

遗传算法 及其在水科学中的应用

GENETIC ALGORITHM AND ITS APPLICATIONS TO WATER SCIENCE

金菊良 丁晶 著

四川大学出版社



国家自然科学基金(编号 49871018)资助项目

Genetic Algorithm and Its Applications to Water Science

遗传算法及其在水科学中的应用

金菊良 丁 晶 著

四川大学出版社

成都

责任编辑：孙康江

封面设计：罗 光

责任印制：吴雨时

图书在版编目（CIP）数据

遗传算法及其在水科学中的应用 / 金菊良，丁 晶著.
成都：四川大学出版社，2000. 8
ISBN 7-5614-1958-9

I. 遗… II. ①金… ②丁… III. 遗传算法-应用-
水科学-研究 IV. TV211. 1

中国版本图书馆 CIP 数据核字（2000）第 40535 号

书 名 遗传算法及其在水科学中的应用

作 者 金菊良，丁 晶著

出 版 四川大学出版社

地 址 成都一环路南一段 24 号（610065）

印 刷 西南民族学院印刷厂

发 行 新华书店经销

开 本 787mm×1092mm 1/16

印 张 11

字 数 260 千字

版 次 2000 年 8 月第 1 版

印 次 2000 年 8 月第 1 次印刷

印 数 0 000~500 册

书 号 ISBN 7-5614-1958-9

定 价 40.00 元

◆读者邮购此书，请与本社发行科
联系。电话：5412526/5414115/

5412212 邮编：610064

◆本社图书如有印装质量问题，请
寄回印刷厂调换。

版权所有◆侵权必究

内容简介

模拟生物界中自然选择和群体遗传机制的遗传算法，能够解决许多常规方法尚无法处理的复杂优化问题，在工程技术、计算机科学、管理科学和社会科学等众多领域中具有非常广泛的应用价值。本书试从工程和技术角度，较为深入系统地探讨了遗传算法的基本理论、运行机制和具体实现技术，研制了标准遗传算法的改进方案——加速遗传算法(AGA)，并把 AGA 系统地应用于水文水资源建模、水环境优化和洪水灾害评估等一系列实际水科学问题中。基本上反映了遗传算法应用于水科学的新发展。根据书中介绍的具体算法和众多实例，读者可以比较方便地掌握遗传算法的基本思想和计算技术，加深对专业问题的理解，进一步提高解决实际问题的能力。

本书除作为高校工科研究生选学教材外，还可供从事水利、环境、灾害、地学和农业等方面的工程技术人员和研究人员阅读参考。

前 言

在当代各门科学技术相互交叉、渗透、融合的过程中,优化已成为系统乃至整个世界发展的趋势和走向。反映到水科学系统中,优化准则日益成为人们分析系统、评价系统、改造系统和利用系统的一种衡量尺度。由于受天文、气候、气象、下垫面、人文等众多因素的综合影响,水科学中的优化问题常常表现出高维、多峰值、非线性、不连续、非凸性、带噪声等等复杂特征。对这些复杂水问题的求解直接影响到水科学理论转化为生产力这一隐形价值的实现及其实现程度,同时也将关系到进一步推动水科学理论的深入发展。但是,许多水问题的复杂程度已超越了传统优化方法的处理能力,以至长期以来始终吸引着众多科学家为之而苦苦探索。而生物界中,一系列具有智能、自组织、自修整的器官通过自然选择和群体遗传机制,在不断产生并在不断进化着,John H. Holland教授等按照类似的机制编制的计算机程序能够解决许多传统优化方法无法处理的复杂优化问题,这类新的优化方法就是遗传算法(GA)。

GA 按不依赖于优化问题本身的方式作用在群体特征串上。GA 搜索可能的特征串空间以找到高适应度值串。为了指导这个搜索,GA 仅用到与在搜索空间检查过的点相联系的适应度函数值。不管求解问题的本身如何,GA 都通过执行同样的、极其简单的选择、杂交和变异三种遗传操作来完成它的搜索。在实际应用中,GA 并不知道问题本身的任何信息,也不了解适应度函数是如何定义的,GA 都能够快速有效地搜索复杂、高度非线性的多维空间,找到最优点。基于这些突出特点,GA 在工程技术、计算机科学、管理科学和社会科学等众多优化领域具有非常广阔的应用前景。

本书是作者近五年来有关研究工作的总结。本书试图从工程和技术角度,较为深入、系统地探讨了 GA 的基本理论和算法的运行机制,提出了对标准遗传算法(SGA)的一种改进方案——加速遗传算法(AGA),并把 AGA 系统地应用于水文水资源建模、水环境优化和洪水灾害评估等一系列水问题中,形成了一种理论研究与应用研究紧密结合的独特体系。本书的主题是 GA 如何在具体水问题中的应用。根据书中介绍的 GA 的具体算法和众多实例,读者可以比较方便地掌握 GA 的基本思想和具体的计算技术,加深对具体水问题的理解,从而能据此解决实际问题。

本书由两部分组成:第1章至第4章组成的理论研究部分和第5章至第7章组成的应用研究部分。第1章详细地介绍了传统优化方法处理水科学复杂优化问题的局限性和 GA 的起源、特点,并概述了 GA 的研究进展,最后提出了本书的主要研究目标。第2

EN17163

章分析了遗传算法的运行过程、基本理论和收敛特性,指出了简单遗传算法在全局优化性能与收敛性方面存在着无法克服的矛盾。第3章归纳了简单遗传算法在应用中表现出来的主要缺点,并结合前人GA研究的成果和作者的认识提出了对简单遗传算法进行全面改进的主要方法;接着,该章着重对标准遗传算法进行了改进,提出了基于二进制编码的加速遗传算法(AGA)、基于实数编码的加速遗传算法(RAGA)以及基于整数编码的遗传算法(IPGA),包括它们的计算原理、算法控制参数的设置技术和算法的理论分析以及算法的测试。该章是本书的核心内容。第4章探讨了AGA与BP网络相结合以改进BP算法这一问题。为此,该章在归纳了BP网络的基本理论之后,详细分析了BP算法的运行过程及实施技术,继而针对BP算法的主要缺点,提出了用AGA改进BP算法的新方法——BP-AGA混合算法,并进行了测试。第5章、第6章和第7章分别是AGA在水文水资源建模问题、水环境优化问题和洪水灾害评估问题中的一系列应用研究,它们汇集了作者近年来的主要研究成果,并力图反映遗传算法应用于水科学的新发展。

本书可供高校有关专业的师生、科研人员和工程技术人员阅读参考。

作为一门新兴的人工智能前沿学科,GA也随着进化论、遗传学、数学和计算机科学的进展而迅速发展着,特别是对GA的一些复杂性能尚待人们去认识,GA的理论基础有待加强,GA的运行规模尚远远小于生物进化的规模。加之作者水平有限,成稿仓促,书中的一些观点和方法可能留有争议或错误,殷切希望同行专家和读者们多予批评指正。

本书参考和引用了国内外许多学者的有关论著,从中受到了莫大的教益。作者在此谨向各位学者表示衷心的感谢!

作者衷心感谢中国水利水电科学研究院李纪人教授,河海大学朱元生教授、刘新仁教授、刘权授教授,四川大学林三益教授、邓育仁教授,中国科学院科技政策与管理科学研究所副研究员魏一鸣博士等所给予的热情指导、支持和帮助。在成书过程中有幸得到国家自然科学基金(编号49871018)、中国博士后科学基金、四川大学高速水力学国家重点实验室开放基金(批准号9904)的大力支持,在此作者谨表示衷心感谢!

作者在此特别感谢四川大学出版社的同志们为本书出版所付出的心血!没有他们的热情帮助以及细致严谨的工作,本书是不会以这样的形式面世的。

作 者

2000年7月于成都

目 录

第 1 章	绪论	1
1.1	水科学优化问题与遗传算法	1
1.2	遗传算法研究进展	9
1.3	本书的目的和内容	21
	参考文献	22
第 2 章	遗传算法分析	25
2.1	遗传算法的运行过程分析	25
2.2	遗传算法的基本理论分析	29
2.3	遗传算法的收敛性分析	32
2.4	小结	34
	参考文献	35
第 3 章	标准遗传算法的改进方案	36
3.1	标准遗传算法的主要缺点及其改进方式	36
3.2	基于二进制编码的加速遗传算法	42
3.3	基于实数编码的加速遗传算法	58
3.4	基于整数编码的遗传算法	61
3.5	小结	65
	参考文献	66
第 4 章	AGA 与 BP 神经网络的结合	68
4.1	BP 神经网络概论	68
4.2	BP 网络分析	76
4.3	BP-AGA 混合算法	84
4.4	小结	88
	参考文献	88

第 5 章	AGA 在水文水资源建模中的应用	91
5.1	水文水资源建模问题	91
5.2	AGA 在门限自回归模型中的应用	93
5.3	AGA 在门限回归模型中的应用	98
5.4	AGA 在水位流量关系曲线优化中的应用	102
5.5	AGA 在水库入库含沙量预报中的应用	104
5.6	AGA 在水库调洪演算中的应用	106
5.7	AGA 在地下水模型优化中的应用	109
5.8	AGA 在水资源工程项目内部收益率估计中的应用	112
5.9	小结	115
	参考文献	116
第 6 章	AGA 在水环境优化中的应用	118
6.1	水环境优化问题	118
6.2	AGA 在水环境质量综合评价中的应用	119
6.3	AGA 在水环境模型参数优化中的应用	122
6.4	AGA 在非线性水环境规划问题中的应用	128
6.5	小结	129
	参考文献	130
第 7 章	AGA 在洪灾评估中的应用	131
7.1	洪水灾害评估问题	131
7.2	AGA 在洪灾危险性分析中的应用	137
7.3	AGA 在洪灾易损性分析中的应用	147
7.4	AGA 在洪水灾情评估中的应用	152
7.5	小结	160
	参考文献	161

Contents

Chapter 1	Introduction	1
1.1	Optimal problems in water science and genetic algorithm	1
1.2	Research advance in genetic algorithm	9
1.3	The aims and contents of the book	21
Chapter 2	Analysis of Genetic Algorithm	25
2.1	The analysis of the run process of genetic	25
2.2	The basic theoretic analysis of genetic algorithm	29
2.3	The convergence analysis of genetic algorithm	32
2.4	Conclusions	34
Chapter 3	Improved Schemes of Simple Genetic Algorithm	36
3.1	Main shortcomings of simple genetic algorithm and their improved ways	36
3.2	Accelerating genetic algorithm based on binary number coding	42
3.3	Accelerating genetic algorithm based on real number coding	58
3.4	Genetic algorithm based on integer coding	61
3.5	Conclusions	65
Chapter 4	Combination of AGA with BP Neural Networks	68
4.1	Introduction of BP neural networks	68
4.2	Analysis of BP neural networks	76
4.3	Hybrid algorithm of BP-AGA	84
4.4	Conclusions	88

Chapter 5	Applications of AGA to Hydrologic Modeling	91
5.1	The problem of hydrologic modeling	91
5.2	Application of AGA to threshold auto-regressive model	93
5.3	Application of AGA to modeling threshold regressive model	98
5.4	Application of AGA to optimal relationship of water level and flow discharge	102
5.5	Application of AGA to forecasting silt charge into reservoir	104
5.6	Application of AGA to calculation of reservoir flood	106
5.7	Application of AGA to optimizing the groundwater models	109
5.8	Application of AGA to estimating internal rate return in hydraulic engineering	112
5.9	Conclusions	115
Chapter 6	Applications of AGA to Water Environment Optimizing	118
6.1	The optimal problems of water environment	118
6.2	Application of AGA to water quality evalution	119
6.3	Application of AGA to optimizing parameters of water environment models	122
6.4	Application of AGA to water quality nonlinear planning	128
6.5	Conclusions	129
Chapter 7	Applications of AGA to Flood Disaster Evaluation	131
7.1	The problem of flood disaster evaluation	131
7.2	Application of AGA to the risk analysis of flood disaster	137
7.3	Application of AGA to the analysis of the vulnerability of flood disaster	147
7.4	Application of AGA to the evalation of flood disaster effect	152
7.5	Conclusions	160

第1章 绪论

1.1 水科学优化问题与遗传算法

从形式上讲, 优化问题(optimal problem)可以描述为: 寻找优化变量各分量的某种取值组合, 使得目标函数在给定约束条件下达到最优或近似最优. 解决这类问题的方法称为(最)优化方法(optimization method). 在当代各门科学技术相互交叉、渗透、融合的过程中, 优化已成为系统乃至整个世界发展的趋势和走向, 反映到水文水资源与水环境系统中, 优化准则日益成为人们分析系统、评价系统、改造系统和利用系统的一种衡量尺度. 由于受天文、气候、气象、下垫面、人文等众多因素的综合影响, 水科学中的优化问题(本书也简称水问题)常常表现出高维、多峰值、非线性、不连续、非凸性、带噪声等复杂特征. 这些复杂特征具体表现在三个方面:

一是水资源工程模型的不确定性. 水资源工程系统一般十分庞大而复杂, 往往难以用精确的数学模型来描述, 现有的建模方法又大多建立在许多假设条件和经验公式之上, 只能对实际水资源工程系统作近似模拟.

二是水资源工程模型的高维高度非线性. 常规优化方法处理线性、低维模型较为成熟, 而对大多数高维高度非线性水资源工程模型往往不能取得满意的结果. 虽然也有一些非线性优化方法可供利用, 但总的来说还很不成熟, 实际应用不多.

三是水资源工程系统庞杂的信息类型. 同一系统中往往既含有大量的确定性信息, 又有众多不确定性信息, 如随机性信息、模糊性信息、灰色信息、混沌性信息, 又如在缺乏水资源或人类活动影响强烈的地区以及在水资源评价和开发利用、水资源工程经济运行和管理中, 往往遇到难以用精确数值来表示的信息, 它们常常以经验性语言、知识或规则的形式出现. 常规的优化方法已无法妥善处理这类信息.

对上述这些复杂水问题的求解直接影响到水科学理论转化为生产力这一隐形价值的实现及其实现程度, 同时也将关系到进一步推动水科学理论的深入发展. 但是, 迄今为止, 尚没有一种对这些复杂水问题都行之有效的优化算法(optimal algorithm).

传统的优化方法, 大致可归结为一类搜索方法, 也就是构造序列 $\{x_n\}$, 使^[1]

$$\lim_{n \rightarrow \infty} f(x_n) = \min_{x \in I^n} f(x). \quad (1-1)$$

当 $f(x)$ 是连续函数时, 有

$$\lim_{n \rightarrow \infty} f(x_n) = f(\lim_{n \rightarrow \infty} x_n) = f(x^*) = \min_{x \in I^n} f(x). \quad (1-2)$$

对有约束问题, 至少当 n 充分大时有 $\{x_n\}$ 在可行域 D 内. 同样有

$$\lim_{n \rightarrow \infty} f(x_n) = f(\lim_{n \rightarrow \infty} x_n) = f(x^*) = \min_{x \in I} f(x). \quad (1-3)$$

构造序列 $\{x_n\}$, 一般用迭代方法(iterative method)(逐步构造法). 其递推公式为

$$x_{n+1} = x_n + \alpha_n P_n, \quad (n=1, 2, \dots, N). \quad (1-4)$$

式中, P_n 为第 n 步的下降方向, α_n 为第 n 步的步长.

各种非线性优化的算法, 大都起源于 P_n 或 α_n 的不同的构造方法.

与初始点有关的优化方法求得的解, 常常是找出在初始点附近的一个极值点来, 至于它是否为全局极值点, 在多数情况下不得而知. 在实际水问题中, 则常希望找到在给定条件下的全局极值点.

求全局极值的方法, 本质上是一种试探性搜索方法. 由于全局极值点 x^* 在可行域 D 中的确切位置事先并不知道, 是需通过构造序列 $\{x_n\}$ 来估计的. 因此必然要求 $\{x_n\}$ 在 D 中分布均匀且有一定的密度. 经典的蒙特卡罗法(Monte Carlo Method)在理论上是能满足这种要求的, 优化问题的维数、几何形状、是否离散等对它影响也不大, 但在实际中则失去使用价值. 这是因为, 在 D 中分布均匀的前提下, 为了提高解的精度, 势必在极值点附近加密投点的同时, 也在整个可行域 D 中盲目地加密了投点, 导致其运算量十分浩大, 该算法的时间复杂性破坏了算法的能行性条件, 因而不合理的. 合理的办法是, 发展一些启发式策略或引入领域知识, 对投点过程给予指导或引导, 即: 在可能出现全局极值点的地方增加投点密度, 而对其它地方只作少量的试探性投点, 特别是对已探明无全局极值点出现的区域不投点, 从而可大大节省计算量, 对该算法的时间复杂性进行了有效压缩, 使该算法可行.

上述传统的优化方法, 可归纳为以下两大类:

一是确定性优化方法.

例如, 用得最广泛的梯度法, 它通过沿目标函数的负梯度方向函数值下降来逐步寻找最小点. 该方法虽然可靠性高, 但要求目标函数的一阶或二阶导数存在, 且和所选取的初始点关系很大, 否则会容易陷入局部最优点. 对于优化变量很多的实际问题, 特别是当目标函数和约束条件具有复杂的非线性(多峰问题)时, 梯度法将变得非常困难和不稳定, 以至无效. 另外, 梯度法在处理离散或离散/连续混合问题时也是十分困难的.

又例如, 在实际工程计算中用得较多的复合形法(complex method), 是单纯形法用于有约束优化问题的推广, 它是用于不等式约束条件下直接搜索的一种确定性优化方法. 在 n 维空间中, 由不处于同一个超平面(超曲面)上的 $(n+1)$ 个点构成的凸多面体是该空间中最简单的几何图形, 称之为单纯形(simplex), 而由 $k(k > (n+1))$ 个顶点构成的图形称之为复合形(complex). 复合形法不必计算目标函数 $f(x)$ 的梯度, 它不是沿着某一方向进行搜索的, 而是对 n 维空间内的 k 个顶点(它们构成一个复合形)上的函数值进行比较,

丢掉其中最差点, 通过反射、延伸、收缩、缩边等算子操作得到一个新点, 从而构成一个新的复合形, 如此反复迭代, 逐步逼近最优点. 复合形法程序简单, 使用方便, 但要求可行域是凸的, 这样才能保证各顶点的“重心”也在可行域内部, 这就大大限制了复合形法的应用范围. 应用复合形法时需对每次求得的“重心点”检验是否满足约束条件, 若不满足(可行域是非凸的)则可作某种摄动, 使迭代得以继续. 复合形法的收敛速度是慢的, 特别是对维数较高或约束条件较多时更是如此.

上述这些确定性优化方法属单路径寻优, 对复杂的非线性优化问题其寻优效率是很低的. 另一类确定性优化方法就是枚举法, 它包括完全枚举法、隐式枚举法(分枝定界法)、动态规划法等, 该类方法的主要缺点是存在“维数灾”问题, 搜索效率不高.

二是随机性优化方法.

又称随机搜索法, 它在问题解空间中随机选定一定数量的点, 从中选优. 设优化问题为

$$\left. \begin{aligned} \min f(x), x \in E^n, \\ g_j(x) \geq 0. \end{aligned} \right\} \quad (1-5)$$

($j=1, 2, \dots, m$)

随机性优化方法的运行过程如下:

步骤 1: 令 $k=0$, F =充分大的正数.

步骤 2: 产生 n 个 $[0, 1]$ 上的均匀随机数 $r_1, r_2, \dots, r_n$. 令

$$x_i^{(k)} = a_i + (b_i - a_i) \cdot r_i, \quad (i=1, 2, \dots, n). \quad (1-6)$$

步骤 3: 若 $g_j(x_i^{(k)}) \geq 0$ ($j=1, 2, \dots, m$), 转步骤 4, 否则令 $k+1$, 转步骤 2.

步骤 4: 计算 $f(x_i^{(k)})$. 若 $f(x_i^{(k)}) < F$, 则令 $F = f(x_i^{(k)})$, $x^* = x_i^{(k)}$, 否则 $k=k+1$, 转步骤 2.

步骤 5: 当 $k=M$ (给定的自然数) 时结束算法的运行, 此时, x^* 即为所求的解.

上述算法中, 每一试验点需要求 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 $f(x)$ 的变化特性, 然后逐渐得到近优点, 这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, 计算量很大, 其搜索效率也是很低的.

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. 例如, 关于水电站的优化运行和管理^[2], 已有的优化方法基本上都出自于数学规划技术, 主要有线性规划、动态规划和非线性规划. 线性规划由于将非线性问题线性化, 误差较大. 非线性规划由于缺乏较成熟算法而很少被使用. 动态规划及其扩展算法广泛用于水库或水电站群的优化调度问题, 但随着状态数目的增加, 它往往伴随着所谓“维数灾”问题, 尤其是当它用于库群联合调度时, “维数灾”问题尤为突出, 以至无法在微机上求解. 为了克服“维数灾”困难, 国内外学者已提出不少方法, 比如增量动态规划方法、离散微分动态规划法、逐次逼近增量动态规划方法、逐步优化算法、分解协调技术等, 这些扩展方法的共同特点是很难保证收敛到全局最优解. 又例如, 马斯京根模型参数最优估计问题, 目前确定该模型参数 x 和 k 的最常用方法是试错法, 它需要大量的试算、作图和主观判断, 从中选取最优值, 而且还受到研究者的经验和判别水准的制约, 因此在多组 x 和 k 都可

接受的情况下, 主观判断不能正确地识别拟合最优的 x 和 k . 另外, 估计 x 和 k 的最小二乘法, 因为 x 和 k 交互作用, 使正规方程变为非线性方程组, 给参数的求解带来一定的困难.

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目的, 更重要的目标往往是解的不断改进的过程, 对于复杂的优化问题更是如此.

1962 年美国 Michigan 大学 John H. Holland 教授发现^[3, 4], 按照类似活的有机体的自然选择(selection)和杂交(crossover)的自然进化(natural evolution)方式, 编制的计算机程序能够解决复杂的优化问题, 这类新的优化方法就是遗传算法(Genetic Algorithm, 简称 GA), 又称基因算法.

本质上, 生物进化过程就是生物群体在其生存环境约束下, 通过各个体的竞争(competition)、自然选择、杂交、变异(mutation)等方式所进行的“物竞天演, 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. GA 正是模拟生物的这种自然选择和群体遗传机理的数值优化方法. 具体说来^[5, 6], GA 把一族随机生成的可行解作为父代群体, 把适应度函数(目标函数或它的一种变换形式)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.

生物进化过程既十分丰富又极为复杂, 有许多方面尚未被人类所认识, 目前已被认识的生物进化过程中的一些基本特征是:

其一, 生物个体的染色体(chromosomes)的结构特征, 即基因码序列(series of genetic code)决定了该个体对其生存环境的适应能力.

其二, 自然选择在生物群体(population)进化过程中起着主导作用, 它决定了群体中那些适应能力(adaptability)强的个体能够生存下来并传宗接代, 体现了“优胜劣汰”的进化规律.

其三, 个体繁殖(杂交)是通过父代个体间交换基因材料来实现的, 生成的子代个体的染色体特征可能与父代的相似, 也可能与父代的有显著差异, 从而有可能改变个体适应环境的能力.

其四, 变异使子代个体的染色体有别于其父代个体的染色体, 从而也改变了子代个体对其环境的适应能力.

其五, 生物的进化过程, 从微观上看是生物个体的染色体特征不断改善的过程, 从宏观上看则是生物个体的适应能力不断提高的过程.

GA 就是用下列计算机技术来模拟上述生物进化特征, 并逐渐发展成为一类新的优化方法^[3, 7, 8, 9]:

不失一般性, 设优化问题为函数极小化问题: $\min f: R^n \rightarrow R$, f 为目标函数; 适应度函数为 $F: I \rightarrow R$, 其中 I 是个体的空间, $I = \{a_i | i = 1, 2, \dots, n\}$; a_i 为个体; $x = \{x_1, x_2, \dots, x_n\}$ 为 R^n 上的目标变量集, $[u_i, v_i]$ 为分量 x_i 的范围; $n_p \geq 1$ 为父代群体规模, $n_c \geq 1$ 为子代群体规模, 即在每一代通过杂交、变异产生的个体数目; 在第 t 次进化迭代时, 群体为 $P(t) = \{a_1(t), a_2(t), \dots, a_{n_p}(t)\}$; $C: I^{n_p} \rightarrow I^{n_c}$ 为杂交算子, 其控制参数为杂交概率 p_c ; $M: I^{n_c} \rightarrow I^{n_c}$ 为变异算子, 其控制参数为变异概率 p_m ; $S: (I^{n_c} \cup I^{n_p}) \rightarrow I^{n_p}$ 为选择算子, 用

于产生下一代父代群体, 其控制参数为选择概率 p_s . 这里 C, M, S 均为宏算子, 即把群体变换为群体, 把相应作用在个体上的算子分别记为 c, m, s. GA 的算法如下:

步骤 1: 用二进制位串(string)来模拟染色体, 用数字串中的某位元模拟基因(gene), 用数量极大的数字串表示优化问题可行域内的试探解. 这种将优化问题的解表示为数字串的过程称为编码(encoding). 每个串代表一个个体(individual), 许多个体的集合称为群体(population). GA 作用于确定长度 e 的二进制位串上, 即 $I=\{0, 1\}^e$. 用一段位串($a_{i1}, a_{i2}, \dots, a_{ie}$)通过如下位段解码函数来表示优化变量的第 i 个分量:

$$x_i = (a_{i1}, a_{i2}, \dots, a_{ie}) = u_i + (v_i - u_i) \left(\sum_{j=1}^e a_{ij} \cdot 2^{j-1} \right) / (2^e - 1), \quad (i=1, 2, \dots, n). \quad (1-7)$$

式中, ($a_{i1}, a_{i2}, \dots, a_{ie}$) 为个体 $a=(a_{i1}, \dots, a_{ie}, a_{21}, \dots, a_{ne}) \in I^e$ 的第 i 段位串.

遗传算法的操作对象是数字串, 而不是优化问题的解形式, 这为遗传算法成为一种通用的优化方法奠定了基础. 若把所有可行解所对应的二进制串比作某一地区, 则那些编码不良答案的串就是该地区的峡谷, 而那些编码优良答案的串就对应山峰. 在该地区搜索寻优的常用方法是登山, 即从某点开始, 若一个小变动能改善解的质量, 则沿该方向继续前进, 否则取其它方向搜索. 但是, 复杂问题会使该地区山峰林立, 随着解空间维数的增加, 该地区的拓扑结构也趋复杂化, 此时寻找正确的山峰、甚至判断哪条路径是向上的都会变得越来越困难. 另外, 这样的搜索空间极大. 例如, 有一种棋, 每步平均有 10 种走法, 每盘棋每方一般走 30 步, 则这种棋共有 10^{60} 种走法, 但是其中大多数走法的效果并不好.

步骤 2: 用目标函数或其变换作为适应度函数(fitness function)来模拟个体对生存环境的适应性. 把上述 n 个位段解码函数组合成一个个体的解码函数, 即 $= 1^2 \dots n$, 则个体 a 的适应度函数可设置为 $F(a)=g(f(a))$, 其中 g 为对目标函数进行变换的函数, 其作用是确保适应度值为非负并且好的个体的适应度值也大, 常见的 g 有线性比例函数、幂比例函数和指数比例函数等^[9].

适应度函数用来计算群体中的每个串的适应度值, 适应度函数值越大的个体被选择的概率越大, 而被淘汰的概率越小, 反之亦然. 个体 a_i 被选中的概率为

$$p(a_i) = F(a_i) / \sum_{j=1}^{np} F(a_j). \quad (1-8)$$

并按照上式概率分布选取 np 个个体作为当代父代群体. GA 象一张网, 罩在优化问题可行域上. 数量极大的位元串组成了一个不断进化的群体, 这些串同时在这一地区的很多区域中进行采样, 且不同区域中的采样频率正比于该区域的平均适应度值.

步骤 3: 父代个体的杂交. 设杂交概率为 p_c , 则当代父代群体中有 $np \cdot p_c$ 个个体进行杂交. 把两个父代数字串配对排列, 沿串的长度方向随机选取一点, 交换两串的左边部分, 产生两个子代个体. 其中一个子代个体含第一个串在交换点以左的符号和第二个串在交换点以右的符号, 另一个子代个体含第二个串在交换点以左的符号和第一个串在交换点以右的符号. 设两父代个体 $a_1=(a_{11}^1, a_{12}^1, \dots, a_{1ne}^1)$ 和 $a_2=(a_{21}^2, a_{22}^2, \dots, a_{2ne}^2)$ 随机地从群体中配对、杂交, 产生两个子代个体

$$b_1=(a_{11}^1, a_{12}^1, \dots, a_{ih}^1, a_{ih+1}^2, a_{ih+2}^2, \dots, a_{ne}^2), \quad (1-9)$$

$$b_2=(a_{11}^2, a_{12}^2, \dots, a_{ih}^2, |a_{ih+1}^1, a_{ih+2}^1, \dots, a_{ne}^1). \quad (1-10)$$

式中, 杂交点 ih 为随机选取. 这种杂交算子称为单点杂交算子, 参见图 1-1. 由此扩展, m 点杂交算子就是随机选取 m 个杂交点, 并交替地互换配对双亲各杂交点之间的位段而生成两个子代个体. 目前尚无明确的理论和可靠的试验证据来判定何种杂交算子是最适宜的, 尽管已有一些试图解决此问题的研究成果.

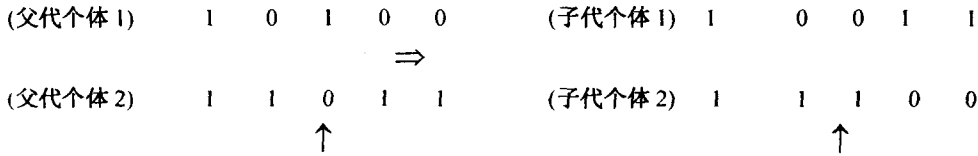


图 1-1 单点杂交算子

遗传算法的杂交算子实际上完全模仿了生物模型. 所产生的子代个体并不取代适应度值高的父代串, 而是取代两个适应度值低的串. 每产生新一代, 都将淘汰适应度值低的串, 以使整个群体的规模 (大小) 保持不变.

步骤 4: 子代数字串变异. 在 GA 中的变异算子是一种辅助算子, 它作用在子代个体位串上, 以小概率 p_m 随机地改变串上的每一位 (即相应位上的 0 变为 1, 或是 1 变为 0), p_m 一般取 0.001 到 0.01 之间^[7]. 变异不依赖于优化变量的维数和位串的总长. 设子代个体为 $a_i=(a'_{11}, a'_{12}, \dots, a'_{ne})$, 经变异算子 $m(a_i)$ 作用后变为 $a_i^1=(a_{11}, a_{12}, \dots, a_{ne})$:

$$a_{jk}^1 = \begin{cases} a'_{jk}, & u_{jk} > p_m, \\ |a'_{jk} - 1|, & u_{jk} \leq 0. \end{cases} \quad (1-11)$$

式中, u_{jk} 为 $[0, 1]$ 上的均匀随机数, $| \cdot |$ 为取绝对值函数, 参见图 1-2. 从优化的角度看, 单靠变异一般不能在求解的过程中取得进展, 但它能确保不产生不能继续进化的单一群体. 杂交与变异增加了产生新的数字串的可能性, 从而使遗传算法有能力搜索那些未被采样的区域.

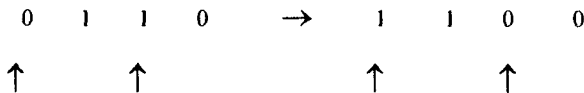


图 1-2 两点变异算子

步骤 5: 进化迭代. 转步骤 2, 如此反复运行, 直至获得满意的解或达到预定进化迭代次数, 结束算法的运行, 并把适应度值最高的数字串所表示的解作为解优化问题的结果. 随着一代代的选择、杂交和变异, GA 能在解空间中发掘出期望值高的区域, 在这些区域中串的数目在不断增加. 表 1-1 为生物进化过程与 GA 的对照表.

表 1-1 生物进化过程与遗传算法的对照表

生物进化	遗传算法
基因(gene)	字符(character)
基因值(等位基因)(allele)	字符值(character value)
基因位置(基因座)(locus)	数串位置(string position)
基因组(genome)	图式(schema)
染色体(chromosome)	数串(string)
基因总集(基因型)(genotype)	数串空间(string space)
表现型(phenotype)	解空间(solution space)
个体适应能力(adaptability)	适应度函数(fitness function)
选择(selection)	选择算子(selection operator)
杂交(crossover)	杂交算子(crossover operator)
变异(mutation)	变异算子(mutation operator)
生物个体的进化过程(process of evolution)	问题解的优化过程(process of optimization)

表 1-1 中, 表现型为基因型与环境的约束之交集。

可见, 遗传算法就是在“优胜劣汰”指导下的一类随机并行自适应优化方法, 所编的码串相当于某群体的个体, 目标函数及解变量的约束条件相当于个体所处的环境, GA 的运行过程就是基于个体与环境的作用进行的。

实际上, 遗传算法可视为介于确定性优化方法与随机性优化方法之间的一类新的优化方法。其确定性成份表现在, 每次选择操作时应用了“优胜劣汰”这一生物进化法则, 也即, 那些适应度值越高的个体越有可能“遗传”到下一轮进化迭代过程; 其随机性成份则表现在选择、杂交、变异操作时都具有一定的随机性, 这正如生物进化过程中的基因遗传、变异和突变也具有随机性一样。

虽然遗传算法本身也是一类随机优化方法, 但它与传统的基于梯度的确定性优化方法相比较, 克服了因线性化引起的不稳定性以及依赖于初始点选择而易限于局部极小点等缺点, 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。它与传统的随机优化方法相比较, 由于 GA 的每步搜索都要充分利用已有寻优信息来指导解空间的搜索, 它把搜索到的优秀点的信息遗传到下一代, 而把劣点予以淘汰, 因而它是一类自适应优化方法; 而传统的蒙特卡罗方法, 搜索完一个点后, 不管该点好坏, 都不保留任何关于该点的信息, 因而降低了搜索效率。此外, GA 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易行, 且可获得近似线性加速的效果。

遗传算法与传统优化方法关于寻优稳健性 (robustness) 的比较参见图 1-3。GA 就是一类理想的稳健优化方法。

归纳起来, 与常规的优化方法相比, 遗传算法具有如下显著特点^[8-12]:

一是适应性强: GA 只要求优化问题是可计算的, 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, 没有利用导数等其它的辅助信息。它与所求解问题的性质无关。