

翟慧,李纪云. BP\_Adaboost 预测模型结合遗传算法在鸡腿菇生长环境寻优中的应用[J]. 江苏农业科学,2013,41(7):383-385.

# BP\_Adaboost 预测模型结合遗传算法 在鸡腿菇生长环境寻优中的应用

翟慧,李纪云

(河南职业技术学院信息工程系,河南郑州 450046)

**摘要:**针对试验中使指标达到最佳值的最优因素组合的问题,提出了一种基于 BP\_Adaboost 的预测模型结合遗传算法的求解方法,并在鸡腿菇生长环境寻优中进行实际应用。结果表明,与 BP 神经网络相比,该模型提高了预测精度和泛化能力,为解决此类寻优问题提供了可靠的参考,具有较好的应用前景。

**关键词:**BP 神经网络;AdaBoost 算法;BP\_Adaboost 预测模型;遗传算法

**中图分类号:** S126 **文献标志码:** A **文章编号:** 1002-1302(2013)07-0383-02

试验研究员一般都希望通过回归方程进行最优控制或寻找最佳试验条件,例如,冶金方面的研究员关心合金强度的最优配方,农业方面的研究员关心使产量达到最优的生产条件等。在试验过程里,不仅须制定优化指标,而且还须设法找到能够让指标获得最优值的优化因素组合,这个问题属于非常典型的非线性优化。目前通用的方法是运用正交设计法来安排试验,用试验数据构建指标和因素之间的回归方程,进而运用某种寻优方法求出最优的因素组合。该方法的缺点是数学模型粗糙,很难表述优化指标与各因素间的非线性关系,结果误差较大。此时,可以借助计算机技术,在已知试验数据的基础上,利用神经网络遗传算法寻找最优的试验条件<sup>[1]</sup>。具体的方法是根据试验条件和试验结果确定 BP 神经网络结构,以试验条件为输入数据、试验结果为输出数据训练网络来预测一定试验条件下的试验结果,然后把试验条件作为遗传算法中的种群个体,把网络预测的试验结果作为个体的适应度,通过遗传算法推导出最优试验结果及与其对应的最优试验条件。试验建模证明神经网络能够逼近任意非线性函数,为研究最广泛的一类建模方法。然而,神经网络的建模一直处于理论仿真阶段,生产实践应用受到限制,主要是因为一直没有能够解决它的 3 个固有缺陷:(1)网络结构的建立没有一定的依据,隐含层的数目仅主观人为地确定;(2)神经网络的学习会陷入局部极小点,导致不能获得问题的最优解;(3)神经网络会出现训练误差过小的情况,进而导致推广能力下降。针对上述问题,本研究提出了一种基于 AdaBoost(adaptive boosting)算法的神经网络预测模型,能较好地解决这 3 个问题。

## 1 AdaBoost 算法原理

AdaBoost 算法是提高预测学习系统预测能力的有效工

具<sup>[2-3]</sup>,算法流程见图 1。该算法的根本原理是:先给出每个弱学习算法与样本空间,然后从样本空间里寻找出  $m$  组训练数据,每组训练数据的权重相同;随后采用弱学习算法进行  $n$  次迭代运算,每次运算后都按照预测结果更新训练数据权重分布,预测失败的训练个体被给予比较大的权重,这些训练个体在下次迭代运算时倍受关注。弱预测器经过多次反复迭代获得了一个预测函数序列,对每个预测函数都赋予 1 个权重,预测的结果越好,对应的权重就越大,经过  $n$  次迭代以后,最终由弱预测函数加权获得强预测函数。AdaBoost 采用单预测器加权投票来构建最后的强预测器,因此泛化能力等到了很好的提升。

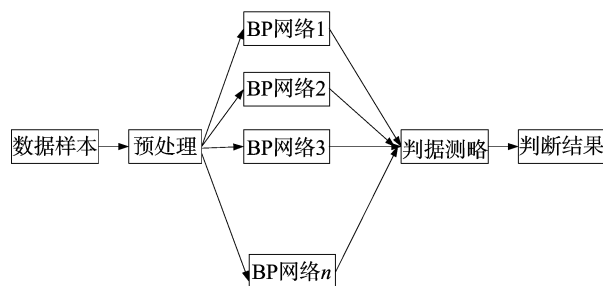


图1 算法流程

## 2 BP\_Adaboost 预测模型

BP\_Adaboost 预测模型是运用 BP 神经网络作为弱预测器,不间断地反复训练 BP 神经网络预测样本输出,采用 Adaboost 算法获得多个 BP 神经网络弱预测器构成的强预测器<sup>[4-5]</sup>。

### 2.1 网络与数据初始化

在样本空间里随机选出  $m$  组训练数据,将测试数据的分布权值初始化  $D_1(i) = 1/m$ ,用样本输入输出的维数来确定神经网络结构,对 BP 神经网络权值与阈值进行初始化。

### 2.2 弱预测器预测

在训练第  $t$  个弱预测器之时,运用训练数据来训练 BP 神经网络并预测训练数据的输出,获得预测序列  $g(t)$  的预测误差  $e_t$ ,  $e_t$  的计算公式是:

收稿日期:2013-01-07

基金项目:河南省高等学校青年骨干教师资助计划(编号:2012GGJS-241);河南省教育科学规划项目(编号:[2011]-JKGHAD-0441)。

作者简介:翟慧(1980—),女,河南周口人,硕士,讲师,主要从事电子技术、计算机应用研究。E-mail:wuna198001@163.com。

$$e_i = \sum_i D_i(i), i = 1, 2, m [g(t) \neq y] \tag{1}$$

式中： $i$  表示第  $i$  组训练数据； $g(t)$  表示弱预测器的估计估计结果； $y$  表示预期估计结果。

2.3 弱预测器权重的计算

依据预测序列  $g(t)$  与预测误差  $e_i$  可以出计算序列权重  $a_i$ ，计算公式是：

$$a_i = \frac{1}{2} \ln(\frac{1 - e_i}{e_i}) \tag{2}$$

2.4 调整测试数据

依据预测序列权重  $a_i$  对下一轮训练样本权重进行调整，调整公式为：

$$D_{i+1}(i) = \frac{D_i(i)}{B_i} \exp[-a_i y_i g_i(x_i)], i = 1, 2, \dots, m \tag{3}$$

式中： $y_i$  表示第  $i$  组训练数据的预期估计结果； $g_i(x_i)$  表示弱预测器的第  $i$  组训练数据估计估计结果； $B_i$  为归一化因子，目的是在权重比例不改变的情况下使分布权值和为 1。

2.5 强预测器函数

训练  $T$  轮以后获得  $T$  个弱预测器函数  $f(g_i, a_i)$ ，组建后得到了强预测器函数  $h(x)$ ：

$$h(x) = \sum_{i=1}^T a_i \cdot f(g_i, a_i) \tag{4}$$

3 BP\_Adaboost 预测模型结合遗传算法在鸡腿菇生长最优环境因素中的应用

鸡腿菇是食用菌的一种，味道鲜美，营养价值高，具有清神益智、益脾胃、助消化、增加食欲等功效，被广大消费者所喜爱；含有 20 种氨基酸，其中 8 种是人体所必需的，还含有抗癌活性物质和治疗糖尿病的有效成分，长期食用，对降低血糖浓度、治疗糖尿病有较好疗效，深受市场欢迎。为了提高鸡腿菇的产量和质量，降低生产成本，达到效益最大化，必须根据鸡腿菇的最佳生长环境，对其生长过程各阶段的环境参数进行调节。鸡腿菇对光照强度要求不高，200 ~ 2 000 lx 之间都可以正常出菇，而温度和湿度这 2 个环境条件对鸡腿菇生长有重大影响。只通过试验得到鸡腿菇的最佳生长温度和湿度，要花费大量的时间和经费，还很有可能找不到最优的结果。根据鸡腿菇栽培经验，在 5 个温度水平（16、18、20、22、24 ℃）和 5 个湿度水平（70%、75%、80%、85%、90%）条件下进行鸡腿菇栽培试验，选取 100 组试验数据，随机取 65 组数据为神经网络的训练数据，剩余 35 组试验数据作为测试数据（表 1）。

鸡腿菇生长试验因素有温度和湿度 2 个，试验结果为鸡腿菇重量 1 个，因此构成 BP 神经网络的结构为 2—5—1，共训练生成 6 个 BP 神经网络弱预测器，最后用 6 个弱预测器组成强预测器来预测鸡腿菇的生长模型。如图 2 所示，BP\_Adaboost 预测模型的误差远小于 BP 神经网络预测模型的误差，提高了预测精度和泛化能力，更能满足具有非线性、时变性和不确定性的非数学模型预测的需要。

4 遗传算法寻优结果分析

遗传算法中把 BP\_Adaboost 预测模型的输出的倒数（鸡腿菇重量的倒数）作为个体适应度，适应度越小，个体越优。

表 1 鸡腿菇栽培试验测试数据

序号	温度 (℃)	湿度 (%)	测量值	序号	温度 (℃)	湿度 (%)	测量值
1	16	70	20.65	19	22	85	44.27
2	16	75	30.76	20	22	90	46.17
3	16	80	40.32	21	24	70	36.37
4	16	85	42.84	22	24	75	38.85
5	16	90	42.44	23	24	80	40.03
6	18	70	28.46	24	24	85	41.48
7	18	75	34.24	25	24	90	42.43
8	18	80	41.22	26	26	70	30.43
9	18	85	44.62	27	26	75	32.51
10	18	90	44.84	28	26	80	34.65
11	20	70	31.63	29	26	85	35.95
12	20	75	37.65	30	26	90	40.43
13	20	80	43.67	31	28	70	28.47
14	20	85	46.54	32	28	75	29.54
15	20	90	47.36	33	28	80	30.11
16	22	70	34.14	34	28	85	30.46
17	22	75	40.71	35	28	90	31.27
18	22	80	43.68				

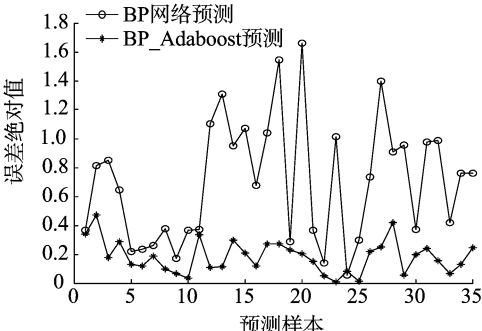


图2 2种预测模型的误差绝对值

由于试验中只有温度和湿度 2 个输入参数，因此个体长度为 2，选择迭代次数 50 次，种群规模是 20，交叉概率 0.4，变异概率 0.2，采用浮点数编码<sup>[6]</sup>。优化过程中最优个体适应度变化曲线如图 3 所示。

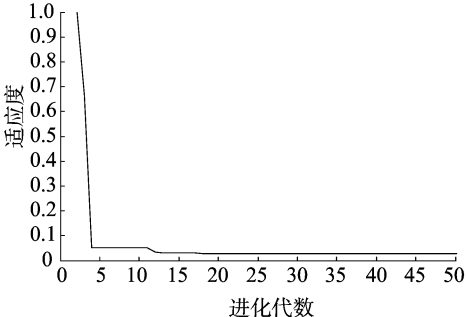


图3 最优个体适应度变化曲线

遗传算法优化的结果为 48.57，对应的试验因素为温度 20.154 ℃、湿度 90%，对比表 1，该结果可以为鸡腿菇最佳生长环境选择提供参考。

5 结束语

BP\_Adaboost 预测模型将 BP 神经网络作为弱预测器，经

谷延霞,王 恒,谷艳云. 基于 MapX 灌区信息管理系统的开发研究[J]. 江苏农业科学,2013,41(7):385-387.

# 基于 MapX 灌区信息管理系统的开发研究

谷延霞<sup>1</sup>, 王 恒<sup>2</sup>, 谷艳云<sup>3</sup>

(1. 河北农业大学理学院物理系,河北保定 071001; 2. 河北农业大学城乡建设学院,河北保定 071001;

3. 河北省临西县河西镇政府,河北临西 054900)

**摘要:**研究了基于 MapX 的灌区管理信息系统的总体设计及采取的开发路线,并对系统的主要功能数据的统计、空间分析、灌溉用水及配水等作了阐述。利用该系统可以及时准确地掌握灌区的水源情况、灌溉情况及其他信息,有利于我国北方缺水地区农业精细灌溉的管理实现智能化。

**关键词:**MapX;地理信息系统(GIS);Visual C++ .NET;数据库;灌溉管理

**中图分类号:**TP311.52 **文献标志码:**A **文章编号:**1002-1302(2013)07-0385-03

中国是世界上 13 个严重缺水的国家之一,水资源短缺在我国北方尤为突出,但在灌溉水资源的利用上传统的灌溉方式存在着严重的浪费现象。据估计,每年农业浪费的水量达 1 000 亿 m<sup>3</sup>。因此,推广数字农业技术已成为中国农业发展的一种趋势。精细灌溉管理技术是根据农作物的需水、耗水规律,来控制、调配水源,以最大限度地满足作物对水分的需要,实现区域最佳的农田水分调控管理。根据作物的实际需求,及时、适量地将水分供给作物,最终达到高产、高效,提高水分利用效率的目的。

## 1 系统采用技术

### 1.1 MapX 概述

MapX 是一个基于 ActiveX(OCX)技术的可编程控件,它不但使用了与 MapInfo Professional 一致的地图数据格式,而且还在这个小的控件中实现了 MapInfo Professional 的大多数功能,例如:地图编辑和空间分析功能。MapX 的内部提供了 40 多个父类为 Microsoft 基础类(MFC)的 OLE 对象,这就使得应用程序员可以使用任何支持 ActiveX 的程序开发环境来采用嵌入 MapX 控件方式,方便快捷地开发 GIS 应用程序。

### 1.2 系统开发平台

系统的运行平台为 Windows 2000 系列操作系统,系统数据库采用 Microsoft Access, GIS 模块采用 MapInfo,编程环境为

面向对象的 VC++ .NET。系统逻辑体系结构如图 1 所示。

在 VC++ .NET 中使用 MapX 的方法不同于 Visual Basic 和 C++ Builder。在 VC++ .NET 中,每个 MapX 对象是一个 C++ 类,这些类在 MapX.h 文件中声明,在 MapX.cpp 文件中实现。这些 C++ 类的名称与 MapX 对象的名称相同,在对象名称前面加上 CMapX 就构成了类名。MapX 对象的属性是通过 C++ 类的成员函数来实现的<sup>[1-2]</sup>。

## 2 系统开发整体过程

利用 MapInfo 公司的 MapX 控件开发 GIS,其具体开发过程如下:首先,安装 MapX 控件。然后,安装 MapInfo 软件。对于图形简单的系统,可以直接利用 MapX 绘图,但功能有限,而且修改困难。在开发灌区 GIS 时,因为系统图形元素较多,还有许多符号需要自定义,所以先用 MapInfo 建立图形和空间数据库,再利用 MapX 提供的 Geoset Manager 形成 MapX 可以调用的 .gst 文件,这样既解决了绘图难的问题,同时方便了 MapX 对图形的管理。最后,应用程序的开发。建立属性数据库,这可以用一般的关系数据库系统来建立,并通过所用开发平台建立对数据库中属性数据的维护工作。以河北省武强胜利灌区为试点,实现了一个小型地理信息系统与移动处理平台,后台数据库采用 SQL Server 2000<sup>[3]</sup>。

## 3 系统数据组织及数据库的建立

### 3.1 系统数据组织

灌区管理信息系统的数据涉及空间矢量数据、属性数据和管理数据。空间数据主要指能标记水利设施的空间位置的

收稿日期:2013-06-06

基金项目:河北省科技计划(编号:12227406)。

作者简介:谷延霞(1982—),女,河北保定人,硕士,讲师,研究方向为农业机械化。Email:guyanxia1209@163.com。

Adaboost 算法得到强预测器,较好地解决了神经网络所面临的 3 个主要问题,同时提高了预测精度和泛化能力,结合遗传算法为解决该类寻优问题提供了可靠的参考,具有一定的应用价值。

## 参考文献:

- [1]王学会. 遗传算法和 BP 网络在发酵模型中的应用[D]. 天津:天津大学,2007.
- [2]孙凤琪. AdaBoost 集成神经网络在冲击地压预报中的应用[J].

吉林大学学报:信息科学版,2009,27(1):79-84.

- [3]毛志忠,田惠欣,王 琰. 基于 AdaBoost 混合模型的 LF 炉钢水终点温度软测量[J]. 仪器仪表学报,2008,29(3):662-667.
- [4]艾小松,黄肇雄,张良春,等. 基于 AdaBoost 算法的高速公路事件检测[J]. 计算机工程与科学,2007,29(12):95-97.
- [5]杨 涛,张良春. 基于 AdaBoost 集成 RBF 神经网络的高速公路事件检测[J]. 计算机工程与应用,2008,44(32):223-225,229.
- [6]吴建生. 基于遗传算法的 BP 神经网络气象预报建模[D]. 南宁:广西师范大学,2004.