基于人工神经网络的农户经济收入预测研究

郝仕龙[1,2], 柯 俊[2], 李壁成[1], 赵小敏[2]
(1. 中国科学院水利部水土保持研究所, 陕西 杨陵 712100; 2. 江西农业大学国土与资源环境学院, 南昌 330045)

摘 要: 研究了利用误差反向传播人工神经网络(BP 网络) 的多变量综合预测问题, 并以研究上黄试区民户经济收入为背景, 建立了相应的多变量综合预测 BP 模型。预测结果表明: 农户经济收入神经网络模型预测精度较高, 开辟了农户经济收入预测的有效途径。

关键词: 人工神经网络; BP 算法; 农户经济收入预测
中图分类号: F323.15 文献标识码: A 文章编号: 1005-3409(2005)03-0117-03

Prediction Study of Farmer Income Based on the Artificial Neural Network

HAO Shi-long[1,2], KE Jun[2], LI bi-cheng[1], ZHAO Xiao-min[2]
(1. *Institute of Soil and Water Conservation, Chinese Academy of Science and Ministry of Water resources, Yangling, Shaanxi* 712100, *China*;
2. *College of Land Resource and Environment, Jiangxi Agricultural University, Nanchang* 330045, *China*)

Abstract: The multiple variable synthetical predication problems with error back propagation training artificial neural network (BP neural network) is researched. To study the farmer income of Shanghuang experimental area is the background, BP's models of multiple variable synthetical predication were established. The results of experimental prediction show that the accuracy of predication by the neural network models is very high. They open up a new way to predict farmer income.

Key words: artificial neural network; BP model; farmer income prediction

1 引 言

国家科技攻关固原上黄试验区地处全国有名的"西海固"老少边穷地区。海拔 1 534.3～1 822 m, 年均气温 6.9℃, 年降水量 420 mm。党中央、国务院对这一地区的经济发展和生态建设十分关心与重视, 曾决定从 1983 年起将"西海固"列为"三西"农业专项计划, 以 20 年时间集中解决这一片贫困问题。中国科学院根据中央领导指示和宁夏自治区要求, 派出一批科学家和科技人员深入宁南山区进行调研与考察, 为中央决策提供了科学依据。并在完成了"固原县农业综合考察与区划"的基础上, 于 1982 年在固原河川乡上黄村建立了科研基点, 进行长期定位试验研究和示范, 拉开了科技攻关的序幕。

为预测当前农户经济收入情况, 用传统的方法传统的回归分析技术对多元非线性关系的模拟效果不佳, 尤其是哪些作为重要影响因子应该予以保留, 哪些因子影响不大而应该舍弃, 一般只能采用逐步回归的方法, 但是在实际运用过程中存在很多多重共线、误差序列相关等问题, 预测的精度长期难以令人满意, 其原因是农户的经济状况受多因素影响, 各种因素之间关系复杂, 并呈现非线性, 预测非常困难。

人工神经网络(artificial neural network) 是基于连接演说构造的智能仿生模型, 它是由大量简单元件—神经元, 广泛相互连接而成的非线性、非局域性、非定常性和非凸性的复杂网络系统, 具有网络系统, 具有并行分布的信息处理结构和自适性的脑模式的信息处理的本质与能力, 它可以通过"自学习"或"训练"掌握大量的知识, 完成特定的工作。

近年来, 人工神经网络已得到了广泛的应用[1-4], 由于人工神经网络模型神经元之间的非线性, 使得模型具有较精度。神经网络的学习算法有多种, 根据所研究问题的性质和神经网络的有关理论, 本文采用 BP 神经网络的结构形式。分析了人工神经网络模型的方法及其过程, 并在此基础上对上黄试区 2003 年农户人均经济收入进行了预测分析。

2 BP 模型方法[5～8]

2.1 BP 模型简介

反向传播神经网络(又称 BP 模型) 是一种适用于非线性的模型识别和分类预测问题的人工神经网络, 它的结构如图 1 所示, 最基本的 BP 网络是由输入层、隐层和输出层组成 3 层前馈网络, 每层有若干个互不连接的神经元节点, 相邻两层节点通过权值连接。

[1] 收稿日期: 2004-12-25
基金项目: 国家"十五"重大科技攻关 (2001BA 606A - 04) 资助项目
作者简介: 郝仕龙(1972-), 男, 江西南昌人, 在读博士, 主要研究研究方向为土地利用/覆被变化等。

2.2 BP 神经网络的学习算法

设输入层为 M,即有 M 个输入信号,其中的任一输入信号用 m 表示;第 1 隐层为 I,即有 I 个神经元,其中的任一神经元 i 表示;第 2 隐层为 J,即有 J 个神经元,其中任一神经元用 j 表示,输出层为 P,即有 P 个输出神经元,其中任一神经元用 p 表示。

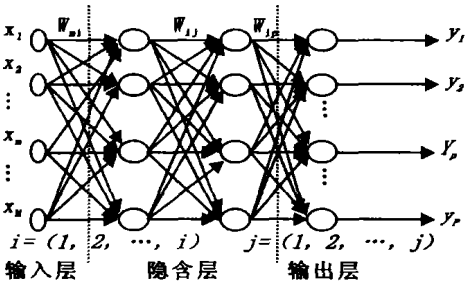

图 1 BP 神经网络模型

输入层与第 1 隐层的突触权值用 w_{mi} 表示,第 1 隐层与第 2 隐层的突触权值用 w_{ij} 表示;第 2 隐层与输出层的突触权值用 w_{ip} 表示。

神经元的输入用 u 表示,激励输出用 v 表示,u,v 的上标表示层,下标表示层中的某个神经元,如 u_i^I 表示 I 层(即第 1 隐层)的第 i 个神经的输入。设所有的神经元的激励函数均用 *Sigmoid* 函数。设训练样本集为 $X=[X_1, X_2, \cdots, X_k, \cdots, X_N]$,对应任一训练样本 $X_k=[x_{k1}, x_{k2}, \cdots, x_{km}]$,$(k=1, 2, \cdots, N)$ 的实际输出为 $Y_k=[y_{k1}, y_{k2}, \cdots, y_{kp}]$,期望输出为 $d_k=[d_{k1}, d_{k2}, \cdots, d_{kp}]$,设 n 为迭代次数,权值和实际输出是的 n 函数。

网络输入训练样本 X_k,由工作信号的正向传播过程可得

$$u_i^I=\sum_{m=1}^{M} w_{mi}x_{km} \qquad v_i^I=f\left[\sum_{m=1}^{I} w_{mi}V_{km}\right] \qquad i=1, 2, \cdots, I,$$

$$v_j^J=f\left[\sum_{i=1}^{I} w_{ij}V_i^I\right] \qquad j=1, 2, \cdots, J,$$

$$u_p^p=\sum_{j=1}^{J} w_{jp}v_j^J \qquad v_p^p=f\left[\sum_{j=1}^{J} w_{ip}v_j^J\right] \qquad p=1, 2, \cdots, P,$$

$y_{kp}=v_p^p=f(u_p^p)=f\sum_{j}^{J} w_{jp}v_j^J$ 输出层第 p 个神经元的误差信号为 $e_{kp}(n)=d_{kp}(n)-y_{kp}(n)$,定义神经元 p 的误差能量为 $\frac{1}{2}e_{kp}^2(n)$,则输出层所有神经元的误差能量总和为 $E(n)$:

$$E(N)=\frac{1}{2}\sum_{p=1}^{p} e_{kp}^2(n)$$

误差信号从后向前传递,在反向传播的过程中,逐层修改联接权值。

BP 网络的具体学习算法步骤如下:

(1) 设置变量和参量:

$X_k=[x_{k1}, x_{k2}, \cdots, x_{km}]$,$(k=1, 2, \cdots, N)$ 为输入向量,或称训练样本,N 为训练样本的个数。

$$W_{MI}(n)=\begin{bmatrix} W_{11}(N) & w_{12}(n) & \cdots & w_{1I}(n) \\ W_{21}(N) & w_{22}(n) & \cdots & w_{2I}(n) \\ & & & \\ W_{IJ}(N) & w_{I2}(n) & \cdots & w_{MI}(n) \end{bmatrix}$$

为第 n 次迭代时输入层与隐层 I 之间的权值向量。

$$W_{IJ}(n)=\begin{bmatrix} W_{11}(N) & w_{12}(n) & \cdots & w_{1J}(n) \\ W_{21}(N) & w_{22}(n) & \cdots & w_{2J}(n) \\ & & & \\ W_{I1}(N) & w_{I2}(n) & \cdots & w_{IJ}(n) \end{bmatrix}$$

为第 n 次迭代时隐层 I 与隐层 J 之间的权值向量。

$$W_{JP}(n)=\begin{bmatrix} W_{11}(N) & w_{12}(n) & \cdots & w_{1P}(n) \\ W_{21}(N) & w_{22}(n) & \cdots & w_{2P}(n) \\ & & & \\ W_{J1}(N) & w_{J2}(n) & \cdots & w_{JP}(n) \end{bmatrix}$$

为第 n 次迭代时隐层 J 与输出层之间的权值向量。

$Y_k(n)=[y_{k1}(n), y_{k2}(n), \cdots, y_{kp}(n)]$,$(k=1, 2, \cdots, N)$ 为第 n 次迭代时网络的实际输出。

$d_k=[d_{k1}, d_{k2}, \cdots, d_{kp}]$,$(k=1, 2, \cdots, N)$ 为期望输出。η 为学习速率;m 为迭代次数。

(2) 初始化,赋给 $W_{MI}(0)$,$W_{IJ}(0)$,$W_{JP}(0)$ 各一个较小的随机非零值。

(3) 随机输入样本 x_k,$n=0$。

(4) 对输入样本 x_k,前向计算 BP 网络每层神经元的输入信号 u 和输出信号 v。其中

$$v_p^p(n)=y_{kp}(n), p=1, 2, \cdots, p$$

(5) 由期望输出 d_k 和上一步求得的实际输出 $Y_k(n)$ 计算误差 $E(n)$,判断其是否满足要求,若满足转至第八步,不满足转至第六步。

(6) 判断 $n+1$ 是否大于最大迭代次数,若大于转至第八步,若不大于,对输入样本 x_k,反向计算每层神经元的局部梯度 δ。其中:

$$\delta_p^p=y_p(n)[1-y_p(n)][d_p(n)-y_p(n)], p=1, 2, \cdots, p$$

$$\delta_j^J(n)=f[u_j^J(n)]\sum_{p=1}^{p}\delta_p^p(n)w_{jp}(n), j=1, 2, \cdots, J$$

$$\delta_i^I(n)=f[u_i^I(n)]\sum_{j=1}^{J}\delta_j^J(n)w_{ij}(n), i=1, 2, \cdots, I$$

(7) 按下式计算权值修正量 Δw,并修正权值:$n=n+1$,转至第四步。

$$\Delta w_{jp}(n)=\eta\delta_p^p(n)v_j^J(n), W_{jp}(n+1)=w_{jp}(n)+\Delta_{jp}(n) \quad j=1, 2, \cdots, J; p=1, 2, \cdots, P$$

$$\Delta w_{ij}(n)=\eta\delta_j^J(n)v_j^I(n), W_{ij}(n+1)=w_{ij}(n)+\Delta_{ij}(n) \quad i=1, 2, \cdots, I; j=1, 2, \cdots, J$$

$$\Delta w_{mi}(n)=\eta\delta_i^I(n)x_{km}(n), W_{mi}(n+1)=w_{mi}(n)+\Delta_{mi}(n) \quad m=1, 2, \cdots, M; i=1, 2, \cdots, I$$

(8) 判断是否所有的训练样本,是则结束,否则转至第三步。

3 BP 神经网络在农户经济收入预测中的应用

3.1 预测指标体系

由于影响农户经济收入的变量较多,我们对各指标值进行了筛选并根据上黄试区实际情况,本文选择影响农户经济收入的土地面积、劳动力文化素质、农业投资、管理水平 4 个变量作为预测农户经济收入的变量指标(即人工神经网络的输入),这 4 个变量与该试区农户经济状况密切相关。

3.2 变量的量化

(1) 土地面积:以农户 2003 年实际调查的土地面积作为样本的输入数据。

(2) 文化素质:以农户受教育的年限来替代,如高中是毕业受 11 年教育,初中毕业为 8 年,小学毕业为 5 年,文盲为 0。并以相应的年限作为输入样本的数据。

(3) 农业投资: 以农户当年农业投资量的多少作为样本的输入数据。

(4) 管理水平: 以单位土地面积产量作为样本数据输入。

3.3　预测指标体系的神经网络结构

首选确定各神经网络单元数, 根据前面所列出影响农户经济收入的变量, 各神经网络的具体参数如表 1 所示。其中隐含层神经单元数可自行设定, 一般来说, 问题越复杂需要的隐含层单元数越多, 但过多的隐含层单元会增加计算量, 目前其还没有有效的方法, 需要根据网络大小来确定。本文根据如下规划确定隐含层的神经元数: 隐含层的神经元数目大于输入层神经元和输出层神经元数目的一半, 小于输入层神经元和输出层神经元数目之和。

表 1　神经网络预测参数表

模型参数	输入层神经单元数	输出层神经单元数	隐含层个数	隐含层神经单元数
农户经济评价神经网络	4	1	1	4

3.4　网络训练

(1) 数据来源。由于评价单元是针对试区农户, 根据上黄试区 2003 年农户经济调查资料, 选取经济收入高、中及低的农户 10 户进行网络训练, 5 户作为网络预测样本。

(2) 数据处理。考虑到原始数据量纲不同和指标数值存在数量级的明显差异, 因此需对原始数据进行归一化处理。这里采用下述方法来进行, 即 X_i 用 $\frac{(X_i - X)}{X}$ 其中 X 为 X_i 的平均数, X_i 表示某一指标的实际值。该法的优点是可以明显地消除原始数据的级差, 并统一量纲。

表 2　输入层与隐含层之间的权值

W	1	2	3	4
1	- 2.58791	1.35277	2.44099	8.40012
2	1.19191	0.88552	- 0.5481	- 0.86322
3	1.22434	2.83605	1.66209	- 1.93305
4	- 4.40356	- 2.85991	0.36649	1.56248

4　模拟结果与讨论

进入 BP 神经网络训练时, 设置各参数如下: 允许误差 0.000 1、输入层节点数 4、隐含层 1 层、最小训练速率 0.1、动态参数 0.6, Sigmoid 参数为 0.9, 最大迭代次数 1 000, 并对各输入节点的数值进行标准化转换后, 启动神经网络进行学习, 经过 584 次循环学习后网络输出结果如表 4 所示, 从表 4 中可以看出, 预测相对误差控制在 5% 左右, 说明该系统的农户经济收入预测结果是比较准确的。

表 3　隐含层与输出层之间的权值

W	1	2	3	4
1	- 10.0686	2.02393	5.07895	- 6.33142

表 4　BP 模型预测结果误差分析

农户	原始值	BP 模拟值	BP 预测值	相对误差/%
1	1295	1294.058		0.07
2	870	869.809		0.02
3	1155	1145.702		0.81
4	3426	3433.702		0.22
5	680	677.191		0.41
6	993	994.631		0.16
7	620	625.593		0.90
8	457	454.595		0.53
9	446	437.065		2.00
10	127	123.586		2.69
11	1044		1034.893	0.87
12	203		192.781	5.03
13	220		210.485	4.32
14	125		127.235	1.79
15	2215		2236.251	0.96

5　结　论

(1) 农户的经济状况受多因子影响, 预测非常困难, 传统的回归分析技术对多元非线性关系的模拟效果不佳, 原因是影响因子大多、太复杂, 关系呈现非线性, 而神经网络擅长于对大量复杂及非线性的数据进行处理, 因此应用神经网络对农户经济收入进行预测可以充分发挥神经网络的优越性。

(2) 人工神经网络预测的准确性样本的多少、指标体系的完善性及数据的可靠性等影响, 本文选取与农户经济收入直接相关的指标, 应用 2003 年农户实地调查数据, 保证了数据的可靠性及预测的准确性, 并消除了年际市场变化对农户经济收入的影响, 使预测结果更加结合实际。

(3) 从 BP 神经网络的模拟及预测结果来看, 预测结果相对误差都控制在 5% 左右, 大部分预测相对误差 5% 以下, 说明该系统的农户经济收入预测结果是比较准确的。

(4) 由于人工神经网络不需给出各影响变量的直接权重, 而是通过网络学习自动映射出农户经济收入与各影响因素之间非线性关系, 因此避免了 Delphi 法确定的人为性和一次性。同时为保证 BP 神经网络预测的准确性, 变量的选取必须与预测目标有直接的联系。

参考文献:

[1]　孙会君, 王新华. 应用人工神经网络确定评价指标的权重[J] . 山东科技大学学报, 2001, 20(3) : 84- 86.
[2]　杨建刚, 等. 利用结构化神经网络识别振动系统非线性特性[J] . 振动工程学报, 1995, 8(3) : 25- 29.
[3]　崔胜民. 神经网络理论在轮胎力学听应用[J] . 农业机械学报, 1995, 26(3) : 147- 148.
[4]　戴文战. 基于三层 BP 网络的多指标综合评估方法及应用[J] . 系统工程理论与实践, 1999, 19(5) : 30- 33.
[5]　潘大丰, 何书金, 郭焕成. 矿区废弃土地资源适宜性评价[J] . 地理科学进展, 1998, 17(4) : 40- 45.
[6]　楼顺天. 基于 MATLAB 的系统分析与设计- 神经网络[M] . 西安: 西安电子科技大学出版社, 1998.
[7]　高隽. 人工神经网络原理及住址实例[M] . 北京: 机械工业出版社, 2003.
[8]　施鸿宝. 神经网络及其应用[M] . 西安: 西安交通大学出版社, 1999.