



计算方法丛书

非数值并行算法

(第二册)

遗传算法

刘勇 康立山 陈毓屏 著



科学出版社

计算方法丛书

非数值并行算法(第二册)

遗传算法

刘勇 康立山 陈毓屏 著

科学出版社

1995

(京)新登字 092 号

内 容 简 介

本书系统地叙述了非数值并行算法之一的遗传算法的基本原理以及最新进展,同时为了便于读者解决实际问题,书中对具体算法的步骤作了详细介绍。本书共分七章,第一章介绍算法的思想、特点、发展过程和前景。第二章介绍算法的基本理论。第三章讨论算法解连续优化问题。第四章利用算法设计和优化神经网络。第五章介绍在组合优化中的应用。第六章介绍应用遗传程序设计解决程序设计自动化问题。第七章对遗传算法和其它适应性算法进行比较。

本书可供高校有关专业的师生、科研人员、工程技术人员阅读参考。

计算方法丛书 非数值并行算法(第二册) 遗传算法

刘 勇 康立山 陈毓屏 著

责任编辑 林 鹏 徐宇星

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码:100717

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

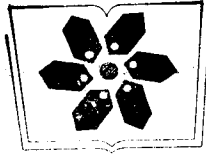
1995 年 1 月第 一 版 开本:850×1168 1/32

1995 年 1 月第一次印刷 印张:7 1/8

印数:1—1 200 字数:181 000

ISBN 7-03-004345-6/O·752

定价:10.60 元



中国科学院科学出版基金资助出版



国家自然科学基金委员会资助出版

《计算方法丛书》编委会

主 编 **冯 康**

副主编	石钟慈	李岳生		
编 委	王仁宏	王汝权	孙继广	李德元
	李庆扬	吴文达	林 群	周毓麟
	席少霖	徐利治	郭本瑜	袁兆鼎
	黄鸿慈	蒋尔雄	雷晋干	滕振寰

序 言

当今计算机科学的各个领域的发展几乎都显示出向并行计算过渡这一趋势,人们开始从并行和分布式处理的观点重新探索计算理论、计算机语言、操作系统、数据库、结构以及应用。本书呈现给读者的正是反映在这场变革中迅速发展的一个全新的科学领域——遗传算法(genetic algorithms),其发展从一开始就是基于并行处理。遗传算法是一种借鉴生物界自然选择和自然遗传机制的高度并行、随机、自适应搜索算法,它主要用在处理最优化问题和机器学习。隐含并行性和对全局信息的有效利用能力是遗传算法的两大显著特点,前者使遗传算法只须检测少量的结构就能反映搜索空间的大量区域,后者使遗传算法具有稳健性(robustness)。遗传算法尤其适于处理传统搜索方法解决不了的复杂和非线性问题。

从1985年到1993年,召开了五届国际遗传算法学术会议,遗传算法已经有了很大的发展,并开始渗透到人工智能、神经网络、机器人和运筹学等领域。遗传算法是多学科相互结合与渗透的产物,它已发展成一种自组织、自适应的综合技术,广泛用在计算机科学、工程技术、管理科学和社会科学等领域。

本书汇集了武汉大学并行计算研究室的科研成果和国外一些有关最新的文献资料,力图反映遗传算法的新发展。我们对遗传算法的基本原理作了较为系统、全面的论述,同时考虑到让读者阅读后能解决实际问题,也对一些具体算法步骤作了较详细的介绍。我们希望本书能吸引更多的科技人才投入到这个研究领域中来。

本书共分七章。第一章绪论中包括了遗传算法的思想、描述、特点、发展过程和前景。第二章论述了遗传算法的基本理论。第三章讨论了遗传算法解连续优化问题。第四章利用遗传算法设计

和优化神经网络.第五章介绍了遗传算法在组合优化中的应用.第六章应用遗传程序设计解决程序设计自动化中的问题.第七章对遗传算法和其它的适应性算法进行了比较.

限于著者的学识水平,书中不妥之处在所难免,非常感谢读者指正.

著者感谢武汉大学软件工程国家重点实验室并行计算研究室的同仁所给予的热情支持和帮助.著者从事遗传算法的研究工作得到了国家自然科学基金、国家 863 高技术计划和国防科工委八五预研项目的资助,本书的出版还得到国家自然科学基金委员会优秀成果专著出版基金与中国科学院科学出版基金的资助,在此谨表示衷心感谢.

刘 勇 康立山 陈毓屏

1993 年 12 月 于武汉大学

目 录

序言

第一章 绪论	1
§ 1.1 自然进化与遗传算法.....	1
§ 1.2 遗传算法的描述.....	4
§ 1.3 表示方案的实例.....	14
1.3.1 工程设计的最优化.....	14
1.3.2 人工蚁问题.....	17
§ 1.4 遗传算法的特点.....	22
§ 1.5 遗传算法的发展简史.....	25
§ 1.6 遗传算法的研究内容及其前景.....	33
第二章 遗传算法的数学理论	38
§ 2.1 遗传算法的基本定理.....	38
§ 2.2 隐含并行性.....	46
§ 2.3 基因块假设.....	48
§ 2.4 最小欺骗问题.....	53
§ 2.5 遗传算法欺骗问题的分析与设计.....	60
§ 2.6 模式的几何表示.....	62
§ 2.7 遗传算法收敛性分析.....	63
2.7.1 基本定义.....	64
2.7.2 守恒杂交算子.....	67
2.7.3 完全变异算子.....	69
2.7.4 遗传算法的马尔柯夫链分析.....	70
第三章 解连续优化问题的遗传算法	75
§ 3.1 基本的遗传算法.....	75
3.1.1 引言.....	75
3.1.2 算法描述.....	76

3.1.3	算法性能分析	78
3.1.4	从目标函数到适应函数	82
3.1.5	基本的选择方法	85
§ 3.2	遗传算法中控制参数的最优化	86
3.2.1	自适应系统模型	86
3.2.2	试验设计	87
3.2.3	试验结果	92
§ 3.3	适应值的比例变换	96
3.3.1	基本的比例方法	96
3.3.2	用于选择比例函数的准则的性质	99
3.3.3	比例函数的一维族	101
3.3.4	比例函数的 m 维族	104
§ 3.4	解函数优化的并行遗传算法	108
3.4.1	遗传算法与并行计算机	108
3.4.2	并行搜索和最优化	110
3.4.3	并行遗传算法的形式描述	110
3.4.4	性能评估	115
3.4.5	数值结果	117
3.4.6	超线性加速比	118
3.4.7	PGA 与一般最优化方法	120
§ 3.5	混合遗传算法	121
3.5.1	混合的原则	122
3.5.2	修改的遗传算子	123
§ 3.6	退火演化算法	124
3.6.1	模拟退火算法概述	124
3.6.2	退火演化算法用于求解连续优化问题	126
3.6.3	比较结果及退火演化算法的并行策略	128
§ 3.7	约束最优化问题	132
第四章	用遗传算法设计神经网络	137
§ 4.1	神经网络概述	137
§ 4.2	感知机结构的设计	140
4.2.1	感知机模型及其学习算法	140

4.2.2	神经网络设计与遗传算法	142
4.2.3	感知机的遗传表示	144
4.2.4	演化过程	145
4.2.5	试验设计	146
§ 4.3	前馈神经网络的设计	150
4.3.1	反向传播法	150
4.3.2	混合学习系统	152
4.3.3	试验结果和结论	154
第五章	遗传算法在组合优化中的应用	157
§ 5.1	基于有序的遗传算法和图着色问题	157
5.1.1	图着色问题	157
5.1.2	基于有序的和遗传算子	159
5.1.3	图着色问题的实例	161
§ 5.2	解货郎担问题的遗传算法	165
5.2.1	货郎担问题与几个常用的遗传算子	165
5.2.2	算法描述	168
5.2.3	货郎担问题的计算结果	171
§ 5.3	解映射问题的并行遗传算法	172
5.3.1	引言	172
5.3.2	遗传表示和并行策略	174
5.3.3	并行遗传算法的执行分析	175
第六章	遗传程序设计与程序设计自动化	178
§ 6.1	引言	178
§ 6.2	遗传程序设计的主要步骤	179
§ 6.3	遗传程序设计的具体描述	182
6.3.1	函数集和端点集	182
6.3.2	初始结构	184
6.3.3	适应值度量	186
6.3.4	主要操作	188
6.3.5	控制参数	191
§ 6.4	解人工蚁问题的遗传程序设计	191
第七章	遗传算法与其它自适应搜索方法的比较	198

§ 7.1 引言	198
§ 7.2 四种自适应搜索方法的比较	200
§ 7.3 结束语	202
附录 Grötschel 货郎担问题的顶点坐标	204
参考文献	206

第一章 绪 论

§ 1.1 自然进化与遗传算法

从本世纪 40 年代,生物模拟就成为了计算科学的一个组成部分,如早期的自动机理论就是假设机器是由类似于神经元的基本元素组成的。这些年来,诸如机器能否思维、基于原则的专家系统是否胜任人类的工作、以及神经网络可否使机器具有看和听的能力等有关生物类比的问题已成为人工智能关注的焦点。最近生物计算在机器昆虫和种群动态系统模拟上所取得的成功激励越来越多的人致力于人工生命领域的研究。当今计算机科学家和分子生物学家已开始携手进行研究,并且类比也得到了更广泛的应用。

计算和生物学之间的类推更为一致,基因和计算机都记录、复制和传播信息。美国 Indiana 大学的 Hofstadter 通过论证明确地指出,在活细胞的繁殖中 DNA 和 RNA 的行为可以解释为自复制 Turing 机的一个实例。

但所有这些模拟都赶不上遗传算法所取得的成功。遗传算法是一族通过模拟自然进化过程搜索最优解的方法。

自从生物变化的进化理论得到人们的接受之后,生物学家就对进化机制产生了极大的兴趣。化石记录表明我们所观察到的复杂结构的生命是在相对短的时间内进化而来的,对这一点包括生物学家在内的许多人都感到惊奇。

虽然目前关于推动这个进化的机制还没有完全弄清楚,但它们的某些特征已为人所知。进化是发生在作为生物体结构编码的染色体上,通过对染色体的译码部分地生成生物体。人们现在还不完全清楚染色体的编码和译码过程的细节,但下面几个关于进化理论的一般特性已广为人们所接受:

(1) 进化过程是发生在染色体上，而不是发生在它们所编码的生物体上。

(2) 自然选择把染色体以及由它们所译成的结构的表现联系在一起，那些适应性好的个体的染色体经常比差的个体的染色体有更多的繁殖机会。

(3) 繁殖过程是进化发生的那一刻。变异可以使生物体子代的染色体不同于它们父代的染色体。通过结合两个父代染色体中的物质，重组过程可以在子代中产生有很大差异的染色体。

(4) 生物进化没有记忆。有关产生个体的信息包含在个体所携带的染色体的集合以及染色体编码的结构之中，这些个体会很好地适应它们的环境。

自然界生物体通过自身的演化就能使问题得到完美的解决，这种才能让最好的计算机程序也相形见绌。计算机科学家为了某个算法可能要耗费数月甚至几年的努力，而生物体通过演化和自然选择这种非定向机制就达到了这个目的。

大多数生物体是通过自然选择和有性生殖这两种基本过程进行演化的。自然选择决定了群体中哪些个体能够存活并繁殖；有性生殖保证了后代基因中的混合和重组。比起那些仅包含单个亲本的基因拷贝和依靠偶然的变异来改进的后代，这种由基因重组产生的后代进化要快得多。自然选择的原则是适应者生存，不适应者淘汰。

自然进化的这些特征早在 60 年代就引起了美国 Michigan 大学的 John Holland 的极大兴趣，那时，他和他的学生们已在从事如何建立能学习的机器的研究。Holland 注意到学习不仅可以通过单个生物体的适应而且通过一个种群的许多代的进化适应也能发生。受达尔文进化论——适者生存的启发，他逐渐认识到，在机器学习的研究中，为获得一个好的学习算法，仅靠单个策略的建立和改进是不够的，还要依赖于一个包含许多候选策略的群体的繁殖。他们的研究想法起源于遗传进化，Holland 就将这个研究领域取名为遗传算法。一直到 1975 年 Holland 出版了那本颇有影

响的专著 *Adaptation in Natural and Artificial Systems*, 遗传算法这个名称才逐渐为人所知。

Holland 创建的遗传算法是一种概率搜索算法,它是利用某种编码技术作用于称为染色体的二进制数串,其基本思想是模拟由这些串组成的群体的进化过程。遗传算法通过有组织地而不是随机地信息交换来重新结合那些适应性好的串,在每一代中,利用上一代串结构中适应性好的位和段来生成一个新的串的群体;作为额外增添,偶尔也要在串结构中尝试用新的位和段来替代原来的部分。遗传算法是一类随机算法,但它不是简单的随机走动,它可以有效地利用已有的信息来搜寻那些有希望改善解质量的串。类似于自然进化,遗传算法通过作用于染色体上的基因,寻找好的染色体来求解问题。与自然界相似,遗传算法对求解问题的本身一无所知,它所需要的仅是对算法所产生的每个染色体进行评价,并基于适应值来选择染色体,使适应性好的染色体比适应性差的染色体有更多的繁殖机会。

遗传算法利用简单的编码技术和繁殖机制来表现复杂的现象,从而解决非常困难的问题。特别是由于它不受搜索空间的限制性假设的约束,不必要求诸如连续性、导数存在和单峰等假设,以及其固有的并行性,遗传算法目前已经在最优化、机器学习和并行处理等领域得到了越来越广泛的应用。

需要说明的是,虽然遗传算法的早期研究从生物进化理论中得到了不少启示,并且生物学家和遗传学家的发现会继续在某种程度上影响这一领域,但是这种影响多半是单向的。至今遗传算法没有应用在遗传学领域中,并且遗传算法的研究也没有对生物学的理论产生影响,在这一点上,遗传算法似乎类似于神经网络和模拟退火算法。这两种算法也是基于对自然界的有效类比。这些算法从自然现象中抽象出来,但研究这些自然现象的科学家们到目前为止还没有受到算法抽象概念的很大影响。经过类比启示的开始阶段后,遗传算法、神经网络以及模拟退火算法已成为沿自身道路发展下去的学科,它们距给它们以启示的学科越来越远。

§ 1.2 遗传算法的描述

本节以一个非常简单最优化问题为例来说明遗传算法。这个例子是为四个连锁饭店寻找最好的经营决策，其中一个经营饭店的决策包括要做出以下三项决定：

- 价格 汉堡包的价格应该定在 50 美分还是 1 美元？
- 饮料 和汉堡包一起供应的应该是酒还是可乐？
- 服务速度 饭店应该提供慢的还是快的服务方式？

目的是找到这三个决定的组合(即经营决策)以产生最高的利润。

因为有三个决定变量，其中每个变量可以假设为两个可能值中的一个，所以对这个问题的每个可能的经营决策可以很自然地用长度 $l = 3$ 、在规模 $k = 2$ 的字母表上的特征串来表示。对于每个决定变量，值 0 或 1 被指定为两个可能选择中的一个。这个问题的搜索空间包括 $2^3 = 8$ 个可能的经营决策。串长($l = 3$)、字母表规模($k = 2$)以及映射组成了对这个问题的表示方案，其中映射把串中具体位上的决定变量规定为 0 或 1。利用遗传算法求解这个问题的第一步就是选取一个适当的表示方案。

按上面描述的表示方案，在 8 个可能的经营决策中表 1.1 给出了其中的 4 个。

表 1.1 饭店问题的表示方案

饭店编号	价格	饮料	速度	二进制表示
1	高	可乐	快	011
2	高	酒	快	001
3	低	可乐	慢	110
4	高	可乐	慢	010

假设这四个饭店的经营决策要由一位没有经验的新手决定，从而他不知道在三个决定变量中哪个是最重要的，也不知道在他

做出最优决策下能得到的最大利润量或者在他做出错误决策下可能招致的损失量，甚至不知道哪个变量的单独改变会产生利润上的最大变化。

新手不知道能否通过下面的逐步调整过程来接近全局最优值，在这个过程中每次改变一个变量，挑选好的结果，然后类似地改变另一个变量，再挑选好的结果。也就是说，他不知道变量能否单独地优化，或者它们是否以高度非线性方式相互联系。

新手面临另外的困难是只有通过每星期各个饭店的赢利情况来获得关于环境的信息。问题是他不清楚影响顾客光顾饭店的确切因素以及每个因素对顾客的决定起作用的程度。在营业过程中所观察到的饭店经营情况只是经营者从环境中得到的反馈，他不能保证经营环境在每个星期都保持不变。顾客的口味是多变的，并且决策的规则可能会突然改变，原来非常好的决策在某个新的环境中可能不再产生同样多的利润。环境的改变不仅可能是突然的，而且是不能预告的。通过观察到当前的经营决策不再产生与以前同样多的利润，经营者才会间接地发现环境的改变情况。

经营者还要面临的是要求立即做出经营决策，没有时间让他有单独的训练或单独的试验，唯一的试验来自实际营业的方式。此外，有用的决策过程必须立即开始产生一连串的中间决策，这些中间决策保持饭店从一开始到后续每个星期都在生存所需的最低水平之上。

因为经营者不了解他所面临的环境，他开始可能会明智地对四个饭店分别采用不同的初始随机决策，可以期望随机决策的获利近似地等于在搜索空间内总体上的平均获利。这种多样性一方面大大增加了获得接近于搜索空间内总体平均利润的机会，另一方面把从第一个星期的实际营业中学到的信息量增加到最大限度。我们将采用表 1.1 中所给出的 4 个不同的决策作为经营决策的初始随机群体。

事实上，饭店经营者是按与遗传算法同样的方式进行决策的。遗传算法的执行开始时是通过检测在搜索空间中随机选取的某些

点来尽量学习关于环境的信息。特别地,遗传算法从第 0 代(初始随机代)开始,初始群体由随机产生的个体组成,在这个例子中,群体规模 N 等于 4。

在遗传算法中,每一代群体中的个体都要在未知环境中进行检测以得到它们的适应值,这里适应值取为利润,它还可以是获利、效用、目标函数值、得分或其它一些值。在这个问题中,初始群体的 4 个个体的适应值由表 1.2 给出,其中适应值简单地定义为每个二进制染色体串所代表的十进制值,所以决策 110 的适应值是 6 美元,全局最优适应值为 7 美元。

表 1.2 初始群体中经营决策的适应值

第 0 代		
i	串 x_i	适应值 $f(x_i)$
1	011	3
2	001	1
3	110	6
4	010	2
总和		12
最小值		1
平均值		3.00
最大值		6

通过检测 4 个随机决策,经营者获悉到什么呢?表面上他知道了搜索空间中被检测的四个特殊点(即决策)的具体适应值(即利润)。特别地,他了解到第 0 代群体中最好的个体 110 每周产生 6 美元的利润,最差的个体 001 每周只产生 1 美元的利润。

在遗传算法中用到的唯一信息是实际出现在群体中个体的适应值。通过模拟生物界自然选择和自然遗传过程,遗传算法把一个群体变换到一个新的群体。一个简单的遗传算法由复制、杂交和变异三个遗传算子组成。

复制算子把当前群体中的个体按与适应值成比例的概率复制