

协同进化理论及其在施肥模型中的应用

郑高伟^{1,2} 李 淼¹ 高会议¹ 李录久³

(1. 中国科学院 合肥智能机械研究所, 合肥 230031; 2. 中国科学技术大学 信息科学技术学院, 合肥 230026;

3. 安徽省农业科学院 土壤肥料研究所, 合肥 230031)

(bamboo@mail.ustc.edu.cn)

摘 要:采用数理统计方法进行施肥模型构造,由于受到固定的数学结构的限制,导致有一些实验结果因不能被模型拟合而被舍弃,造成了一些数据的浪费。针对这些问题,提出了基于协同进化理论的施肥模型构建算法,将模型构建问题分解为模型结构构建与模型参数优化两个子问题,并将这两个子问题抽象成多种群间协同进化。使用遗传规划算法进行模型结构构建,使用遗传算法对模型参数进行优化,两个过程协同进行。实验结果表明,该算法能够在历史实验数据的基础上自动生成动态模型,同时具有较好的准确度。

关键词:协同进化; 遗传规划; 遗传算法; 施肥模型; 适应度函数

中图分类号: TP399 **文献标志码:** A

Co-evolution theory and its application in fertilization model

ZHENG Gao-wei^{1,2}, LI Miao¹, GAO Hui-yi¹, LI Lu-jiu³

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

2. School of Information Science and Technology, University of Science and Technology of China, Hefei Anhui 230026, China;

3. Institute of Soil and Fertilizer, Anhui Academy of Agricultural Science, Hefei Anhui 230031, China)

Abstract: Due to the limitation of the fixed mathematical structure, it can not fit all the data; as a result, some experimental results may be discarded in constructing fertilization model, which results in a waste of some data. To solve these problems, the algorithm of fertilization construction model based on co-evolution theory was proposed. The algorithm divided the whole construction model problem into two sub-problems: model structure construction and model parameter optimization. In addition, the two sub-problems were abstracted as co-evolution among various groups. The processes of using genetic programming to construct the model structure and genetic algorithm to optimize the model parameters were carried out in collaboration. The results show that the algorithm not only generates the dynamic model automatically based on the historical experimental data but also has higher accuracy.

Key words: co-evolution; Genetic Programming (GP); Genetic Algorithm (GA); fertilization model; fitness function

0 引言

施肥是促进粮食增产保收的重要手段,大量研究结果表明,化肥在粮食增产中的贡献率高达40%~50%^[1-2]。在我国人口众多、耕地等资源严重不足条件下,通过施肥等技术实现作物高产和水土资源替代战略,对我国农业可持续发展具有特殊的重要意义。长期以来我国在肥料施用方面存在许多突出问题(化肥投入量大、施肥结构不合理、使用方法不当等),导致肥料利用率低,主要粮食作物肥料利用效率较低,其中氮肥利用率为26.1%~28.3%,磷肥利用率为7.3%~20.1%,钾肥利用率为21.2%~35.9%,远低于国际水平,与20世纪80年代相比呈下降趋势^[3]。不仅造成农业生产成本提高和资源浪费,而且对农业环境造成了污染,是影响水质恶化的面源污染中的重要方面。因此,合理施肥是关系到我国农业可持续发展的重大技术问题,其重要性和紧迫性随着农业生产的发展和化肥施用量的增加而显得越来越突出。

合理构造施肥模型是施肥技术的核心内容之一,目前国内正在大面积推广应用的测土配方施肥方法多是采用数理

统计方法来进行施肥模型构造,该方法首先设定施肥模型结构为三元二次方程,然后利用施肥实验数据进行模型拟合。在拟合过程中,由于模型拟合成功率不高,会出现一些实验数据不能拟合而被舍弃,这在一定程度上降低了对施肥数据信息的有效利用,进而造成用部分实验结果反映整体上的问题,造成信息失真;同时也增加了实验成本和难度^[4-6]。

本文针对施肥实验数据的特点提出基于协同进化理论进行施肥模型构建,使施肥模型在拟合的过程中根据特定的实验数据动态选择模型结构,同时在模型参数优化过程中引入氮磷钾适宜配比因子,控制参数优化过程。

1 协同进化理论

协同进化最早由 Ehrlich 和 Raven 提出,用于描述植物和植食昆虫相互之间的进化影响^[7];是指一个物种的性状作为对另一个物种性状的反映而进化,而后一物种的这一性状本身又是作为对前一物种性状的反映而进化^[8-9]。这一定义要求每一个性状的进化都是由于另一个性状;并且两个性状都要进化,具有相互性。协同进化的研究内容极为广泛,包括:

收稿日期:2010-11-30;修回日期:2011-01-19。 基金项目:国家自然科学基金资助项目(30871451)。

作者简介:郑高伟(1984-),男,河南新蔡人,硕士研究生,主要研究方向:模式识别、智能系统;李淼(1955-),女,安徽庐江人,研究员,博士生导师,主要研究方向:人工智能、农业知识工程;高会议(1982-),男,安徽砀山人,博士,主要研究方向:生态环境与农业信息;李录久(1962-),男,安徽合肥人,研究员,主要研究方向:土壤资源、环境、植物营养与施肥。

竞争物种间的协同进化、捕食者与猎物系统的协同进化、寄生生物与寄主系统的协同进化、拟态的协同进化、相互作用的协同进化等^[10]。

协同进化算法是近年来在协同进化理论上提出的一类新的进化算法,协同进化算法与传统的进化算法的区别在于:协同进化算法在进化算法的基础上,考虑了种群与环境之间、种群与种群之间在进化过程中的协调。文献[11]提出了协同进化增广 Lagrangian 方法求解约束优化问题;文献[12-13]中提出了基于生态模型的协同进化算法,实现了根据适应度情况动态调整各个模式在种群中的比例的目的;文献[14]提出了用于函数优化问题的协同进化遗传算法(Cooperative Coevolutionary Genetic Algorithm,CCGA)。

协同进化计算在数据挖掘、数值优化、多目标优化等方面都有很广泛的应用。文献[10]针对数据挖掘的特点,结合协同进化机制,提出了组织协同进化分类算法(Organizational CoEvolutionary algorithm for Classification,OCEC),进一步将协同进化理论应用到数值优化问题上,提出了组织进化算法(Organizational Evolutionary Algorithm,OEA),理论分析证明该算法能够收敛于全局最优解;文献[14]最早使用合作型协同进化算法解决函数优化等问题;文献[15]从协同进化算法的共性入手,以适应度评估过程中的个体相互作用为出发点,提出了动态种群规模的协同进化算法模型(Model of Co-Evolutionary Algorithm with Dynamic population size,CEAD),并从动力学、稳定性、收敛性等方面对 CEAD 进行理论分析,结果表明 CEAD 模型具有全局渐近稳定性,在其指导下的协同进化算法可以保证种群规模能够自适应地调节至合适的水平;文献[16]分析了智能计算技术应用于农业施肥模型构造中的可行性,研究表明,使用智能计算技术构造施肥模型的关键在于构造模型结构和参数优化。

2 基于协同进化理论构建施肥模型

施肥模型的构建一般包括施肥模型结构构建和参数优化两部分。由于遗传算法(Genetic Algorithm,GA)的主要特点是可以自适应地调整搜索方向,不需要先验规则。遗传规划(Genetic Programming,GP)的特点是擅长解决种群结构优化^[16-17]。文献[16]使用遗传规划算法进行模型结构的构造,然后使用遗传算法对模型参数进行优化。本文基于文献[16]提出了使用协同进化理论改进文献[16]中算法进行施肥模型的构造,将施肥模型的构建问题分解为模型结构的构造和模型参数优化两个子问题,采用遗传规划算法进行模型结构构造;采用遗传算法进行模型参数优化,并将遗传算法的参数优化过程嵌入到遗传规划过程中,实现遗传算法与遗传规划的协同进化。算法流程如图1所示。

算法使用 GP 算法进行施肥模型结构的构造,使用 GA 进行模型参数优化,通过氮磷钾适宜配比和适应度函数将两个过程进行协同,最终产生最优施肥模型。在使用遗传规划算法进行模型结构构造时,只对构成模型结构的变量进行编码,而在使用遗传算法进行参数优化时,只对模型参数变量进行编码。在进行个体评价时,需要综合利用两个过程的信息,以构成一个完整的决策变量编码,才可以利用适应度函数进行评价^[18]。

使用 GP 进行模型结构构建首先选择变量和函数,即构成模型的结构包括变量和函数。由于在农业生产的施肥过程中,氮磷钾养分是作物生长发育必需的三大营养元素,缺一不可。有研究表明,作物吸收养分与土壤有效养分利用率呈显著负相关,与土壤供应量呈对数曲线关系,与当季肥料养分利用率呈负相关^[1]。因此,选择氮(N)、磷(P)、钾(K)作为 GP 变量,选择 +、-、×、/、log、exp 为函数,并引入氮磷钾适宜配比领域因子对模型构造过程进行约束。选择估计值与实际值之间的均方误差(Mean Squared Error,MSE)作为个体的适应度因子,以预测值与实验值之间的相对误差作为终止判定条件,即:

$$\text{函数集 } F = \{ +, -, \times, /, \log, \exp \}, \text{ 变量集 } N_f = \{ N, P, K \};$$

$$N : P : K = n_N : n_P : n_K \quad (1)$$

$$RE_i = \frac{|y_i - y'_i|}{y_i} \quad (2)$$

$$Fiction_GP(i, j) = \lg \left(r_{\max} - \sqrt{\frac{\varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_n^2}{n}} \right) \quad (3)$$

其中: F 为函数集, N_f 为变量集, $Fiction_GP(i, j)$ 为第 i 代群体第 j 个个体适应度函数, ε_i^2 为预测值与实际值之间的方差, $\varepsilon_i^2 = (y_i - y'_i)^2$, y_i 为实际值, y'_i 为预测值, r_{\max} 为均方误差所能达到的最大值。

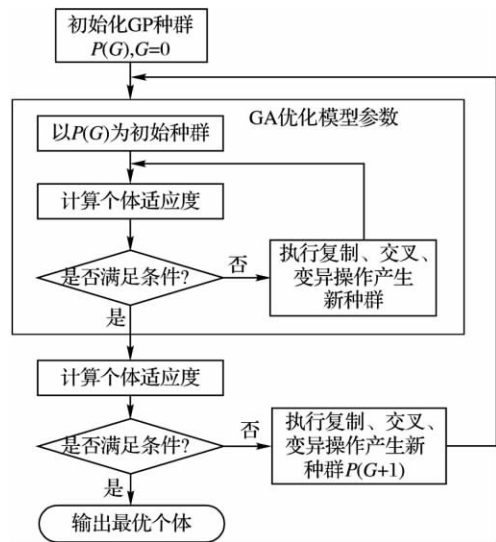


图1 协同进化施肥模型构造算法流程

使用 GA 对 GP 构造的模型结构个体进行参数优化时,引入氮磷钾适宜配比领域因子对优化过程进行约束。氮磷钾的影响可以通过作物的产量来反映,即氮磷钾施用量越合理,作物产量越高,反之亦然。优化结果根据模型对实验数据拟合程度进行评价,即根据模型拟合误差进行评价模型参数优化结果。则 GA 个体的适应度可以设计为:

$$Fiction_GA(i, j) = \lg \left(r_{\max} - \sqrt{\frac{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}{n}} \right) \quad (4)$$

其中: σ_i^2 为预测值与实际值之间的方差, $\sigma_i^2 = (y_i - y'_i)^2 \times N_y$; y_i 为实际值, y'_i 为预测值, r_{\max} 为均方误差所能达到的最大值, N_y 为对应作物产量值。

基于协同进化理论施肥模型构造算法分为算法 I 和算法 II,即模型结构构造算法和模型参数优化算法,两个过程协同

进行, 具体如下。

算法 I:

1) 根据给定参数, 随机产生初始种群 $P(G)$, 即产生有函数和变量构成的初始模型结构, 在选择变量时根据氮磷钾适宜配比进行选择, 选择变量 N 的概率为 $P(N) =$

$$\frac{n_N}{n_N + n_P + n_K}$$

选择变量 P 的概率为 $P(P) = \frac{n_P}{n_N + n_P + n_K}$, 选

$$择变量 K 的概率为 P(K) = \frac{n_K}{n_N + n_P + n_K}, G = 0。$$

2) 将种群 $P(G)$ 传给算法 II 进行参数优化, 获得返回后的种群 $P'(G)$ 。

3) 在 $P'(G)$ 中使用赌轮法选择具有较高适应度的个体, 并形成新的子中群体 $P''(G)$ 。

4) 在群体 $P''(G)$ 中执行复制、交叉、变异操作, 生成群体 $P(G+1)$, 复制、交叉、变异算子均采用自适应算子进行。其中复制操作根据适应度—比例原则(即个体适应度越高, 参与复制的可能性越大), 个体复制概率为 $p_{si} = \frac{Fiction_GP(i, j)}{\sum_{j=1}^M Fiction_GP(i, j)}$ 种群规模为 M ; 交叉操作也根据适应度—比例原则进行, 采用单点交叉, 个体进行交叉概率与复制概率计算公式相同; 变异算子根据交叉所得两个新个体的海明距离进行变化, 海明距离(即新个体编码不同的位数)越小, P_m 越大, 反之 P_m 越小。

5) 终止判定条件: 满足终止条件, 则输出结果, 不满足条件则返回到步骤 2) 继续迭代。

算法 II:

1) 根据算法 I 的种群 $P(G)$ 作为初始种群。

2) 计算种群个体适应度。

3) 对种群执行复制、交叉、变异操作, 并使用精英法(即保留最佳个体)选择具有较高适应度个体, 作为子群体。其中交叉、变异原则与算法 I 相同。

4) 终止条件判定: 满足终止条件, 将结果输出给算法 I, 不满足条件则返回到步骤 2) 继续迭代。

由上可知, 协同进化算法将施肥模型构建问题分解为建模和优化两个相互关联的子问题, 每个子问题由一个种群表示。各个种群可以有相同或不同的基因表达方式, 取决于具体的问题以及进化算法的特性, 种群间存在相互合作或竞争关系。协同进化算法相对于传统的智能算法来说具有两个优点: 1) 使用不同的基因表现形式表达求解问题有助于不同的智能算法之间互相弥补短板; 2) 一部分种群可以针对问题求解中的一个子问题进行求解而不需要对整个问题进行搜索。这样可以更好地发挥不同智能算法的优点, 通过竞争与协作求得最优解。同时, 由于文献[16]中算法是首先使用 GP 进行模型结构构造, 然后使用 GA 对模型参数优化, 两个过程独立进行, 由于 GP 算法对参数优化能力较弱, 就会导致 GP 算法会抛弃一些结构较好、但由于参数不好的模型结构, 从而可能导致模型不能找到全局最优模型。本文将两个算法协同进行, 就可以避免文献[16]中算法的不足, 得到较好的结果。

3 实验及结果分析

本文实验数据采用 4 组安徽淮北小麦单季“3517”施肥

实验数据, 6 组安徽临泉单季玉米“3517”施肥实验数据, 共 10 组实验数据, 6 组用于训练样本(包括 2 组安徽淮北小麦实验数据, 4 组安徽临泉玉米实验数据), 另外 3 组作为验证样本(包括 2 组安徽淮北小麦实验数据, 2 组安徽临泉玉米实验数据)。

首先利用前面提出的基于协同进化理论的施肥模型构造算法对 8 组数据构建施肥模型(其中一组数据见表 1), 然后利用验证样本数据对模型进行检验, 同时与传统数理统计方法以及文献[16]中的方法构造的模型进行比较。其中 GP 的参数设置见表 2, GA 的参数设置见表 3。

表 1 安徽淮北小麦各处理肥料用量和产量

序号	处理	小区 I	小区 II	小区 III
1	N ₀ P ₀ K ₀	6299.7	6271.5	6348.3
2	N ₀ P ₇₅ K ₉₀	6296.9	6346.5	6311.5
3	N ₉₀ P ₇₅ K ₉₀	7598.5	7762.8	7320.1
4	N ₁₈₀ P ₀ K ₉₀	7188.7	6707.5	6836.1
5	N ₁₈₀ P _{37.5} K ₉₀	7944.6	7274.8	7688.5
6	N ₁₈₀ P ₇₅ K ₉₀	8202.9	8021.7	8028.2
7	N ₁₈₀ P _{112.5} K ₉₀	8118.3	7128.2	7374.1
8	N ₁₈₀ P ₇₅ K ₀	7144.3	7232.7	7278.4
9	N ₁₈₀ P ₇₅ K ₄₅	7356.1	7664.0	7776.6
10	N ₁₈₀ P ₇₅ K ₁₃₅	7571.9	7683.5	8274.1
11	N ₂₇₀ P ₇₅ K ₉₀	7577.9	8084.0	7767.0
12	N ₉₀ P _{37.5} K ₉₀	6907.4	6719.0	5348.2
13	N ₉₀ P ₇₅ K ₄₅	6110.9	7001.9	5943.2
14	N ₁₈₀ P _{37.5} K ₄₅	7056.6	7027.2	7192.4
15	N ₃₆₀ P ₇₅ K ₉₀	7043.5	7227.3	7569.7
16	N ₁₈₀ P ₁₅₀ K ₉₀	7304.8	7365.6	7632.0
17	N ₁₈₀ P ₇₅ K ₁₈₀	7990.5	7552.9	7578.5

表 2 GP 参数设置

参数	描述
函数集	+ , - , * , / , log , exp
适应度函数	Fiction_GP
最大演化代数	1000
种群规模	100
交叉概率	P_{si}
变异概率	P_{si}
选择方法	轮赌选择

表 3 GA 参数设置

参数	描述
种群大小	100
交叉率	P_{si}
变异率	P_{si}
选择策略	精英选择

使用本文协同算法、数理统计方法和文献[16]方法对安徽淮北小麦实验数据进行模型拟合, 小麦产量误差如表 4、图 2 所示。

表 4 模型拟合误差——安徽淮北小麦

算法	I			II		
	最大误差	最小误差	平均误差	最大误差	最小误差	平均误差
本文算法	0.165	0.000	0.060	0.104	0.02	0.046
数理统计算法	1.109	0.089	0.430	0.386	0.06	0.173
文献[16]算法	0.357	0.013	0.146	0.309	0.03	0.053

其中,在安徽淮北小麦第 I 组数据中,使用数理统计方法进行拟合的相关系数为 0.130,使用本文方法进行拟合的相关系数为 0.980,使用文献[16]中的算法进行拟合的相关系数为 0.937;第 II 组数据中,使用数理统计方法进行拟合的相关系

数为 0.976,使用本文方法进行拟合的相关系数为 0.984,使用文献[16]中的算法进行拟合的相关系数为 0.982。

分别使用三种算法对安徽临泉地区小麦实验数据进行模型拟合,实验结果如表 5 所示。

表 5 模型拟合误差——安徽临泉玉米

算法	I			II			III			IV		
	最大误差	最小误差	平均误差	最大误差	最小误差	平均误差	最大误差	最小误差	平均误差	最大误差	最小误差	平均误差
本文算法	0.125	1.023	0.179	0.141	0.735	0.163	0.093	2.357	0.079	0.128	0.483	0.201
数理统计算法	0.007	0.024	0.017	0.003	0.079	0.008	0.000	0.006	0.002	0.006	0.086	0.007
文献[16]算法	0.040	0.471	0.053	0.041	0.159	0.069	0.026	0.932	0.034	0.045	0.109	0.059

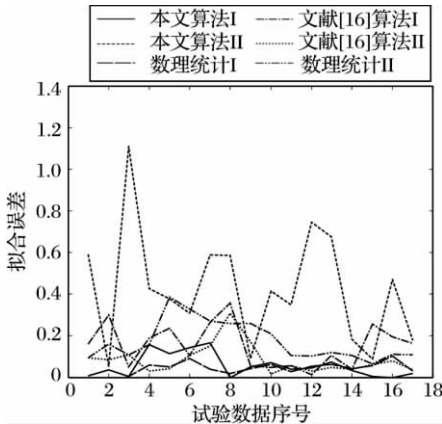


图 2 模型拟合误差

由表 5 可以看出,使用本文算法构造的模型与数理统计方法和文献[16]中方法构造的施肥模型相比,对原始数据的拟合度更好,能够更好地反映原始数据包含的信息。同时模型构造算法具有很好的自适应性,对不同地区、不同农作物具有自适应选择模型结构,有效地解决了数量统计计算方法的结构限制问题(预先设计一元一次、一元二次或三元二次方程结构),也在一定程度上提高了文献[16]中算法的性能。

使用另外 1 组安徽淮北小麦实验数据和 2 组安徽临泉玉米实验数据进行模型验证,得到的小麦和玉米产量预测结果如表 6 所示,可以看出,模型的最大预测误差为 0.373,在误差允许范围之内,说明模型具有较好的预测能力,能够为施肥提供较好的施肥模型。

表 6 模型预测误差

误差	淮北小麦	临泉玉米 1	临泉玉米 2
拟合平均误差	0.008	0.046	0.072
预测平均误差	0.045	0.194	0.373

4 结语

基于协同进化理论构造施肥模型算法,解决了数理统计构建施肥模型中存在的问题(即数据浪费,模型不能拟合),与数理统计方法相比,具有:

1) 在使用数理统计方法进行拟合的情况下,本文方法能够给出更好的拟合结果;

2) 在使用数理统计方法无法拟合时,本文算法拟合结果不受数据影响,能够拟合出很好的结果。

由此可知,本文算法比数理统计方法具有更好的普适性,拟合结果不受实验数据的约束。

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] 张福锁.测土配方施肥技术要览[M].北京:中国农业大学出版社,2006:1-2.
- [3] 张福锁,王激清,张卫峰,等.中国主要粮食作物肥料利用率现状与提高途径[J].土壤学报,2008,45(5):915-924.
- [4] 侯彦林.“生态平衡施肥”的理论基础和技术体系[J].生态学报,2000,20(4):653-658.
- [5] COVOLAN ULSON J A, BENEZ S H, NUNES DA SILVA I, et al. Nitrogen content identification in crop plants using spectral reflectance and artificial neural networks [C]// IJCNN01: Proceedings of International Joint Conference on Neural Networks. Washington, DC: IEEE, 2001:2088-2092.
- [6] 王圣瑞,陈新平,高祥照,等.“3414”肥料实验模型拟合的探讨[J].植物营养与肥料学报,2002,8(4):409-413.
- [7] EHRlich P R, RAVEN P H. Butterflies and plants: A study in co-evolution[J]. Evolution, 1964, 18(4): 586-608.
- [8] 蓝盛芳.试论达尔文进化论与协同进化论[J].生态科学, 1995(2): 167-170.
- [9] 李振基,陈小麟,郑海雷.生态学[M].北京:科学出版社, 2000: 89-193.
- [10] 刘静.协同进化算法及其应用研究[D].西安:西安电子科技大学, 2004: 29-95.
- [11] TAHK M J, SUN B C. Coevolutionary augmented Lagrangian methods for constrained optimization [C]. IEEE Transactions on Evolution Computation, 2000, 4(2): 114-124.
- [12] 曹先彬,罗文坚,王煦法.基于生态种群竞争模型的协同进化[J].软件学报, 2001, 12(4): 556-562.
- [13] 郑浩然.基于生态特征的进化与协同研究[D].合肥:中国科学技术大学, 2000: 27-75.
- [14] POTTER M A, De JONG K A. Cooperative coevolution: An architecture for evolving coadapted subcomponent [J]. Evolutionary Computation, 2000, 8(1): 1-29.
- [15] 郭圆平.动态种群规模的协同进化算法模型——理论与应用[D].合肥:中国科学技术大学, 2008: 25-98.
- [16] 袁媛,李森,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.
- [17] EBERBACH E. The role of completeness in convergence of evolutionary algorithms [C]// The 2005 IEEE Congress on Evolutionary Computation. Edinburgh: IEEE Press, 2005: 1706-1713.
- [18] 巩敦卫,孙晓燕.协同进化遗传算法理论及应用[M].北京:科学出版社, 2009: 6-15.