

改进遗传神经网络在甘蔗产量预测中的应用

徐永春^{1,2}, 张森文³

(1 华南农业大学 工程学院, 广东 广州 510642; 2 广东理工职业学院 计算机系,
广东 广州 510091; 3 暨南大学 应用力学研究所, 广东 广州 510632)

摘要:传统甘蔗产量预测方法对外界多因素关联作用的农作物产量预测的难度大、精度差、准确度低,本文提出改进自适应交叉和变异算子的遗传BP算法,多元逐步回归简化BP网络的输入变量,应用改进的遗传BP算法策略,并以甘蔗产量实例数据进行了验证和分析,结果表明,改进的遗传BP算法总体效果最优。

关键词:改进遗传算法; BP网络; 甘蔗; 预测

中图分类号: TP306.1

文献标识码: A

文章编号: 1001-411X(2010)03-0102-03

The Application on the Sugar Cane Forecast with the Neural Networks Using Improved Genetic Algorithms

XU Yong-chun^{1,2}, ZHANG Sen-wen³

(1 College of Engineering, South China Agricultural University, Guangzhou 510642, China;

2 Department of Computer, Guangdong Polytechnic Institute, Guangzhou 510091, China;

3 Applied Mechanical Institute, Jinan University, Guangzhou 510632, China)

Abstract: The features of traditional method on sugar cane production forecast were discussed, and an improved intelligence genetic algorithms (GA) strategy and a method of simplified input variables by the multiple stepwise regression are presented, which can avoid the shortage of trained BP and enhance the BP model learning efficiency on the sugarcane data. An improved GABP model was simulated and verified in the forecast example of sugarcane data.

Key words: improved genetic algorithm; BP network; sugar cane; forecast

甘蔗产量的预测是制糖工业的一项重要技术管理工作,对优化提高甘蔗生产作业及统筹管理具有指导作用。在甘蔗产量预测中,农作物产量、气象等是重要影响因子,应用传统数理统计方法对农作物产量、气象等的预测预报的研究已较为成熟,但该法对受外界多因素关联作用的农作物产量预测存在预测难度大、精度差、准确度低等问题,在实际应用中难以符合需求。近年来,神经网络在人工智能领域的应用研究十分活跃,但仍存在训练时间长、收敛于局部极小值等不足,使得传统的神经网络算法在网络学习训练中所获得的网络权、阈值在训练效率和精度上都有所下降^[1-2],而与之对比的人工智能算

法——遗传算法则具有简单通用、鲁棒性强和并行运算等优点,二者的结合将强化BP学习训练、优化BP神经网络权阈值,促进BP神经网络的快速收敛,提高BP网络神经预测的效率与精度。但传统遗传算法存在一些具有超高适应度的个体,它们在选择算子时,会以很大的概率参加繁殖,引起群体平均适应度饱和、个体间竞争力减弱,导致运算收敛速度下降,甚至是未成熟收敛,出现“早熟”现象^[3-4]。为有效发挥遗传算法的特色优势,在应用中需对算子选择方法、判别做适度改进。

甘蔗生产是市场调节、种植规模、气候调节多因素作用关联的大系统。在甘蔗产量预测中,简化农作

物产量、气象等输入因子对于构建规模适度的 BP 网络模型,提高 BP 网络的训练效率和精度有直接的作用. 本文通过多元逐步回归简化输入参量,有效结合遗传算法与神经网络算法,构建改进的遗传神经网络甘蔗产量预测模型,并以实例验证.

1 改进遗传神经网络设计

1.1 遗传算法改进设计方案

神经网络的优化设计问题属于高维连续的寻优问题. 根据 BP 网络输入节点设计的需要,本文采用搜索精度高、搜索空间大的实数编码方案,可以较好地解决因空间不足而导致算法收敛缓慢问题.

在遗传算子的改进方案中,遗传参数中的交叉概率 p_c 和变异概率 p_m 是决定遗传算法(GA)性能的关键,无论过大还是过小,都会直接影响算法的收敛性和收敛速度,但很难找到一个适应每个问题的最佳值,对于不同的优化问题需要反复试验来确定 p_c 和 p_m . Scinivas 等^[5] 和 Fogarty^[6] 提出了种群随迭代次数增加而收敛调整 p_c 和 p_m 的概率值,控制 p_c 和 p_m 最小范围,但算法的自适应能力较差. 本文在文献[7]基础上,改进自适应交叉和变异算子,设计自适应的交叉和变异公式(1)和(2),在预测仿真前实例验证,证明变异概率随指数下降有很好的性能,其自适应的交叉和变异公式为:

$$p_c = \begin{cases} \frac{p_{c1}}{1 + \exp(10\lambda(t-1)/2(t_{\max}-2)) + k} & 0.5 < p_c < 1, \\ 0.5 & p_c < 0.5. \end{cases} \quad (1)$$

$$p_m = \frac{p_{m1}}{1 + \exp(-2t/t_{\max}) - 0.45}, \quad (2)$$

式中, t 为遗传迭代次数; $t_{\max}$ 为最大遗传代数; λ 、 k 为常数,根据种群分布大小取值,一般为 $[5, 10]$; p_{c1} 、 p_{m1} 分别为固定交叉概率和变异概率. 通过式(1)和式(2)可控制 p_c 取值在 $[0.5, 1]$ 之间,而 p_m 取值在 $[0, 0.5]$ 之间. 通过自适应交叉和变异处理的遗传算法,出现适应值高于群体平均值的个体有更大的概率进入下一代,而适应值低于群体平均值的个体则将被淘汰,从而可大大加快遗传算法收敛.

1.2 基于改进遗传算法优化的 BP 网络设计

利用改进遗传算法与 BP 网络形成混合训练算法,使每个个体包含若干条染色体,每条染色体对应 1 个权值或阈值,染色体采用实数编码,通过上述改进策略,则改进的遗传 BP 网络模型设计框架主要是以权阈值训练为主. 首先由改进 GA 进行全局搜索,

将搜索范围缩小后,交与 BP 网络执行局部寻优,发挥其局部寻优的高精确性,具体步骤如下:

- 1) 初始化网络,初始化群体;
- 2) 计算神经网络误差函数个体适应度,若达到预定精度要求则转 5);
- 3) 遗传操作(选择、交叉、变异),产生新一代个体;
- 4) 改进自适应交叉和变异算子更新;
- 5) 解码,获得优化的 BP 网络权阈值;
- 6) 再以优化的权、阈值进行 BP 模型的数据仿真预测.

2 甘蔗生产预测仿真

在甘蔗产量预测模型研究中,外界因子如气象、品种资源分布率、化肥、人工技术等均对甘蔗产量预测有显著的关联作用. 因此本文构建的改进遗传算法 BP 模型选取对产量有显著影响的气象、优化品种种植分布率作为输入外界因子. 因气象、品种资源等外界因子在预测过程中,不能够独立提供预测模型参数,故气象、品种资源需要进行自我学习、训练完成下一步的预测工作^[8-9].

在模型仿真训练中,本文采用 1995—2004 年 10 年间的气象、品种资源分布率等作为外界因子.

气象因子主要包括温度、湿度、降雨量和日照,而在每年甘蔗生产中,只有若干季度是甘蔗最主要的外界生产影响条件,每年 12 个月中有 48 个外界气象因子、若干品种资源分布数据及年甘蔗产量,通过组合改进遗传算法 BP 模型输入因子达到 58 个,使预报因子数目过大,导致计算复杂,降低了模型的预报精度,因此根据 BP 模型训练需要,通过多元逐步回归简化而得到甘蔗生产最终输入的关键输入向量,其数学方程如下:

$$Y = -233\,091.152\,1 - 43.206\,3x_3 + 44.555\,2x_4 + 260.737\,2x_{11} + 643.584\,4x_{19} + 2\,893.212\,4x_{27} + 2\,532.370\,4x_{30} - 2\,428.027\,4x_{34} + 375.794x_{46} - 351.145\,6x_{48} + 236.4x_{53}. \quad (3)$$

在式(3)的输入向量中, $x_1 \sim x_{12}$ 分别为 1—12 月的气温输入向量因子, $x_{13} \sim x_{24}$ 分别为 1—12 月的日照输入向量因子, $x_{25} \sim x_{36}$ 分别为 1—12 月的降雨输入向量因子, $x_{37} \sim x_{48}$ 分别为 1—12 月的湿度输入向量因子, $x_{49} \sim x_{57}$ 分别为品种资源输入向量因子. 最后通过多元逐步回归方程简化可得其最终决策外界因子为: $x_3, x_4, x_{11}, x_{19}, x_{27}, x_{30}, x_{34}, x_{46}, x_{48}$ 及 x_{53} 等 10 个输入因子.

由于采用自学习样本,因此本文的模型数据设计即为以 1994—2003 年训练样本由甘蔗均产及外界因子等 9×11 个变量所构成,同理,2004—2006 年 3 组样本(3×11)作为教师样本.

为验证预测仿真的精度及效率,本文同时采用灰色线性外推法、SBP、SGA-BP、SBP、RINAGA-BP 等多种算法对甘蔗产量进行预测.各 BP 模型参数的设置均为:动量系数(lr) = 0.95,迭代次数(MAXGEN) = 10 000,误差精度(e) = 0.000 01;不同模型所获得预测及误差如表 1、2.

表 1 不同模型预测的甘蔗产量
Tab. 1 Result by different model $kg \cdot hm^{-2}$

年份 ¹⁾	历史数据	灰色线性	标准 BP	SGABP	改进 GABP
2003	73 244.00	75 321.64	73 267.65	73 199.05	73 202.95
2004	69 033.22	73 280.86	69 447.47	68 944.55	69 042.83
2005	86 530.00	71 295.38	86 908.79	86 535.58	86 524.83
2006					88 524.83
2007					95 805.86
2008					92 802.65
2009					94 408.51
2010					95 302.22

1)2006—2010 年为预测年份.

表 2 不同模型的预测训练时间及误差和
Tab. 2 Training time and erroneous sum with different model of forecast

模型名称	年份	绝对误差	训练时间/s
灰色线性	2003	95 345	0.1
	2004	95 224	
	2005	97 006	
	小计	287 575	
SBP	2003	23.65	7.75
	2004	414.25	
	2005	378.79	
	小计	816.69	
SGA-BP	2003	44.95	4.57
	2004	88.67	
	2005	5.58	
	小计	139.20	
改进 GABP	2003	41.05	5.28
	2004	9.61	
	2005	5.17	
	小计	55.83	

由表 2 可知,灰色线性外推法误差和最大,时间最短,但决策可用性极低;BP 模型可显著降低预测误差.不同 BP 模型训练误差、训练时间差异较大,SGA-BP 训练时间消耗最小,改进遗传算法的 BP 预测其次,但其误差累计和最小,其模拟预测的精度最高,综合而言,效果最优.

3 结论

通过编码、遗传算子的改进,提高传统遗传算法的全局搜索和收敛效率,并以改进遗传算法对 BP 网络权阈值进行全局优化,提高了 BP 神经网络运行速度,从而有效克服了传统 BP 网络方法的缺点;针对甘蔗复杂的输入因子进行多元逐步简化,优化和提高了改进遗传 BP 在甘蔗生产预测仿真的学习效率.该预测模型与多种预测方法对比,验证了在多因素作用关联下的甘蔗产量预测效率和精度.结果显示,改进遗传算法在 BP 方法上的优越性和合理性,这对于甘蔗产量的多因素关联作用的预测预报应用具有重要参考价值.

参考文献:

[1] 沈清.神经网络应用技术[M].长沙:国防科技大学出版社,1993:45-85.

[2] 徐永春,张森文.遗传 K 均值方法在品种资源分类中的应用[J].华南农业大学学报,2009,30(2):97-100.

[3] 陈国良,王熙法,庄镇泉,等.遗传算法及其应用[M].北京:人民邮电出版社,1996:80-81.

[4] 谢凯.排挤小生境遗传算法的研究与应用[D].合肥:安徽理工大学计算机科学与工程学院,2005:43-53.

[5] SCINVIVAS M, PATNAIK L M. Adaptive probabilities of crossover and mutation in genetic algorithms[J]. IEEE Trans SMC,1994,24(4):656-666.

[6] FOGARTY T C. Varying the probability of mutation in genetic algorithms[C]//Anon. Proceedings of the 3rd International Conference on Genetic Algorithms. Boston: Boc., 1989:104-109.

[7] 徐永春.基于智能算法的甘蔗生产决策模型研究与应用[D].广州:华南农业大学工程学院,2008.

[8] 高大启.有教师的线性基本函数前向三层神经网络结构研究[J].计算机学报,1998,21(1):80-86.

[9] 陈磊,张士乔,吕谋,等.自适应遗传算法优化管网状态估计神经网络模型[J].水利水电技术,2004,35(10):61-63.

【责任编辑 周志红】