。而平均海拔的关联程度较大, 因此, 分别去掉海拔高差和海拔标准差, 进行训练和检测, 比较他们的 SSE 和预测精度。再将海拔高差和海拔标准差同时去掉, 比较前后的 SSE 和预测精度, 得出表 4。

表 4 去掉不同因子后的结果
Table 4 The results after cancellation of different input factors

去掉不同的因子	SSE	误差百分比 (第 15 个样本)	误差百分比 (第 16 个样本)	误差百分比 (第 14 个样本)	误差百分比 (第 12 个样本)
去掉海拔标准差	0.939 743 94%	- 0.262 880 888	0.115 360 212	- 0.030 356 376	0.037 716 333
去掉海拔高差	2.278 466 03%	- 0.280 062 941	0.104 390 936	- 0.026 853 834	0.039 224 142
去掉海拔高差、海拔标准差	2.634 537 99%	- 0.263 333 047	0.105 251 272	- 0.021 828 449	0.041 938 198

通过这三种结果的比较, 可以发现, 去掉海拔标准差与海拔高差之后, 网络训练的速度大大加快, 网络的精度没有太大变化。这说明, 海拔标准差与海拔高差对网络输出没有多大贡献, 可以忽略不计。所以, 对于海拔高差、平均海拔、海拔标准差这 3 个强相关的因子, 可以去掉海拔标准差和海拔高差, 用平均海拔来代替。

对筛选后的 6 个地形因子重新进行训练, 得出各个地形因子对地面坡度关联程度的排序结果(表 5)。
2.2.2 带有隐含层的 BP 网络

根据不带隐含层的 BP 网络的实验结果, 将筛选出的这 6 个地形因子作为网络的输入因子, 平均坡度作为网络的输出, 建立模型, 对这几个样区的地

形因子间的内在关系进行模拟。

首先确定网络的训练函数、仿真函数(即预测函数)、传递函数及相应的各参数。网络参数确定后, 建立模型, 根据给出的输入值进行仿真, 并计算仿真结果的误差值, 从而实现对未知样本的预测(仿真结果见表 6)。

表 5 各因子按权值排序的结果
Table 5 Factors arranged in weight order

地形因子	地面粗糙度(平均)	平均剖面曲率	平均坡长	平均平面曲率	地面起伏度(平均)	平均海拔
权值	1. 616 5	0. 418 14	0. 392 19	0. 289 51	0. 134 9	0. 126 83
排序结果	1	2	3	4	5	6

表 6 由带隐含层的 BP 网络得出的仿真结果
Table 6 Simulating results from the BP network with a latent layer

SSE	第 15 个样本的误差	第 16 个样本的误差	第 14 个样本的误差	第 12 个样本的误差
0. 000 989 334	- 3. 001 885 50%	- 6. 746 104 29%	- 0. 340 203 39%	0. 484 609 79%

由表 6 可以看出, 网络的预测精度可达到 93. 2% 以上, 网络误差的平方和(SSE) 达到 0. 00099 以下, 说明了得出的地形因子影响程度的正确性。

3 结论与讨论

1. 在黄土丘陵沟壑区, 各个地形因子与平均坡度的关联程度有较大的差异。地面粗糙度、平均剖面曲率和平均坡长的权值较大, 即对 DEM 所提取的平均坡度的影响较大, 而平均平面曲率、地面起伏度、平均海拔的影响相对较小。

2. 神经网络由其本身类似于黑箱操作的特点, 在数据处理中可避免数据分析和建模中的困难。特别适用于不确定性和非结构化信息处理, 对地学中各种未知信息的预测有着较好的适用性。同时, 便于模拟出多种因素间的复杂关系, 这是传统的统计分析与建模方法所难以实现的。本研究在模型的建立过程中显示, 带隐含层的 BP 网络预测精度高; 不带隐含层的 BP 网络权值清晰, 可揭示各地形因子对平均坡度关联程度的大小。两种方法的结合使用, 可取得理想的模拟效果。

3. 本研究所得到的模拟结果与先前进行统计分析结果基本吻合, 但研究中并未涉及对黄土丘陵沟壑区多地形要素关联程度内在机理的分析。拟在今后的研究中进行黄土高原多地貌类型区的地形定量因子关联性试验, 在此基础上, 从地形地貌学与数字地形表达多角度揭示其内在的机理与外部的条件, 以深化对黄土高原地形地貌规律的认识。同时,

也为神经网络技术在数字地形分析中的应用提供有益的实践。

参考文献(References):

[1] Li Zhilin, Zhu Qing. The Digital Elevation Model[M]. WuHan University Press. 2001. [李志林, 朱庆. 数字高程模型[M]. 武汉大学出版社, 2001.]

[2] Tang Guoan, Zhao Mudan. The Geographical Information System [M]. Science Press. 2000. [汤国安, 赵牡丹. 地理信息系统 [M]. 科学出版社, 2000.]

[3] Wang Xuhong. Application of Multivariate Statistical Analysis in Regionalization Study [J]. *Saentia Geographica Sinica*. 2003, (1): 66~ 71. [王秀红. 多元统计分析在分区研究中的应用[J]. 地理科学. 2003, (1): 66~ 71.]

[4] Zhang Yuanming, Chen Yaning, Zhang Daoyuan. Plant Communities and Their Interrelations with Environmental Factors in the Middle Reaches of the Tarim River [J]. *Acta Gaograp hica Sinica*. 2003, (1): 109~ 118. [张元明, 陈亚宁, 张道远. 塔里木河中游植物群落与环境因子的关系[J]. 地理学报. 2003, (1): 109~ 118.]

[5] Mc Clellan G E, De Witt R N, Hemmer T H, *et al.* Multi-spectral image processing with a three-layer back-propagation network [J]. *Proc IJCNN*. 1989, (1): 151~ 153.

[6] Ding Jian-li, Tashpolat. TIYIP, Liu Chuan-sheng. Artificial Neural Network Model and it's Application in Remote Sensing [J]. *Journal of Xinjiang University*. 2001, (3): 269~ 276. [丁建丽, 塔西甫拉提·特依拜, 刘传胜. 人工神经网络模型及其在遥感中的应用[J]. 新疆大学学报(理工版). 2001, (3): 269~ 276.]

[7] VITTORIO MANIEZZO. Genetic Evolution of the Neural Networks [J]. *IEEE, Trans on Neur al Networks*, 1994, 5 (1): 39~ 53.

[8] GILDERG D E. Genetic Algorithm in Search, Optimization, and Machine Learning [M]. Addison-Wesley, Reading, MA, 1989.

- [9] Peng Qing-e, Cao Shu-you, Liu Xing-nian, *et al.* BP Neural Network Models of Soil Erosion on Slope[J]. *Journal of Soil and Water Conservation*. 2002, (3): 79~ 82. [彭清娥, 曹叔尤, 刘兴年, 等. 坡面产沙BP神经网络模型研究[J]. 水土保持学报. 2002, (3): 79~ 82.]
- [10] Wang Ying, Li Jia-biao, Han Xi-qiu, *et al.* The influence of terrain slope on the distribution of polymetallic nodules [J]. *Acta Oceanologica Sinica*. 2001, (1): 60~ 65. [王英, 李家彪, 韩喜球, 等. 地形坡度对多金属结核分布的控制作用[J]. 海洋学报. 2001, (1): 60~ 65.]
- [11] Zeng-ren Yuan. Artificial Neural Network and it's Application [M]. Beijing: Tsinghua University Press. 1999. [袁曾任. 人工神经网络及其应用[M]. 北京: 清华大学出版社. 1999.]
- [12] Xin Wen, Lu Zhou, Dan-li Wang, *et al.* MATLAB Application Design in Neural Network [M]. Beijing: Science Press. 2002. [闻新, 周露, 王丹力, 等. MATLAB神经网络应用设计[M]. 北京: 科学出版社. 2002.]
- [13] Li cheng Jiao. Neural Network Calculation [M]. Xi'an: Xi'an Electronic University Press. 1995. [焦李成. 神经网络计算[M]. 西安: 西安电子科技大学出版社. 1995.]
- [14] Zeng-liang Liu. Indefinite Technology and Application Collection (3) [C]. Beijing: Beijing Aerospace and Aviation University Press. 1998. [刘增良. 模糊技术与应用选编(3)[C]. 北京: 北京航空航天大学出版社. 1998.]
- [15] Shun-tian Lou, Yang Shi. The Systematical Analysis and Design of Neural Network Based on MATLAB [M]. Xi'an: Xi'an Electronic University Press. 1998. [楼顺天, 施阳. 基于MATLAB的系统分析与设计——神经网络[M]. 西安: 西电出版社. 1998.]
- [16] Li-ming OuYang. Dominative System Control of MATLAB [M]. Beijing: National Defense Industrial Press. 2001. [欧阳黎明. MATLAB控制系统设计[M]. 北京: 国防工业出版社. 2001.]
- [17] Yun Zhou Studio. The Basical Tutorial of MATLAB6 Mathematical Modelling [M]. Beijing: Demos Post Press. 2001. [云舟工作室. MATLAB6数学建模基础教程[M]. 北京: 人民邮电出版社. 2001.]
- [18] Du Yajun, Hou Lanjie, Lihui. The Application of BP Neural Network to Geology[J]. *Journal of Sichuan Geology*. 1999, (1): 71~ 74. [杜亚军, 侯兰杰, 李辉. BP人工神经网络在地学中的应用[J]. 四川地质学报. 1999, (1): 71~ 74.]

Correlation Between Mean Slope and Other Quantitative Terrain Factors Based on the Neural Network ——A Case Study in the Gully Hill Areas of China Loess Plateau

ZHANG Ting¹, TANG Guoan^{1,2}, WANG Zheng³, CHEN Li³, WANG Chun¹

1. Department of Urban and Resource Science, Northwest University, 710069, Xi'an, P. R. China;

2. Pivot Laboratory of Geographical Science in Jiangsu Province, Nanjing Normal University, 210097, Nanjing, P. R. China; 3. Department of Computer Science, Northwest University, 710069, Xi'an, P. R. China)

Abstract: Different factors reflect the relief of landform, with remarkable relationships between them. Analysis of these relationships is important in topographic study. This paper focuses on the relationships between the mean slope and other factors by means of the Back Propagation model of Neural Network. Sixteen Loess gully-hill areas are selected as test areas, and the relevant 1: 10 000 scale DEMs of high resolution and high precision are also used as basic data. Results show that this method can effectively evaluate the relevancy degree between terrain factors on the mean slope extracted from DEMs. This paper provides a new method for selecting multiple topographical factors and the estimation of the relevancy between these factors.

Keywords: Neural Network; the Loess Plateau; terrain factor; DEM; BP Network