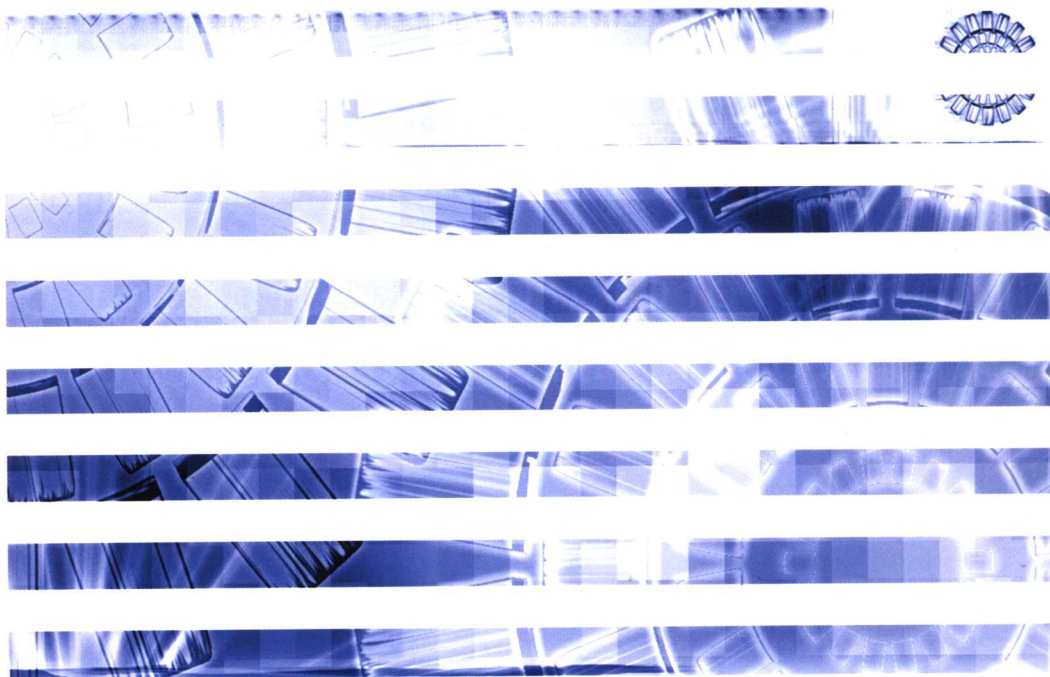


付 强◇著

数据处理方法及其 农业应用



科学出版社

www.sciencep.com

数据处理方法及其农业应用

付 强 著

科 学 出 版 社

北 京

内 容 简 介

本书针对农业大系统复杂、不确定、模糊、随机等诸多特点,将最新的数据处理方法、建模与优化技术全面应用于其中.以智能优化算法(遗传算法、免疫算法等)及其在农业系统中的应用为基础,详细介绍了多种数学分析与处理方法,包括灰色理论、模糊数学理论、人工神经网络模型、数学规划、时间序列分析模型、回归分析、混沌分析理论、分维与分形、集对分析与粗糙集理论、投影寻踪技术、物元可拓理论、小波分析等.在介绍各种数据分析方法基本原理的同时,还进行了各种模型理论之间的耦合分析.针对每种方法列举了大量的典型例题,系统阐述了现代数学分析方法在农业科学领域中的应用.

本书对于多种数学分析方法的介绍力求浅显易懂,使读者能够很快掌握并得以应用.数据处理方法的介绍以及大量的实例分析对于解决农业大系统中诸多问题具有较强的理论指导意义及应用价值,为农业科学研究提供了诸多理论参考.

本书介绍的理论方法具有广泛的适用性和推广性,可供从事农业工程、环境工程、系统工程、水文学及水资源、农林经济管理及其他相关专业的教学、科研、管理等工作的读者借鉴与参考.

图书在版编目(CIP)数据

数据处理方法及其农业应用/付强著. —北京:科学出版社, 2006

ISBN 7-03-017124-1

I. 数… II. 付… III. 数学分析-分析方法-应用-农业 IV. S11

中国版本图书馆CIP数据核字(2006)第031869号

责任编辑: 鄢德平 范庆奎 彭克里 刘 晶 / 责任校对: 张 琪

责任印制: 安春生 / 封面设计: 王 浩

科学出版社出版

北京东黄城根北街16号

邮政编码: 100717

<http://www.sciencep.com>

丽源印刷厂印刷

科学出版社发行 各地新华书店经销

*

2006年8月第一版 开本: B5(720×1000)

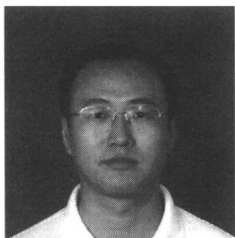
2006年8月第一次印刷 印张: 31 1/2

印数: 1—3 000 字数: 603 000

定价: 65.00元

(如有印装质量问题, 我社负责调换〈新欣〉)

作者简介



付强,男,1973年生,汉族.中共党员,博士,教授,博士生导师,东北农业大学农业水土工程学科带头人.

1995年和1997年在东北农业大学分别获得农田水利工程专业和农业水土工程专业工学学士、硕士学位.2000年6月获东北农业大学农业机械化工程专业(农业系统工程方向)工学博士学位.2002年5月,在四川大学水利工程博士后流动站水文水资源专业完成博士后研究工作.2004年6月至今在北大荒农垦集团公司博士后科研工作站做二站博士后研究工作.

主要从事节水灌溉理论与技术、水资源系统分析、湿地水管理及农业系统工程建模与优化技术方面研究.先后为本科生、硕士研究生、博士研究生讲授专业课程9门,指导硕士研究生40名,已毕业并获学位15人,指导博士研究生1名,协助指导博士研究生8名.

现为国际水土保持协会终生会员,中国农业工程学会高级会员,黑龙江省农业工程学会理事,全国中文核心期刊《农机化研究》、《黑龙江水专学报》、《水利科技与经济》编委,水利与建筑学院学位委员会、教授委员会委员,黑龙江省博士研究会委员.

主持国家自然科学基金、国家“863”专项、中国博士后科学基金、黑龙江省青年基金等国家及省部级科研项目13项,主要参加国家级、省部级科研项目15项,生产项目5项.先后在国内及荷兰、美国、英国等40余种学术刊物发表学术论文160余篇,其中EI收录17篇,SCI收录5篇,ISTP收录6篇,CSCD收录32篇,CSCD引用27篇,出版专著3部,合作出版专著2部,主编教材1部,副主编教材3部.获省市级各项科研奖励12项.

前 言

作为自然科学与社会科学的基础,数学及其应用已渗透到社会的各个方面,而且将继续形成更多新的交叉学科.掌握现代数学的理论和方法是 21 世纪各类专门人才所必须具备的基本条件之一.数学与应用数学本身也在其深度和广度上向前发展.随着现代科学技术的发展,对数学的要求越来越多,也越来越高.过去,很多专业,如农学类研究需要的数学内容很少,但是,伴随着现代农业和生态农业的发展,农学和生物类专业对数学的需求也有很大的增加,特别是生物信息学的发展对数学要求也是很高的.因此,应用数学模型、方法及其应用的研究也成为今后重点发展的研究方向.

目前,随着社会的发展和人类的进步,在各种外部环境因素的影响下,社会大系统变得越来越复杂.探索社会与自然规律,保证社会经济、人口、资源、环境的可持续发展成为当前与今后相当长一段时间内的研究主题.无论是自然科学还是社会科学,探索事物与现象的产生、发展和变化的过程,进而揭示事物与问题的本质,都是科学研究的根本目的.在这个过程中,科研工作者通过大量的观测、试验,获取了大量的数据,通过数据的分析与处理,将量化的数据进行定性化分析,进而探求事物与现象的内在规律,通过现象挖掘其本质.数据处理的方法很多,回归分析是最常用的方法之一,它通过试验观测,获取一定的原始数据,通过数据处理,寻找单个或多个自变量与单个或多个因变量之间的关系,为生产实践提供科学理论依据.对于简单的现象,我们可以采用较常规的数据处理方法来解决,而数据内部间存在关联时,再采用传统的数据处理方法就不合适了.例如,传统的多元回归分析模型,是在各个自变量之间相关性不显著的情况下,采用最小二乘法进行参数估计,从而建立数学模型.但是若各个自变量之间存在严重相关性时,仍采用以往的数据处理方法,虽然也可以建立这样一个多元回归方程,但是自变量的符号、系数等均会出现与现实不相符的问题,而且模型的精度很差,稳健度减低.为此,我们可采用最新的被称为第二代回归分析技术——偏最小二乘回归技术(PLS)来有效地解决这个问题.另外,系统本身尤其是农业大系统内部存在越来越多的随机、模糊、不确定性的问题,这些问题本身可能同时具有多个特点,即多种复杂现象与特点交织在一起,若简单使用单一的数学模型很难处理,这就需要对现有的数学理论进行改进与完善,最常用的就是数学模型理论之间的耦合分析.例如,我们可以根据洪水数据的特征,将灰色理论与模糊数学理论进行耦合分析,从而实现洪水的预报;或者将灰色理论与偏最小二乘回归技术耦合,共同对某一事物进行预报.一种好的、实用的数据方法可以最大限度地揭示事物与现象的本质和内在规律,进而为所研究的问题奠定扎实

的理论基础,推动科学研究的不断深入.

与此同时,随着知识的更新,涌现出越来越多的数学新方法,这些方法在处理实际问题时非常必要和有效.尤其是农业大系统中内部与外部环境之间关系越来越复杂,诸多复杂、不确定、模糊、随机等现象交织在一起,若采用传统的数据处理方法,将很难解释与表述复杂系统的内部规律.另外,许多学生与科研工作者的大部分精力放在试验研究上,而忽视了定性化基础研究,等到处理数据时却苦于寻找不到比较合适的关于数据处理的图书;即使找到了,其理论深度也较深,不适合应用.而本书所介绍的多种数据处理方法浅显易懂、深入浅出,使读者能够很快掌握并得以应用,为读者提供了研究思路与模式.大量的实例分析在帮助读者解决农业大系统中诸多问题方面具有较强的理论指导意义及应用价值,同时也为农业科学研究领域提供了诸多理论参考.

近年来,随着人工智能的发展、交叉学科的不断涌现,又出现了许多新的数学模型理论与优化技术,如人工神经网络技术、小波分析技术、分维与分形、混沌理论、智能优化算法(遗传算法、免疫算法、蚁群算法、模拟退火算法等).这些理论是解决越来越复杂的农业大系统中存在的复杂问题的有效工具与手段,也使复杂烦琐的数据处理和揭示事物内部规律成为可能.为此,本书中各个章节将目前最新的数据处理方法向读者逐一介绍,并将各种方法结合农业大系统中的实际问题,通过实例分析的形式展示给读者,将现代建模与优化技术系统全面应用到农业科学领域,为从事农业科学研究的读者提供新的研究思路与数据处理模式,从而解决农业生产中的实际问题.同时,模型理论本身也可以应用推广到国民经济其他相关研究领域.

本书是作者近十年的有关研究工作的总结,并在总结多年科研经验的基础上,对多种数学方法进行了系统归纳与总结,为该领域研究提供了新的研究思路与研究模式,可作为广大在读研究生、科研工作者处理农业实际问题的理论指导,从而解决农业生产中的实际问题.同时,书中介绍的诸多模型理论有较多的改进方案,对于理论本身及其实际应用均有较高的学术价值.由于时间与精力有限,本书中还有许多研究工作需要今后不断补充和完善.

本书作者参考和引用了大量国内外学者的有关论著,吸收了同行们的辛勤劳动成果,从中得到了很大的教益与启发,在此谨向各位学者表示衷心的感谢!

在本书的研究过程中,东北农业大学、合肥工业大学、四川大学、华北水利电力学院、中国科学院地理科学与资源研究所、河海大学、黑龙江大学等单位的同志给予了帮助与指导;在具体撰写过程中,东北农业大学水利与建筑学院赵小勇、贺延国、刘仁涛、冯艳、李国良、李伟业参与了部分章节的编写与校对工作,在此一并表示衷心的感谢!

由于水平有限,书中不足之处在所难免,恳请广大读者批评指正!

注：本书得到国家自然科学基金 (No.30400275)、国家“863”计划项目 (No.2002AA2Z4251-09)、中国博士后科学基金 (No.2004035167)、黑龙江省青年基金 (No.QC04C28)、黑龙江省教育厅科研基金 (No.10541033)、黑龙江省教育厅人文社科基金 (No.1054xy006)、黑龙江省博士后科学基金 (LHK-02001)、北大荒集团公司博士后科研工作站博士后科研基金 (No.LRB04-069)、东北农业大学博士后科学基金 (No.240009) 的联合资助。

作 者

目 录

前言

第 1 章 遗传算法	1
1.1 遗传算法简介.....	1
1.2 改进的遗传算法.....	4
1.3 大型灌区非均匀给水系统中分区给水优化的新方法.....	14
参考文献.....	18
第 2 章 灰色系统理论及其农业应用	20
2.1 灰色系统理论的概念与基本原理.....	20
2.2 灰色关联分析及其农业应用.....	23
2.3 灰色预测模型在农业系统中的应用.....	31
2.4 灰色聚类分析与灰色评价.....	40
参考文献.....	45
第 3 章 时间序列分析及其农业应用	47
3.1 时间序列·时间序列分析.....	47
3.2 时间序列分析在农业水资源预测中的应用.....	49
3.3 时间序列分析的研究进展.....	97
参考文献.....	98
第 4 章 模糊数学理论及其农业应用	100
4.1 模糊数学基本概念和原理.....	101
4.2 模糊聚类分析及其农业应用.....	105
4.3 模糊综合评判及其农业应用.....	123
4.4 模糊数学模型与其他模型的耦合.....	137
参考文献.....	162
第 5 章 人工神经网络模型及其农业应用	164
5.1 人工神经网络模型的基本原理.....	164
5.2 人工神经网络模型的分类.....	171
5.3 人工神经网络模型的农业应用.....	171
5.4 人工神经网络模型的农业应用前景展望.....	202

参考文献	203
第 6 章 回归分析及其农业应用	205
6.1 回归分析基本原理与方法	205
6.2 一元与多元回归分析及其农业应用	205
6.3 主成分分析及其农业应用	217
6.4 偏最小二乘回归模型	224
6.5 结论	249
参考文献	249
第 7 章 数学规划及其农业应用	250
7.1 线性规划及其农业应用	250
7.2 非线性规划及其农业应用	264
7.3 基于实数编码的加速遗传算法的多维动态规划法在推求水稻非充分灌溉下优化灌溉制度中的应用	268
7.4 加速遗传算法在小流域水土保持多目标综合优化模型中的应用	276
参考文献	282
第 8 章 投影寻踪模型及其应用	283
8.1 投影寻踪简介	283
8.2 投影寻踪回归模型及其应用	291
8.3 投影寻踪聚类模型及其应用	297
参考文献	323
第 9 章 物元模型及其在农业系统评价中的应用	326
9.1 物元模型原理	326
9.2 物元模型的农业应用	328
9.3 物元模型的耦合分析及其应用	333
参考文献	344
第 10 章 集对分析与粗糙集理论在农业中的应用	345
10.1 集对分析基本原理	345
10.2 集对分析的农业应用	349
10.3 粗糙集理论基本原理	357
10.4 粗糙集理论及其应用	362
参考文献	369

第 11 章 混沌分析及应用	370
11.1 混沌分析基本原理	370
11.2 混沌分析应用	385
参考文献	398
第 12 章 分维与分形理论及其应用	400
12.1 分形与分维简介	400
12.2 分维与分形应用实例	405
参考文献	428
第 13 章 小波分析理论及其应用	430
13.1 小波分析的发展历史	430
13.2 小波分析的基本理论	430
13.3 小波分析的应用	433
参考文献	455
第 14 章 其他数据处理方法及其农业应用	456
14.1 属性识别模型及其应用	456
14.2 风险决策分析方法及其应用	460
14.3 熵权系数法	466
14.4 逻辑斯蒂曲线模型	469
14.5 工程经济风险分析的蒙特卡洛法	473
14.6 Shepard 插值模型	478
14.7 改进的层次分析法	483
参考文献	491

第1章 遗传算法

农业水土资源大系统中存在诸多随机、模糊、不确定性的复杂问题, 这些问题往往是多维、多峰的复杂函数, 寻求其最优化结果是我们所需要的. 传统的优化方法处理多变量并同时寻优时, 往往易陷入局部最优、早熟或提前收敛, 寻求不到真正的最优解, 这样不利于实际问题的解决. 优化问题贯穿于农业系统分析的全过程, 为此, 这里先介绍一种新的优化方法——遗传算法 (genetic algorithm, GA).

1.1 遗传算法简介

1.1.1 遗传算法原理

生物进化过程本质上就是生物群体在其生存环境约束下通过个体的竞争 (competition)、自然选择 (selection)、杂交 (crossover)、变异 (mutation) 等方式所进行的“适者生存, 不适者淘汰”的一种自然优化过程. 因此, 生物进化的过程, 实际上可以认为是某种优化问题的求解过程. 遗传算法是由美国密歇根大学 John Holland 教授于 1962 年提出的^[1,2], 该方法按照类似有机体的自然选择 (selection) 和杂交 (crossover) 的自然进化 (natural evolution) 方式, 借助计算机程序可以有效地解决较复杂的非线性组合问题及多目标函数优化问题. 遗传算法 (genetic algorithm) 正是模拟生物的这种自然选择和群体遗传机制的数值优化方法. 它把一组随机生成的可行解作为父代群体, 把适应度函数 (目标函数或它的某种变形) 作为父代个体适应环境能力的度量, 经选择、杂交生成子代个体, 后者再经变异, 优胜劣汰, 如此反复进化迭代, 使个体的适应能力不断提高, 优秀个体不断向最优点逼近^[1~5].

下面是标准遗传算法 (SGA) 的计算原理.

设模型的参数优化问题为

$$\begin{aligned} \min f &= \sum_{i=1}^m \|F(C, X_i) - Y_i\|^q, \\ \text{s.t. } &a_j \leq c_j \leq b_j, \quad j = 1, 2, \dots, p. \end{aligned} \quad (1.1)$$

式中, $C = [c_j]$ 为模型 p 个待优化参数 (优化变量); $[a_j, b_j]$ 为 c_j 的初始变化区间 (搜索区间); X 为模型 M 维输入向量; Y 为模型 M 维输出向量; F 为一般非线性模型, 即 $F: R^N \rightarrow R^M; \{(X_i, Y_i) | i = 1, 2, \dots, m\}$ 为模型输入、输出 m 对观测数据;

$\| \|$ 为取范数; q 为实常数, 如当 $q = 1$ 时为最小一乘准则, $q = 2$ 时为最小二乘准则, $\dots\dots$, 可视建模要求而定; f 为优化准则函数。

上述遗传算法中, 当两点交叉改为单点交叉时, 就成为简单遗传算法 (simple genetic algorithm, SGA), 又称标准遗传算法^[1~5]。

SGA 计算主要是由基因编码、产生初始群体、评价个体优劣、选择、杂交、变异五个系列演变过程组成, 其核心技术包括两方面内容: 一是选择方法, 选出的解应具有良好的特征或适应值, 以便产生优良的后代, 同时解空间应相当分散, 以保证求得全局最优解; 二是遗传算子应具备良好的计算特征, 既要保留原有解的优良特性, 又要有恢复已丢失的重要信息或优良特征的功能。

1.1.2 遗传算法的特性

以水资源系统分析为例, 由于受水文、气象、地形、地质条件的综合影响, 水资源工程中的优化问题常表现出多维、多峰值、非连续性等复杂特征。这些复杂特征具体表现为: ① 水资源工程模型的不确定性; ② 水资源工程模型的高维、非正态、非线性; ③ 水资源工程系统庞杂的信息类型。

对上述这些问题目前尚无一种行之有效的优化算法, 传统的方法大致可归纳为确定性优化方法和随机优化方法 (也称随机搜索法) 两类。

一类确定性优化方法属于单路径寻优, 对复杂的非线性优化问题的寻优效率很低。另一类确定性优化方法就是枚举法, 包括完全枚举法、隐式枚举法 (分枝定界法)、动态规划法等。它们的主要缺点是存在“维数灾”问题, 搜索效率不高。随机优化方法算法中, 每一尝试点需要求 n 个随机数, 可见随机性优化方法是通过随机变量的大量抽样, 以得到目标函数的变化特性, 然后逐渐得到近似最优点。该类方法只要求目标函数和约束条件是可计算的, 寻优范围大, 不会陷入局部最优点, 但属“盲目”寻优, 计算量大, 搜索效率低。

由此可见, 传统的优化方法尚无法满足许多复杂水问题的要求。实际中经常遇到的优化问题使人们逐渐认识到, 用某种优化方法寻求最优点不是唯一的目的, 更重要的目的往往是不断改进解的过程, 对于复杂的优化问题更是如此。

遗传算法本身也是一类随机优化方法, 但与传统的基于梯度的确定性优化方法相比, 克服了因线性引起的不稳定性以及依赖于初始点的选择而易陷入局部极小点等缺点, 并且它本身是一类全局寻优方法, 不需计算目标函数的偏导数, 其定义域可任意设定, 只要求对于输入可计算出加以比较的正的输出即可。与传统的优化方法相比, 遗传算法的每步搜索都要利用已有寻优信息来指导解空间的搜索, 它把搜索到的优秀信息遗传到下一代, 而把劣点予以淘汰, 因而是一类自适应优化方法; 遗传算法在运行过程中保持多个当前解, 这样不仅使近似解的优化程度有所提高, 同时

也使得并行计算容易进行,且可获得近似加速的效果.遗传算法与传统优化方法关于寻优表现出较好的稳健性.也就是说 GA 是一种理想的鲁棒优化方法.

归纳起来,遗传算法有如下显著特性^[1~6]:

(1) 适应性强. GA 只要求优化问题是可行的,对搜索空间没有任何特殊要求,可以是离散的、非线性的、多峰值的或高维的、带噪声的.在算法运行中只利用了目标函数值信息,没有利用导数等其他信息.它与所示的问题的性质无关.

(2) 全局优化. GA 是多点、多路径搜索寻优,且各路径之间有信息交换,而不是单点、单路径“登山”.它同时从一代个体点群开始并行攀登多峰,并通过杂交算子在各个可行解之间交换信息,这使得它可以有效地在整个解空间寻优,能以很大的概率找到全局最优解或准全局最优解,即使在所定义的适应度函数是不连续的、非规则的或有噪声的情况下.因此,GA 是一类稳健的全局优化方法.

(3) 编码特征. GA 通过编码将优化变量转换成与基因类似的数字编码串结构,遗传信息储存在其中,可进行各种遗传操作,相应地有解码过程. GA 的操作对象就是这些数字编码串,而不是变量本身,且编码技术在 GA 中一般是不变的,基于编码机制的 GA 用简单的杂交算子、变异算子等模拟了人类探索和发明创造等思维过程中存在的信息交换、渗透和激励机制,从而可以方便地处理离散性问题和连续性问题.

(4) 概率搜索. GA 在选择操作时,用概率规则而不是确定性规则来引导搜索过程向适应度函数值逐步改善的搜索区域方向发展,这就克服了传统随机性优化方法的盲目性,只需较少的计算量就能找到问题的全局近似解,在杂交、变异操作过程中也是采用随机方式进行的.由于 GA 使用概率规则指导搜索,因此能搜索离散的、有噪声的或多峰的复杂空间.

(5) 隐含并行性. GA 通过控制群体中 n 个串,实际上能反映出 $o(n^3)$ 阶个图式,这个性能使 GA 能利用较少的数字串来搜索可行域中的大量区域,从而只花较少的代价就能找到问题的全局解. GA 这种隐含并行性是它优于其他优化算法最主要的因素,因此特别适合于处理复杂的优化问题.

(6) 自适应性. GA 具有潜在的学习能力,能把注意力集中在解空间中适应函数值最高的部分,挖掘出目标区域,因此它适用于具有自适应与学习能力的系统.

(7) 应用的广泛性. GA 兼有确定性优化方法与随机性优化方法的长处,只要求目标函数和约束条件具有可计算性,不要求梯度存在,因此它的适应范围很广.与传统的非线性方法相比,GA 利用选择、杂交、变异操作有可能在更加广阔的范围内寻找问题的潜在解,故它适于处理各类非线性问题,并能有效地解决传统方法难以解决的某些复杂问题.

(8) 算法的简单性和通用性. GA 易于写出一个通用算法,以求解许多不同的优

化问题, 它只需做很小的修改即可适应新的问题。

遗传算法本质上是一种智能优化方法, 直接面向优化问题, 与传统的优化方法相比, 它具有一系列优点, 其结果是一组好的解而不是单个解, 这为解的使用者提供了可选择的机会, 所以它特别适合于处理复杂的非线性优化问题; 对于一个具体问题, 只需选择或编写一种具体的遗传算法方案, 按待求问题的目标函数定义一个适应度函数, 然后就可以用 GA 来求解了, 而不管实际问题的解空间是否连续、线性或可导; GA 还有全局优化的能力。这一系列的优化特征是 GA 在优化问题中能广泛应用的理论依据。

1.1.3 遗传算法的研究动态

目前, 遗传算法在许多领域中得到了重视。同时, 一部分学者也认识到求解复杂问题最优解是不现实的, 转而求其满意解, 而在这方面遗传算法也是最佳的工具之一。在农业大系统中, 优化准则日益成为人们分析系统、评价系统、改善系统和利用系统的一种衡量标准。由于土壤、植被、水文、气候、气象、地质条件、经济、人文等因素的影响及其相互之间的作用, 使得农业系统工程中的优化问题常常具有高维、多峰值、非线性、不连续、非凸性、带噪声等复杂特征。对于这些复杂问题的求解最优程度将影响到农业科学理论的深入发展, 同时影响到农业科学理论转化为实践的程度。

近年来, 农业工程界的国内外专家学者已经在农业水土资源模型参数识别、水土资源环境优化问题、农业自然灾害预测与分析、灌溉制度优化以及多目标函数优化问题等方面进行了初步探讨, 并取得了一定成果, 这对提高水的利用效率、生产效益, 对农业水资源的可持续利用与发展, 进而维护自然生态平衡都具有深远的意义。

目前遗传算法的研究尚处于初期, 其本身在理论上和方法都尚待完善, 因而在应用上受到较大的限制。例如, 在 SGA 拓扑结构方面还没有指导性的理论; 由于群体的有限性与解空间的高维性之间的矛盾, 使得搜索的全局能力受到极大限制; 不可避免地使 SGA 存在“近亲繁殖”、早熟收敛的缺陷; 对实变量的优化问题不太适合; SGA 通常需要比较长的计算时间等问题。所有这些均需进一步深入研究和改进, 以完善其理论基础, 拓广应用领域^[7~21]。

1.2 改进的遗传算法

1.2.1 遗传算法可行的改进措施

由于遗传算法涉及精度、可靠性、计算时间、探索与开发等诸多问题, 通过改

进遗传算法本身,在某种程度上可以提高遗传算法求解问题的性能.针对各种情况有不同的改进,可克服遗传算法中存在的主要问题.为提高遗传算法的质量,对基本遗传算法可作以下几点改进.

1) 控制参数的设置

SGA 中需要设置的参数主要有编码长度 e , 群体规模 n , 杂交概率 p_c , 变异概率 p_m 等,这些参数的设置对 SGA 的运行性能影响很大.

在 SGA 中这些控制参数是不变的. Schaffer 建议 SGA 的最优参数范围: $n = 20 \sim 30$, $p_c = 0.75 \sim 0.95$, $p_m = 0 \sim 0.05$. 目前常用的范围: $n = 20 \sim 200$, $p_c = 0.5 \sim 1$, $p_m = 0 \sim 0.05$.

目前许多学者认识到这些算法参数需要随 GA 的运行进程而做自适应变化,以使 SGA 具有更好的鲁棒性、全局最优性和寻优效率.例如,根据操作串的适应度值来调整参数 p_c, p_m 的大小; p_c, p_m 随进化迭代次数而变化.

2) 编码方式的改进

编码是 GA 应用中首要问题,也是 GA 理论中的基础.不合适的编码不仅影响 GA 的收敛速度,而且也会极大地影响 GA 的搜索效率,因此在应用 GA 时要认真考虑编码方案.对于具体问题,选择或设计一种便于 GA 求解的编码方法经常需要对问题有深入的了解.

常用的修改方式是:对变异算子,个体的每个分量以完全相同的概率在约束范围内随机取值;对杂交算子,用一点或两点交叉,则杂交点位置需处在各分量之间,而更为常用的杂交方式是采用两个配对个体的线性组合.另外,编码时也必须考虑所要求解的问题的特征,如变量的约束条件,所采用的编码方式,除了必须保证不丢失全局最优解外还应该考虑 GA 的求解效率,并尽量避免产生不可行解,这样可以提高计算速度.除此以外,GA 的编码方式也可根据应用问题的具体环境而作相应的变化,可突破传统的一维数字串编码形式,而采用二维数字矩阵或更高维的数字立方体编码方式.考虑问题的专门知识而设计的编码方式常常比通用的编码方法效率更高.

3) 选择算子的改进

选择算子的操作主要源于生物进化过程中“适者生存、不适者淘汰”的规则.在选择中,适应度值低的个体趋向于淘汰(删除),而适应度值高的个体将趋于被保留(复制),所以选择算子的作用效果是提高了群体的平均适应度值,但同时也可能损失群体的多样性.选择操作在总体上决定着个体向着目标函数值改善的方向前进.选择算子并没有产生新个体,且群体中最佳个体的适应度值也不会得到提高.改进

选择算子的目的是为了有效避免基因的缺失,提高 GA 的全局收敛性和搜索效率.选择操作与编码方式无关,而与适应度函数有关.由于适应度函数的分布特性与具体问题不同,因此一律采用 SGA 的比例选择方式是不恰当的.可以采用适当的适应度函数变换,排序选择方式与适应函数的分布和取值无关,故常被采用.例如,现代数学手册采用计算基于序的评价函数值,然后计算每个个体的累积概率进而进行选择操作,使选择只与个体的序号有关,避开适应度函数的影响.

4) 杂交算子的改进

杂交算子的操作主要源于生物群体内部染色体的信息交换机制,即通过两个父代个体的杂交产生新的个体,杂交产生的子代一般与其父代不同,并且彼此也不相同,每个子代都包含两个父代个体的遗传因子.杂交算子的作用是可以产生新个体,从而检测搜索空间中的新点,它有可能使群体中最佳个体的适应度值有所提高.同时需降低对有效基因的破坏率,以免杂交后的子代反而不如父代的生存适应能力.

变异算子的操作可采用单点、两点、均匀等变异方式.采用变异算子,即先在个体串上随机选取一些位置,然后把这些位置上的值用随机选取的值来替换.此外,还可根据优化问题的领域知识来设计变异算子.

5) 算法终止条件的改进

多数改进方法是基于某种判定标准,判定群体是以已收敛并不再有进化趋势作为终止条件.如根据连续几代个体的平均适应度值之差小于某个较小的正数值 ϵ ,也可根据群体中最佳个体适应度值与平均适应度值之差小于某个极小正值 ϵ ,作为终止条件.但是,由于实际优化问题的复杂性和 GA 本身的运行机理尚不完全清楚,应用中常用经验固定进化迭代次数作为遗传算法的终止条件.

6) 改进父代替换方式

可按一定的比例从父代群体中选择部分最佳个体直接进入下一代个体而成为其一部分.

遗传算法经过了近四十年年的发展,开始逐渐走向成熟,尤其在数值优化领域得到了广泛应用.目前,人们对 SGA 已进行了大量改进,并应用于更广泛的领域.这些改进的 GA 之间及其与 SGA 之间已有很大的差别,甚至与其他的进化算法的界限难以区分.事实上,GA 只是提供了一类基本框架,它是一种算法体系,根据不同观点,针对不同类型的问题,结合不同的算法可以编制出不同的遗传算法.这也是 GA 具有较强生命力的原因之一.

1.2.2 基于实数编码的加速遗传算法

1.2.2.1 RAGA 的计算原理

标准遗传算法的编码采用二进制编码,它所构成的基因是一个二进制编码符号串.编码过程烦琐,且精度受到字符串长度的限制,如要求更高的精度,则不得不以增加字符串的长度为代价,计算量大,其结果使进化过程变得十分缓慢,有时易出现早熟收敛.同时,二进制编码不便于反映所求问题的特定知识,因此不便于开发针对专门问题的遗传运算算子.这里提出了一种改进的、基于实数编码的加速遗传算法,使算法的寻优性能大大地增强,克服了二进制编码的缺点,简称为 RAGA(real coded accelerating genetic algorithm).具体改进方案如下^[1~16]:

(1) 采用实数编码.其优点有:①适合于在遗传算法中表示较大的数;②适合于精度要求较高的遗传算法;③便于搜索较大空间的遗传算法;④便于遗传算法与经典优化方法混合使用;⑤便于设计针对专门问题的遗传算子;⑥便于处理复杂的决策变量约束条件.

(2) 在个体适应度评价时采用了基于序的评价函数,使其不受实际目标值的影响.

(3) 在进化迭代时,把每次遗传操作所产生的子代保存下来,即各种遗传操作是并行进行的,尔后将所有子代统一进行评价,再从中依据适应度值选取与群体总数相同的个体作为下一次进化的父代,因此从整体上看实际搜索的范围比 SGA 广,实现了 GA 的并行计算,这样能尽可能地保证个体的多样性,选出更优越的个体解,并能加速进化时间.

(4) 在应用中发现 SGA 对各种实际优化问题的搜索空间(优化变量的范围空间)的大小变化适应能力较差,计算量大,容易出现早熟收敛现象.而利用 SGA 运行过程中搜索到的优秀个体所囊括的空间来逐步调整优化变量的搜索空间,可使算法的寻优速度大大提高,即加快收敛速度,称之为加速遗传算法(accelerating genetic algorithm, AGA).经过大量实例验证,结果表明 AGA 比 SGA 在收敛速度和全局优化性能方面均有明显的提高.

将以上四种方案综合便形成了改进的、基于实数编码的加速遗传算法(简称 RAGA).

RAGA 的建模步骤如下^[3~6]:

一般优化问题多为如下最小化问题:

$$\begin{aligned} \min f(x), \\ \text{s.t. } a(j) \leq x(j) \leq b(j). \end{aligned} \quad (1.2)$$