

基于智能计算的施肥模型算法研究

袁媛¹, 李淼^{1*}, 李录久², 张国兵^{1,3}, 陈曦^{1,3}, 万莉¹

(1. 中科院合肥智能机械研究所, 合肥 230031; 2. 安徽省农业科学院土壤肥料研究所, 合肥 230031;
3. 中国科学技术大学信息科学技术学院, 合肥 230026)

摘要:传统的数理统计方法构造施肥模型时, 其结构设计和因子选择多依赖先验知识, 存在一定的偶然性和主观性。该文针对这些问题, 采用 GP 构造初始施肥模型, 再利用 GA 优化初始施肥模型参数, 得到最优施肥模型。同时改进 GP/GA 初始种群的产生方法, 减少了 GP/GA 进化代数, 提高收敛效率。结果表明, 该方法能够在不给定任何先验知识的条件下, 得到很准确的模型, 为施肥模型在广大农村的应用奠定了理论基础。

关键词:施肥模型, 智能计算, 遗传规划, 遗传算法, 适应度函数

中图分类号: S14-33

文献标识码: A

文章编号: 1002-6819(2008)-12-0116-04

袁媛, 李淼, 李录久, 等. 基于智能计算的施肥模型算法研究[J]. 农业工程学报, 2008, 24(12): 116-119.

Yuan Yuan, Li Miao, Li Lujiu, et al. Algorithm of fertilization model based on intelligent computing[J]. Transactions of the CSAE, 2008, 24(12): 116-119.(in Chinese with English abstract)

0 引言

在农业生产过程中, 针对不同地区、不同作物种类和土壤状况, 不同的施肥量对作物的产量有很重要的影响。如果施肥不合理, 不仅作物产量下降, 而且容易造成作物品质下降、环境污染、土地发生次生盐渍化并使作物容易发生病害^[1]。目前, 我国大面积推广测土配方施肥技术, 多采用数理统计方法构建施肥模型, 但是, 其结构设计和因子选择多依赖领域专家的先验知识, 主观性比较强。依照现有的农村生产发展现状和广大农民现有的文化素质, 大范围推广有很大的阻力。针对这样的问题, 国内的罗长寿、朱艳、陆海燕^[2-9]等相继采用神经网络以及模糊理论解决施肥模型的自动构造问题。国外自 2001 年以来也有不少这样的研究, 美国坦普尔大学的 Dragoljub Pokrajac^[10,11]、美国圣保罗大学 JosC Alfred、Covolan Ulson^[12]等分别开展了利用神经网络构建施肥模型, 进行提高小尺度地块肥料利用率、识别作物的含氮量等研究。上述这些研究在传统的回归拟合数理统计基础上, 尝试探索一种新的解决施肥模型构建技术与理论。但是, 在应用中发现神经网络在施肥模型构造方面存在 3 个问题: ①现在普遍采用的数据都是测土配方施肥数据, 而其田间试验都是采用多点分块处理, 这些数据都处在分散非均匀的函数空间中, 给网络训练达

到要求的精度造成一定困难; ②神经网络在使用时需要针对数据量大小和分布的特点选择不同的隐含层节点, 仍然需要具有专业知识的农业专家来确定, 专家的主观性容易造成模型构造的不准确; ③使用神经网络训练田间数据, 一方面需要大样本来提高普适性, 但大数据量又会降低神经网络拟合性能^[13]。因此, 我们提出利用混合 GP/GA 来构建施肥模型, 解决施肥模型构建过程中的科学和技术问题。通过 GP 进化计算的优胜劣汰方法规避数理统计中因子选择的主观性, 探询作物施肥与土壤养分状况之间的内在相互影响, 以解决传统施肥模型构造中的数据结构问题。

1 智能计算

遗传算法 (Genetic Algorithm, GA) 是美国的 J.Holland 教授于 1975 年首先提出的, 它是一种借鉴生物界的进化规律 (适者生存, 优胜劣汰遗传机制) 来模拟随机化搜索的方法。其主要特点是可以自适应地调整搜索方向, 不需要先验规则^[14]。1992 年美国的 Koza 教授对 GA 做出了重大改进, 提出了遗传规划 (The Genetic Programming, GP)。该方法与 GA 最大的不同是“擅长解决种群的结构优化而非参数优化”^[15]。

1999 年, 英国的 Hani Hagaras 等在当年的 IEEE 工作会议上, 提出采用 GA 解决施肥、农药喷洒等的模型构造问题^[16]。2006 年加拿大的 Clément Chion 等人开展了使用 GP 构造施肥模型建模的研究, 并构建了玉米田氮含量描述性模型^[17]。据此, 我们提出了通过智能计算对施肥数据进行模型构建、模型优化等方面的研究, 探询作物产量与施肥量 (氮磷钾) 之间的内在相互作用, 探索一种通用的施肥模型构建方法。

本文提出的算法思想是利用 GP 从历史数据中先学习出一个模型框架, 由于 GP 中的参数是随机产生的, 有很大的随机性, 导致搜索空间增大, 有些结构很好的

收稿日期: 2008-01-24 修订日期: 2008-07-18

基金项目: 中国科学院知识创新工程重要方向项目 (KGCX2-SW-511); 国家自然科学基金项目 (30871451) 资助

作者简介: 袁媛 (1981-), 女, 安徽肥东人, 主要从事农业信息化及智能化应用研究。合肥 中科院合肥智能机械研究所, 230031。

Email: ahhfuyuan@yahoo.com.cn

*通讯作者: 李淼 (1955-), 女, 安徽庐江人, 研究员, 主要从事人工智能和农业知识工程研究。合肥 中科院合肥智能机械研究所, 230031。

Email: mli@iim.ac.cn

模型可能由于参数选择不合理而被抛弃，因此，我们引入GA对其参数进行优化，进一步优化模型。图1是混合GP/GA算法的流程图。

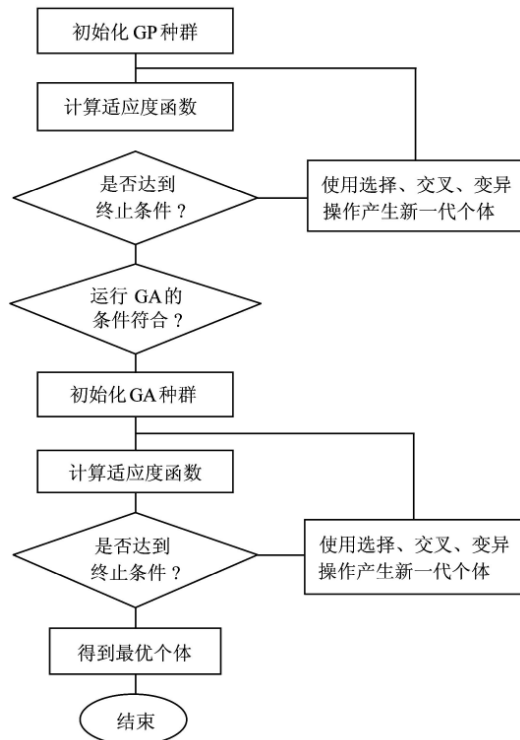


图 1 混合 GP/GA 算法流程图

Fig.1 Flow chart of hybrid GP/GA

2 改进的智能计算算法

介绍了混合GP/GA智能计算算法的基本流程，但是其中存在代码膨胀，初始种群的随机产生导致GP/GA要经过好几代的进化才能够达到一个比较好的平均适应度值，收敛速度慢等问题。

针对代码膨胀问题，我们参照唐丽珏等人^[18]提出在适应度函数中加入简约压力项。在适应度函数中加入简约因子，通过改进适应度函数，防止代码膨胀的产生，更好地驱动自然选择。简约压力项使用基于大小的适应值惩罚量来阻止代码增长。对于个体带有简约压力项的适应度函数表示如：

$$fitness_new(i) = \log(fitness(i) + \sigma S(i)) \quad (1)$$

式中 $fitness(i)$ ——个体 i 的原始适应度； $S(i)$ 给出了 i 的树状表示的复杂度，算法中指节点个数； σ ——简约系数。适应值函数包含了适应精度和个体树的大小，既要求最小适应误差，也要求模型的最简单结构，以方便遗传算法进一步优化结构分离所得的参数。同时，对求出的适应度函数值求对数，这样的做法对适应度值较小时的变化趋势有一个很好的描述，比如， $fitness(i) + \sigma S(i)$ 的值从1000到100，其值变化为900，但其对模型的意义不是很大； $fitness_new(i)$ 的值从3变为2，而当 $fitness(i) + \sigma S(i)$ 值从0.1到0.01时，虽然变化仅为0.09，但其对模型的意义能够比得上从1000到100的变化，

$fitness_new(i)$ 的值则从-1变为-2，因此，该方法能够很好的照顾适应度较小的个体，有利于个体进化。

针对初始种群随机产生导致平均适应度较低的缺点，我们提出了种群分组产生的方法。具体产生初始群体的方法为：根据初始群体的规模，按照树的深度进行分组，确定每组个体的数目，按组生成初始群体，每组初始群体的生成方法如下（其中， $pre_fitness$ 为前一组适应度函数值的平均值， $cur_fitness$ 为当前组已产生个体的适应度函数值的平均值， $depth$ 为树的深度， $group$ 为组数， max_depth 为最大树深度， max_group 为最大组数）：

(0) $pre_fitness = 0$, $cur_fitness = 0$, $depth = 1$,

$group = 0$, $max_depth = N$, $max_group = M$;

(1) 如果 $depth < max_depth$, 则随机产生两个深度为 $depth$ 的个体， $cur_fitness$ = 这两个个体的平均适应度， $group++$ ；转（2）；

(2) 随机产生一个新个体，转（3）；

(3) 比较新产生个体的适应度与 $\max(pre_fitness, cur_fitness)$ 的大小，当新个体的适应度大于等于 $\max(pre_fitness, cur_fitness)$ 时，该个体被保留下来， $cur_fitness$ = 此时本组个体的平均适应度， $group++$ ；转（4）；当新个体的适应度小于 $\max(pre_fitness, cur_fitness)$ 时，舍弃该个体，转（2）；

(4) 如果 $group = max_group$, 则

$pre_fitness = cur_fitness$, $cur_fitness = 0$, $depth++$, $group = 0$, 停止；否则转（2）；

采用这种分组方法生成的初始群体平均适应度高，且对结构简单的个体优先选择，这样不仅可以减少进化代数，提高收敛效率，而且能够得到结构较优的个体，在此基础上进行遗传、交叉、变异操作，能够快速得到最优解。

3 试 验

构建土壤施肥模型的样本数据采用的是安徽省临泉县提供的7组小麦单季“3414”施肥数据，4组用于训练样本，另外3组作为验证样本。

首先利用前面提出的改进混合GP/GA智能算法对4组数据构建施肥模型（其中一组数据见表2），其中，GP的参数设置见表1。具体构建模型过程见图1。然后统计其模型拟合误差（表3）。由表3可以看出，构造的模型对原数据的拟合度比较好，其拟合误差最大的为0.0699，最小的为0.0117，能够较好的反映原数据包含的最优施肥量。构造施肥模型的过程相比传统的数理统计方法更为简单，它可以自动获取历史试验数据所反映的客观规律，解决数理统计计算方法的结构限制问题（预先设定好的一元一次、一元二次、三元二次方程固定结构）。

另外使用3组数据进行了模型检验，得到的试验结果如表4所示，可以看出，模型的最大预测误差也仅为

0.03964, 在误差允许范围之内, 说明模型有比较好的预测能力, 运用到测土配方施肥的试验数据拟合中, 能够提供较优的施肥模型。

表1 GP参数设置情况

Table 1 Settings of GP parameters

运行参数	详细描述
函数集:	+, -, *, /, sin, cos, log
终止符集:	x, y, R ∈ [-2, 2]
适应度函数:	(1)
最大演化代数:	50
种群规模:	100
初始种群产生方法:	分组产生
最大初始树深:	6
内节点交叉概率:	90%
任意节点交叉概率:	10%
变异概率:	5%
选择方法:	赌轮选择
终止规则:	GP运行100代或适应度值为0

表2 各处理肥料用量和产量

Table 2 Designs of fertilizer inputs and grain yields for different treatments

编号	处理	N	P	K	/kg · km ⁻²		
					小区 I	小区 II	小区 III
1	N ₀ P ₀ K ₀	0	0	0	6153.08	5952.98	5902.95
2	N ₀ P ₂ K ₂	0	90	90	7053.53	6153.08	6953.48
3	N ₁ P ₂ K ₂	90	90	90	6703.35	6503.25	6553.28
4	N ₂ P ₀ K ₂	180	90	90	6303.15	6253.13	6303.15
5	N ₂ P ₁ K ₂	270	90	90	6353.18	6353.18	6603.30
6	N ₂ P ₂ K ₂	180	0	90	5802.90	6203.10	6503.25
7	N ₂ P ₃ K ₂	180	45	90	6553.28	6953.48	6603.30
8	N ₂ P ₂ K ₀	180	135	90	6303.15	6753.38	7053.53
9	N ₂ P ₂ K ₁	180	90	0	6053.03	6503.25	7253.63
10	N ₂ P ₂ K ₃	180	90	45	6353.18	6753.38	6203.10
11	N ₃ P ₂ K ₂	180	90	135	5802.90	7003.50	6203.10
12	N ₁ P ₁ K ₂	90	45	90	6103.05	6153.08	5902.95
13	N ₁ P ₂ K ₁	90	90	45	6303.15	6653.33	6303.15
14	N ₂ P ₁ K ₁	180	45	45	6303.15	6403.20	6953.48

表3 模型拟合误差

Table 3 Model fitting errors

	I	II	III	IV
最大拟合误差	0.0699	0.0143	0.0171	0.0185
最小拟合误差	0.0403	0.0117	0.0145	0.0151
平均拟合误差	0.0515	0.0131	0.0156	0.0167

表4 模型预测误差

Table 4 Model prediction errors

	1	2	3
拟合平均误差	0.0110	0.0169	0.0119
预测平均误差	0.0354	0.02844	0.03964

4 结 论

本文提出了一种基于改进的混合 GP/GA 的施肥模

型构建算法, 并从适应度函数、初始群体产生等方面作了改进。试验结果表明:

1) 该算法在构建施肥模型时能够达到较高的拟合精度, 利用该模型进行预测时, 预测误差也在允许的范围之内。这为当前测土配方施肥提供了一种智能化的施肥模型构建方法。

2) 能够在没有任何先验知识的条件下达到这样的精度, 说明该算法具有很好的泛化性。

但是, 由于农业生产的复杂性, 影响施肥的因素很多, 比如水分、光照、病虫害等。限于试验条件, 未考虑这些因素, 这将在以后的研究中进一步探讨和完善。

[参 考 文 献]

- [1] 侯彦林, 陈守伦. 施肥模型研究综述[J]. 土壤通报, 2004, 35(4): 493—501.
- [2] 罗长寿, 左 强, 李保国, 等. 基于遗传算法的人工神经网络模型在冬小麦根吸水模型中的应用[J]. 土壤通报, 2003, 34(4): 250—252.
- [3] 朱 艳, 曹卫星, 王绍华, 等. 小麦栽培管理知识模型系统的设计与实现[J]. 南京农业大学学报, 2002, 25(3): 12—16.
- [4] 林忠辉, 莫兴国. 作物生长模型研究综述[J]. 作物学报, 2003, 29(5): 750—758.
- [5] 陆海燕, 李萍萍. 基于人工神经网络技术的蔬菜施肥决策研究[J]. 中国农机化, 2004, (5): 20—22.
- [6] 米湘成, 马克平, 邹应斌. 人工神经网络模型及其在农业和生态学研究中的应用[J]. 植物生态学报, 2005, 29(5): 863—870.
- [7] 杨宇姝, 王福林. 基于神经网络的模糊土壤平衡施肥模型系统的研究[J]. 农机化研究, 2007, (10): 49—50.
- [8] 滕青芳, 秦春林. 利用神经网络建立土壤施肥模型的应用研究[J]. 兰州铁道学院学报(自然科学版), 2002, 21(4): 54—57.
- [9] 兰维娟, 毛鹏军. 基于径向基函数网络的变量施肥决策研究[J]. 安徽农业科学, 2007, 35(21): 6505—6507.
- [10] Pokrajac D, Obradovic Z. Neural network-based software for fertilizer optimization in precision farming[A]. In: Proc. International Joint Conference on Neural Networks[C]. Washington, DC, USA, 2001: 2110—2115.
- [11] Pokrajac D, Obradovic Z. A neural network-based method for site-specific fertilization recommendation[R]. ASAE Annual International Meeting. Sacramento, California, USA, 2001.
- [12] Covolan Ulson J A, Benez S H, de Silva I N, et al. Nitrogen content identification in crop plants using spectral reflectance and artificial neural networks[A]. In: International Joint Conference on Neural Networks[C]. Washington, DC, USA, 2001: 2088—2092.
- [13] Roy A. Artificial Neural Networks - A science in trouble[J]. ACM SIGKDD Explorations Newsletter, 2001, 1(2): 33—38.
- [14] Eberbach E. The role of completeness in convergence of evolutionary algorithms[J]. Evolutionary Computation, 2005. The 2005 IEEE Congress on, Sept, 2005, 2(2-5): 1706—1713.
- [15] Luger, George F. Genetic Algorithm & Genetic Programming, Artificial Intelligence, Structures and

- Strategies for Complex Problem Solving, Fourth Edition[M]. Harlow, England: Addison-Wesley, 2002: 471.
- [16] Hagras H, Victor Callaghan. A fuzzy-genetic based embedded-agent approach to learning & control in agricultural autonomous vehicles[A]. In: IEEE International Conference on Robotics and Automation[C]. Detroit, USA, 1999: 1005–1010.
- [17] Chion C, LE Da Costa, Jacques-André Landry. Genetic programming for agricultural purposes[A]. In: Proceedings of the 8th annual conference on Genetic and evolutionary computation[C]. Washington, USA, 2006: 783–790.
- [18] 唐丽珏, 李 淼, 张 建. 混合 GP-GA 用于信息系统建模预测的研究[J]. 计算机工程与应用, 2004, 40(25): 43–48.

Algorithm of fertilization model based on intelligent computing

Yuan Yuan¹, Li Miao^{1✉}, Li Lujiu², Zhang Guobing^{1,3}, Chen Xi^{1,3}, Wan Li¹

(1. Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, China;

2. Soil and Fertilizer Institute, Anhui Academy of Agricultural Science, Hefei 230031, China;

3 School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China)

Abstract: When traditional mathematical statistic method is used to build fertilization model, the structure design and factor choosing largely rely on prior knowledge. Consequently, the result is somewhat casual and subjective. In order to settle this problem, this paper applies Genetic Programming(GP) to construct initial fertilization model and then applies Genetic Algorithm(GA) to optimize the parameters of the initial fertilization model. In this way, an optimal fertilization model comes into being. Meanwhile, the authors present a method for improving the production of GP/GA initial population, the method decreases evolution generation and improves convergence efficiency. Experimental results show that this approach can build very accurate model without any prior knowledge, which contributes as theoretical foundation to the application of fertilization model in vast rural areas.

Key words: fertilization model, intelligent computing, Genetic Programming (GP), Genetic Algorithms (GA), fitness function