

基于 BP 神经网络的森林生态系统健康预测

武巧英, 陈丽华, 李晓凤, 王 鹏, 卞西陈

(北京林业大学 水土保持及荒漠化教育部重点实验室, 北京 100083)

摘 要: 森林的健康状况将直接影响全球生态稳定和全人类的可持续发展, 而对森林健康预测是人们了解森林生态状况的重要手段。在选定森林健康评价指标的基础上, 建立了一个 3 层 BP 神经网络模型, 并应用于鹫峰国家森林公园森林健康预测。预测验证结果表明, 基于神经网络预测森林生态系统健康具有良好的合理性与可靠性。并为了构建一个准确稳定的森林健康预测模型, 给出了一些相关建议。

关键词: 森林生态系统; BP 神经网络; 健康预测

文献标识码: A

文章编号: 1000-288X(2011)02-0150-05

中图分类号: S718.55

Prediction of Forest Ecosystem Health Based on BP Neural Network

WU Qiao-ying, CHEN Li-hua, LI Xiao-feng, WANG Peng, BIAN Xi-chen

(Key Laboratory of Soil and Water Conservation and Desertification Combating, Ministry of Education, Beijing Forestry University, Beijing 100083, China)

Abstract: The health of forests affects directly global ecological stability and sustainable development. The prediction of forest health is an important means for people to understand forest health status. Based on the selected indicators of forest health, a model has been constructed to predict the health of Jiufeng National Forest Park using a three-layer BP neural network. The results show that the adoption of neural network in forest ecosystem health forecasting was reasonable and reliable. Some important conclusions and recommendations have also been given for constructing a precise and stable prediction model.

Keywords: forest ecosystem; BP neural network; health prediction

森林是维持地球动态平衡的根本保障, 是一个无价的生态、经济、文化系统^[1]。森林的健康状况将直接影响全球生态稳定和全人类的可持续发展, 而对森林健康预测是人们了解森林生态状况的重要手段, 因此对森林健康进行预测有着非常重要的意义^[2-4]。森林健康全称为森林生态系统健康和恢复, 是近几年来新发展起来的一个学科术语^[5]。其基本理念主要是由西方国家针对人工造林林分结构单一、森林病虫害防治能力和水土保持能力薄弱等问题提出的^[6]。森林健康同时受多种因素的影响, 各因素之间存在复杂的相互作用。因此, 运用传统的方法很难建立起一个精确和完善的预测模型。神经网络算法具有对非线性复杂系统预测的良好特性, 可以有效地描述森林生态系统健康预测中具有的不确定性、多输入等复杂的非线性特性。

本研究在选择森林健康评价指标后, 通过对训练样本的学习来构造一个 BP 神经网络, 待网络模型稳

定后, 以新的测试样本作为输入数据进行模拟, 将得到的结果与实际值进行比较, 从而验证了基于神经网络预测森林生态系统健康的合理性与可靠性。

1 材料与方法

1.1 研究区概况及数据采集

鹫峰国家森林公园位于北京市西北部, 太行山北部与燕山东端交汇处, 北纬 39°53', 东经 116°28', 距市区约 30 km², 面积 811.73 hm², 横跨海淀和门头沟 2 个区。境内山峦绵延, 地形复杂, 最低处海拔仅 70 m, 最高处达 1 100 m, 坡度在 36°的山地占 1/4。岩石有花岗岩、石灰岩、凝灰岩、砂岩、砾岩等。海拔 900 m 以下以山地淋溶褐土为主, 900 m 以上为棕壤, 土层多较薄, 石砾含量大, pH 值为 6.4~8.4。年平均气温 12.2 °C, 年均降水量 700 mm, 多集中在 7—8 月。植被类型属于温带落叶阔叶林、山地栎林和油松林带, 森林覆盖率达 60%。

收稿日期: 2010-09-06

修回日期: 2010-12-10

资助项目: 林业公益性行业科研专项经费项目“典型区域森林生态系统健康维护与经营技术研究”(200804022A)

作者简介: 武巧英(1984—), 女(汉族), 山西省朔州市人, 硕士研究生, 研究方向为森林健康经营。E-mail: wqysm@163.com。

通信作者: 陈丽华(1957—), 女(汉族), 北京市人, 博士, 教授, 研究方向为森林水文、流域治理。E-mail: C_ljhua@bjfu.edu.cn。

试验数据来源于 2009 年 7 月 9 日至 7 月 19 日于北京市鹫峰国家森林公园的实地采集(样地规格设为 25 m×25 m)资料。试验区中多为人工林,人工林以侧柏〔*Platycladus orientalis*〕,油松(*Pinus tabulaeformis*),栓皮栎(*Quercus variabilis*)为主,高海拔的地方有人工落叶松林。在几个主要林分中设置 5 块样地进行调查,包括栓皮栎人工林,油松、侧柏混交林,油松人工林,原生灌木林,落叶松人工林。各样地的基本情况如表 1 所示。采集内容包括:样地基本

情况,如纬度、海拔、坡度、坡向、树种组成、林分郁闭度、林龄结构、灌木总盖度、草本总盖度等。植被调查以 10 m×10 m 的小样方为单位,对高度大于 1 m 的乔木进行每木检尺,记录其树种、胸径、树高、冠幅和生长状况(腐倒木、枯立或正常);样地内设置 4 个 5 m×5 m 的灌木样方及 4 个 1 m×1 m 的草本样方,对林内灌木及草本植物,分别记录其种类、多度、盖度、平均高度、株数、生长状况等;在样地内挖掘土壤剖面,将土壤样品带回分析。

表 1 试验区样地基本情况

林分类型	起源	主要乔灌木
栓皮栎混交林	人工林	栓皮栎,刺槐,构树,小叶鼠李,荆条,孩儿拳头。
油松、侧柏混交林	人工林	油松,栎类,侧柏,孩儿拳头,荆条,小叶鼠李。
油松人工纯	人工林	油松,山杏,落叶松,刺槐,三裂绣线菊。
原生灌木林	天然林	雀儿舌头,土庄绣线菊,三裂绣线菊,荆条,酸枣。
落叶松人工纯林	人工林	落叶松,白蜡,油松,山荆子,胡枝子。

1.2 预测方法和模型构建

美国学者 Rumelhart 等人于 1985 年提出了误差反传递算法(error back propagation),简称 BP 算法,它由信息的正向传播和误差的反向传播两个过程组成。输入层各神经元负责接收来自外界的输入信息,并传递给中间层各神经元;中间层是内部信息处理层,负责信息变换,根据信息变化能力的需求,中间层可以设计为单隐层或者多隐层结构;最后一个隐层传递到输出层各神经元的的信息,经进一步处理后,完成一次学习的正向传播处理过程,由输出层向外界输出信息处理结果。

当实际输出与期望输出不符时,进入误差的反向传播阶段。误差通过输出层,按误差梯度下降的方式修正各层权值,向隐层、输入层逐层反传。周而复始的信息正向传播和误差反向传播过程,是各层权值不断调整的过程,也是神经网络学习训练的过程,此过程一直进行到网络输出的误差减少到可以接受的程度,或者预先设定的学习次数为止^[7]。

1.2.1 构造网络输入层和输出层 在前人研究的基础上^[8-11],选取“物种多样性”、“土壤厚度”、“叶面积指数”、“土壤有机质”、“病虫害等级”、“林分受损程度”作为森林健康评价指标。并按照高志亮在“北京森林健康示范区森林生态系统健康评价与预警研究”一文中的指标测度分析原则^[12],对各指标进行赋值量化。

表 2 为归一化后的各指标的得分因子,以这 6 个指标作为神经网络的输入项。

将森林的健康等级划为 5 级,并且以向量编码(1 0 0 0 0),(0 1 0 0 0),(0 0 1 0 0),(0 0 0 1 0),

(0 0 0 0 1)的形式表示。按照李静锐等人对森林健康等级的划分评价标准^[13],这 5 个等级分别对应优质健康(>7.0 分),良好健康(6.0~7.0 分),健康(5.0~6.0 分),亚健康(4.8~5.0 分),不健康(<4.8 分)。各样地的实际健康状况如表 3 所示,以森林的健康等级作为网络的输出项。

BP 网络的输入层和输出层的神经元数目是由输入向量和输出向量的维数确定的。输入向量的维数也就是森林健康评价指标的个数,所以输入层的神经元个数为 6。将前 4 个样地(栓皮栎混交林,油松、侧柏混交林,油松人工纯林,原生灌木林)数据作为输入样本数据来训练网络,将最后一个样地(落叶松人工纯林)数据作为测试样本数据来验证网络。

健康等级作为网络的目标输出,因此输出层神经元个数为 5。由于输出向量的元素为 0~1 值,因此,输出层神经元的传递函数可选用 S 型对数函数 lgsgn()。

1.2.2 构造网络隐含层 隐含层数目越多,BP 网络的非线性映射能力越强。但是隐含层数目不宜过大,否则会降低网络的性能。由于单隐层的 BP 神经网络可以逼近一个任意的连续非线性函数,因此采用单隐层的 BP 网络。根据 kolmogorov 定理设定网络的隐含层神经元个数为 13。按照一般的设计原则,隐含层神经元的传递函数为 S 型正切函数 tansig()。

1.2.3 选择训练函数 网络结构确定后,需要根据一定的学习规则对网络进行训练,直到网络收敛。选用的训练函数为 traingdx(),该函数以梯度下降法进行学习,并且学习速率是自适应的^[14]。

表 2 归一化后的各评价指标的得分因子

样地类型	样方号	物种多样性	土壤厚度	叶面积指数	土壤有机质	病虫害等级	林分受损程度
栓皮栎 混交林	1	0.925 0	0.476 1	0.213 0	0.354 7	1.000 0	0.000 0
	2	0.474 0	0.243 6	0.418 6	0.267 2	0.000 0	0.500 0
	3	0.620 0	0.523 8	0.345 5	0.254 6	0.000 0	0.000 0
	4	0.740 0	0.324 0	0.264 7	0.541 2	0.500 0	0.000 0
油松侧柏 混交林	5	0.944 0	0.619 0	0.684 4	0.475 4	0.000 0	0.500 0
	6	0.573 0	0.241 0	0.426 6	0.132 2	0.500 0	1.000 0
	7	1.000 0	0.523 8	0.545 5	0.354 6	0.500 0	0.500 0
	8	0.384 5	0.000 0	0.263 0	0.358 5	0.500 0	0.000 0
油松人工 纯林	9	0.184 4	0.619 0	0.684 4	0.000 0	0.000 0	0.000 0
	10	0.073 0	0.543 0	0.926 6	0.132 2	0.500 0	0.500 0
	11	0.272 7	0.523 8	0.445 5	0.254 6	0.500 0	0.500 0
	12	0.074 0	0.724 0	1.000 0	0.386 6	0.500 0	1.000 0
原生灌木林	13	0.138 3	1.000 0	0.754 2	0.742 0	1.000 0	0.500 0
	14	0.124 4	0.945 5	0.426 6	1.000 0	0.500 0	1.000 0
	15	0.095 4	0.523 8	0.000 0	0.654 6	1.000 0	0.000 0
	16	0.321 0	0.624 0	0.243 0	0.793 2	0.500 0	0.500 0
落叶松 人工纯林	17	0.274 0	0.693 6	0.418 6	0.267 2	0.500 0	1.000 0
	18	0.092 4	0.512 0	0.845 3	0.162 0	0.500 0	0.500 0
	19	0.093 8	0.424 0	0.942 5	0.386 6	0.500 0	1.000 0
	20	0.143 0	0.347 0	0.326 6	0.232 2	0.000 0	0.000 0

表 3 各样地的实际健康状况

样地类型	样方号	健康指数	健康等级	健康状况
栓皮栎 混交林	1	4.84	00010	亚健康
	2	5.45	00100	健康
	3	5.94	00100	健康
	4	6.30	01000	良好健康
油松、侧柏 混交林	5	7.22	10000	优质健康
	6	5.20	00100	健康
	7	7.40	10000	优质健康
	8	4.62	00001	不健康
油松人工 纯林	9	4.08	00001	不健康
	10	4.25	00001	不健康
	11	4.70	00001	不健康
	12	4.88	00010	亚健康
原生 灌木林	13	7.80	10000	优质健康
	14	7.65	10000	优质健康
	15	5.74	00100	健康
	16	6.43	01000	良好健康
落叶松人 工纯林	17	5.36	00100	健康
	18	4.48	00001	不健康
	19	4.94	00010	亚健康
	20	4.31	00001	不健康

2 结果与分析

2.1 试验结果

经过 BP 网络建立及训练的 matlab 程序,如图 1 所示,网络经过 509 次迭代后趋于稳定。网络训练结

束后,还必须选用一组样地数据对其进行测试,所谓测试,就是利用仿真函数得到网络输出,然后通过计算出与实际值的误差,来检验测试结果的正确性。取第 5 组样地(落叶松人工纯林)数据为测试样本数据。

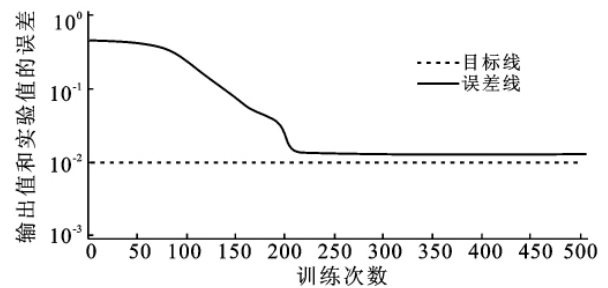


图 1 BP 网络模型训练误差曲线

输出结果为:

$$y = \begin{bmatrix} 0.000\ 0 & 0.000\ 0 & 0.000\ 0 & 0.000\ 0 \\ 0.000\ 0 & 0.000\ 0 & 0.000\ 0 & 0.000\ 0 \\ 0.899\ 7 & 0.000\ 0 & 0.000\ 0 & 0.387\ 2 \\ 0.020\ 1 & 0.000\ 0 & 0.999\ 9 & 0.000\ 0 \\ 0.038\ 8 & 1.000\ 0 & 0.000\ 0 & 1.000\ 0 \end{bmatrix}$$

2.2 结果分析

重复 2.1 中的试验 9 次,发现每一次试验得到的输出矩阵 y 都是不同的,这是因为每次重复模拟都会得到不同的权值矩阵,而输出是由输入矩阵和权值矩阵决定的。通过 10 次试验模拟,得到了 40 组预测健康等级(表 4)。

表 4 健康等级预测结果

试验次数	样方号	实际健康等级	预测健康等级					预测结果
第 1 次	17	0 0 1 0 0	0.0000	0.0000	0.8997	0.0201	0.0388	正确
	18	0 0 0 0 1	0.0000	0.0000	0.0000	0.0000	1.0000	正确
	19	0 0 0 1 0	0.0000	0.0000	0.0000	0.9999	0.0000	正确
	20	0 0 0 0 1	0.0000	0.0000	0.3872	0.0000	1.0000	正确
第 2 次	17	0 0 1 0 0	0.0788	0.0000	0.7268	0.0590	0.1286	正确
	18	0 0 0 0 1	0.0071	0.0002	0.0129	0.0073	0.9860	正确
	19	0 0 0 1 0	0.0040	0.0003	0.0414	0.7617	0.0985	正确
	20	0 0 0 0 1	0.0001	0.0136	0.0850	0.0003	0.9805	正确
第 3 次	17	0 0 1 0 0	0.0000	0.0000	0.0008	0.6961	0.0016	错误
	18	0 0 0 0 1	0.0000	0.0000	0.0390	0.0000	1.0000	正确
	19	0 0 0 1 0	0.0000	0.0000	0.0000	0.9958	0.0017	正确
	20	0 0 0 0 1	0.0000	0.0000	0.9453	0.0000	0.9998	错误
第 4 次	17	0 0 1 0 0	0.4908	0.0193	0.0015	0.4194	0.0390	错误
	18	0 0 0 0 1	0.0087	0.0000	0.0272	0.0208	0.9996	正确
	19	0 0 0 1 0	0.1990	0.0284	0.0007	0.9397	0.0000	正确
	20	0 0 0 0 1	0.0097	0.0062	0.6559	0.0000	0.1306	错误
第 5 次	17	0 0 1 0 0	0.0099	0.0044	0.9256	0.0471	0.0522	正确
	18	0 0 0 0 1	0.0002	0.0005	0.0453	0.0434	0.9744	正确
	19	0 0 0 1 0	0.0038	0.0002	0.0129	0.9659	0.0540	正确
	20	0 0 0 0 1	0.0037	0.0522	0.0664	0.0010	0.7322	正确
第 6 次	17	0 0 1 0 0	0.0085	0.0001	0.9146	0.0001	0.0097	正确
	18	0 0 0 0 1	0.0040	0.0000	0.0733	0.0268	0.9893	正确
	19	0 0 0 1 0	0.0087	0.0007	0.1139	0.8681	0.0462	正确
	20	0 0 0 0 1	0.0002	0.0137	0.9661	0.0002	0.5175	错误
第 7 次	17	0 0 1 0 0	0.0000	0.0000	0.0008	0.1048	0.0019	错误
	18	0 0 0 0 1	0.0000	0.0000	0.0000	0.0370	1.0000	正确
	19	0 0 0 1 0	0.0000	0.0000	0.0000	1.0000	0.2760	正确
	20	0 0 0 0 1	0.0000	0.0000	0.0650	0.0080	0.9996	正确
第 8 次	17	0 0 1 0 0	0.2714	0.0001	0.1498	0.6274	0.0342	错误
	18	0 0 0 0 1	0.0111	0.0000	0.0123	0.2349	0.9882	正确
	19	0 0 0 1 0	0.1163	0.0005	0.0066	0.6371	0.1292	正确
	20	0 0 0 0 1	0.0002	0.0387	0.0903	0.0008	0.9629	正确
第 9 次	17	0 0 1 0 0	0.0361	0.0030	0.7275	0.0485	0.0570	正确
	18	0 0 0 0 1	0.0000	0.0101	0.0045	0.0699	0.9741	正确
	19	0 0 0 1 0	0.0032	0.0311	0.0286	0.7576	0.1108	正确
	20	0 0 0 0 1	0.0000	0.0221	0.1091	0.0113	0.9854	正确
第 10 次	17	0 0 1 0 0	0.0023	0.0009	0.9266	0.0416	0.0144	正确
	18	0 0 0 0 1	0.0171	0.0000	0.0000	0.0315	0.9353	正确
	19	0 0 0 1 0	0.0656	0.0020	0.0000	0.9680	0.1671	正确
	20	0 0 0 0 1	0.0001	0.0042	0.0957	0.0000	0.7029	正确

对比表 4 中的实际健康等级与预测健康等级,可知网络输出的预测健康等级与实际健康等级都为 5 维向量。为便于结果分析,本研究做一些定义。

(1) 令 5 维向量 $T_n = [t_{n1}, t_{n2}, t_{n3}, t_{n4}, t_{n5}]$, $n = 1, 2, \dots, 40$ 和 $S_{ni} = [s_{n1}, s_{n2}, s_{n3}, s_{n4}, s_{n5}]$ 分别表示第 n 次预测得到的健康等级和对应的实际健康等级;

(2) 令 $f_{ni} = \frac{t_{ni}}{t_{n1} + t_{n2} + t_{n3} + t_{n4} + t_{n5}}$, $n = 1, 2, \dots,$

40 ; $i = 1, 2, 3, 4, 5$, 则 f_{ni} 表示第 n 次预测得到的健康等级向量中第 i 维的“得分”($0 \leq f_{ni} \leq 1$);

(3) 当下面两个条件同时成立时,说明预测健康等级与实际健康等级相符,预测结果是合理的。

① $f_{ni} \geq \xi$ ($n = 1, 2, \dots, 40$; $i = 1, 2, 3, 4, 5$);

② 与 f_{ni} 对应的 $S_{ni} = 1$ ($n = 1, 2, \dots, 40$; $i = 1, 2, 3, 4, 5$);

这里 ξ 为一阈值参数,决定着预测结果与实际情

况的匹配程度,本文取 $\xi=0.70$ 。

以第 8 次试验中 19 号样方的预测结果为例加以说明,可以看到:预测结果向量 $[0.1163 \quad 0.0005 \quad 0.0066 \quad 0.6371 \quad 0.1292]$ 中 $f_{81}=0.13, f_{82}=0.00, f_{83}=0.01, f_{84}=0.72, f_{85}=0.14$, 由于 $f_{84} > \xi$ 且对应的 $S_{84}=1$, 因此预测结果是正确的。

利用上述方法对 10 次试验得到的 40 组预测健康等级进行判定,仅有 7 组与实际健康等级不符。因此,本模型的预测正确率为 $33/40$,即 82.5% ,这说明运用 BP 神经网络得到的预测结果是可靠的。

3 结论

人工神经网络应用于森林健康预测只需以森林健康评价指标为基础生成学习样本,通过训练就可以对森林健康的趋势进行预测。因此,只要训练好了一个稳定的网络模型,就可以广泛应用于森林健康预测中。

BP 算法具有自学习和自适应能力,通过比较网络的实际输出和期望输出的误差,反复调整网络的权值和阈值,逐步缩小误差从而达到理想的精度,因此利用训练好的网络模型来预测森林的健康趋势是合理的,且预测结果是可靠的。

为了构建一个准确稳定的森林健康预测模型,根据对神经网络输入层节点数、输出层节点数、隐含层节点数以及训练函数的构造和选取原则,给出一些建议。(1) 反映森林健康的各指标之间的关系比较复杂,且各指标对森林生态的影响具有不确定性,这就要求选择能够准确反映森林健康状况的评价指标来构建网络训练样本集;(2) 当输出层的节点数较少时,若预测结果较多,采用较少的输出层节点是不够的,此时神经网络的训练是不能收敛的,因此应根据预测样本集的大小规模设置输出层的节点个数(本研究采用 5 个输出层节点);(3) 网络中隐层神经元的数目对模拟结果也有一定的影响,因此应根据实际问题选用不同的经验函数或通过试验模拟来确定隐层神经元的最佳个数;(4) 根据问题的规模和实际情况

选择适当的训练方法,以平衡空间需求(内存空间)和时间需求(收敛速度)。

[参 考 文 献]

- [1] Macro Ferretti. Forest health assessment and monitoring-Issues for consideration[J]. Environmental Monitoring and Assessment, 1997, 48:45-72.
- [2] Samuel A, Craig J. Forest health monitoring in the United States: First four years[J]. Environmental Monitoring and Assessment, 1999, 55(2):267-277.
- [3] Angus J. Fungal damage recorded in eucalypt plantations during forest health surveys and their management [J]. Australasian Plant Pathology, 2007, 36(3): 225-239.
- [4] Sampson R N, David L A. Assessing Forest Ecosystem Health in the Inland West [M]. Binghamton: The Haworth Press, 1994.
- [5] 高均凯. 森林健康基本理论及评价方法研究[D]. 北京:北京林业大学,2007:18-21.
- [6] 甘敬,张振明,余新晓,等. 森林健康监测与评价研究[J]. 水土保持研究,2006,13(3):177-178.
- [7] 罗四维. 大规模神经网络理论基础[M]. 北京:清华大学出版社,2004:57-63.
- [8] 甘敬,朱建刚,张国祯,等. 基于 BP 神经网络确定森林健康快速评价指标[J]. 林业科学,2007,43(12):1-7.
- [9] 肖风劲,欧阳华,傅博杰,等. 森林生态系统健康评价指标及其在中国的应用[J]. 地理学报,2003,58(6):803-808.
- [10] 马立. 北京山地森林健康综合评价体系的构建与应用[D]. 北京:北京林业大学,2007:24-32.
- [11] 周曼,周明全. 基于 BP 神经网络的水稻害虫自动识别[J]. 北京师范大学学报:自然科学版,2008,44(20):165-167.
- [12] 高志亮. 北京森林健康示范区森林生态系统健康评价与预警研究[D]. 北京:北京林业大学,2009:45-50.
- [13] 李静锐,张振明. 森林生态系统健康评价指标体系的建立[J]. 水土保持研究,2007,14(3):173-175.
- [14] 张德丰. MATLAB 神经网络应用设计[M]. 北京:机械工业出版社,2009:214-243.