

基因算法及其应用

JIYINSUANFAJIQIYINGYONG

孟庆春 著

山东大学出版社

基因算法及其应用

孟庆春 著

江苏工业学院图书馆
藏书章

山东大学出版社

内容简介

基因算法具有许多传统的优化搜索技术所不具有的优点,是80年代世界上兴起的一种主要算法技术。其最主要的特点是通过模拟生物界的进化过程,从解空间的多点搜索问题的解入手,最终追求解的全局合理性。本书首先介绍了基因算法的发展过程、原理和特点以及在控制工程、人工智能、工程设计、优化技术、计算机技术、模式识别、医学、生物、化学、社会科学等众多领域中的应用,然后概述了门限变换思想和对称编码理论,并且在问题求解和机器人的动力学优化控制及路线规划两个方面,对新算法进行了模拟,得到了远好于常规基因算法的结果。同时,书中还介绍了有关机器人技术方面的研究工作。

本书引用了作者近年来的研究成果和国内外最近的文献资料,可供高等院校和科研机构从事计算机科学、控制工程、人工智能、机器人技术、工程设计、计算数学、统计物理、生物、化学、医学和决策科学等专业的教师、研究人员、研究生和本科毕业生参阅使用。

鲁新登字 09 号

责任编辑 张秉尧

封面设计 于岱利

基因算法及其应用

孟庆春 著

*

山东大学出版社出版发行
山东省泰安市第三印刷厂印刷

*

787×1092毫米 16开本 9.25印张 210千字

1995年8月第1版 1995年8月第1次印刷

印数 1—2000

ISBN7-5607-1504-4

0·94 定价: 17.80元

序 言

本书是我国第一部系统介绍基因算法及其应用的论著。孟庆春博士在书中吸收了他的许多研究成果,并介绍了世界这一领域中的最新成就。

基因算法是 Holland 教授在 70 年代创立的一项优化搜索技术。这是一类模拟生物进化过程和基于统计随机理论的组合算法。这类算法自 80 年代中后期引起了人们的广泛兴趣,并取得了快速的发展。它们在计算机科学、控制工程、人工智能、模式识别、工程设计、计算数学、生物学、统计物理、化学及社会科学等众多领域中得到了广泛的应用,并且取得了许多令人鼓舞的成就,其中对许多问题的求解远优于传统的优化搜索技术。与此同时传统的优化搜索技术面对许多具有高算法复杂度的问题表现的束手无策,人们被迫进行新的探索,把注意力转向以基因算法为主的组合算法的研究。基因算法已成为当今世界上主要热门算法领域之一。

孟庆春博士自从法国学习期间就致力于基因算法及其应用的研究。他的研究成果受到了国内外专家的高度评价。其主要贡献有以下几点:

1. 首次将基因算法应用于机器人的动力学优化控制和路线规划技术中,并得到了可以满足实时控制要求的良好结果。
2. 创立了基因算法中的对称编码理论,提出了均匀编码和偏值码技术。
3. 发展了带有对称编码的基因算法。在这一算法中采用了许多新的技术,如部分基因保留技术、优良个体和希望型个体保护技术等。这一算法技术在求解机器人控制和路线规划高算法复杂度的问题时的性能远优于传统的基因算法。
4. 提出了带有门限变换函数的基因算法,并将这一算法应用于解是随机分布的问题的求解。通过定义和发展门限变换函数,提高了算法的有效搜索空间。
5. 通过对编码理论的研究,提出了一个对基因算法中编码过程的约束条件。
6. 在将对称编码理论应用于优化控制过程中,提出了两个满足终点速度约束的充要条件。

这些工作都是基因算法领域中具有开创性的研究。通过本著作的出版,将进一步推动基因算法的发展,促进这一研究在我国的开展。基因算法正在迅速地发展,其应用领域十分广泛,努力开展这一研究具有十分重要的学术和应用价值。

孙增圻

一九九五年五月于清华大学

ABE 40/01

前 言

自 Holland 教授于 1975 年对基因算法做了较为系统的总结和阐述以来,此后的一段时期内,并未引起多大的轰动。自从 80 年代,特别是 80 年代中后期,这类算法才得到了较大的发展,引起众多欧美学者的极大兴趣,取得了许多引人注目的成就,成为 80 年代世界上的主要热门算法技术之一。现在每年都有大量研究成果发表。

作为一种新的优化搜索技术,基因算法得到了广泛的应用,诸如计算机科学、控制工程、人工智能、工程优化设计、社会科学、决策系统、模式识别、统计物理、生物学、化学等等,涉及面极为广泛,是许多优化技术所不及的。基因算法之所以受到青睐并得到了如此广泛的应用,就在于这种算法克服了某些传统的解析算法所具有的局限性。

基因算法是一种基于生物进化原理的统计启发式优化搜索技术。这一算法技术利用一个变量群体从多点搜索问题的解,追求问题的解的全局合理性,而非局部优化。一个简单的基因算法由四步组成:

- 编码(产生初始群体)(Coding)
- 群体的再生产(或选优)(Reproduction)
- 个体成员间的基因交换(Crossover)
- 个体变异(Mutation)

对这四步运算中的每一步,人们都提出了多种实现或改进技术。目前有关这一领域的研究,正随着其应用范围的扩大,不断地向深度和广度两个方向发展。

本专著共分五章和二个附录。第一章综述了基因算法的发展过程、算法原理和算法特点、基因算法的主要应用和有待研究的课题。本章是对基因算法整个领域的一个简介。第二章讲述基因算法在问题求解中的应用。这一章中介绍了从前人们提出的两种思想,即分享思想(Sharing Idea)和聚集思想(Crowding Idea),为克服这些算法思想中存在的问题,我们提出了门限变换思想(Gate Limit Change Idea)。这一技术被应用于求解问题的解是随机分布于解空间的情况。第二章还对基因算法的编码理论作了进一步的发展。在定义了十几个基因运算后(如位置换、链、活性块、死亡块、个体成员“与”运算、群体“与”运算等等),我们提出了两个推论和两个定理,为基因算法的编码问题提供了一个充分条件。第三章介绍了基因算法较早期的一些应用和人们提出的基因交换、再生产和个体变异技术等。第四章首先介绍了机器人领域中的有关动力学优化控制、路线规划和工作空间建模问题,提出了二次安全扩展障碍法,扩展量由定理 3 给出。然后针对移动式机器人的特点,提出了带有对称编码的基因算法。在这一算法中,我们发展了多种算法技术,如优良个体保存技术、“移民”技术和部分基因保留技术等等,其中最主要的是对称编码技术。在定义了两种对称编码后,给出了对称编码的 8 个性质并加以证明。根据对称编码理论,最后给出了两个满足双轮驱动式机器人的终点速度约束条件的两个定理。第五章是对第四章提出的算法理论的模拟实验。实验设计共分四个阶段,应用了四种基因算法和三种编码技术。实验结果表明带有对称编码的基因算法大大好于其它的基因算法。最后在二个附录

中分别介绍了机器人的工作空间模型和图论问题。

本书内容取自我们近几年的研究成果,并参阅和引用了国内外较新的(直至1995年)文献资料。这些研究受到了博士生基金、清华大学智能技术与系统国家重点实验室、机械部北京自动化所机器人中心和法国高级电子和电技术工程师学校的资助和支持。

由于基因算法是一个较新的领域,有些内容只是初步的探索,许多问题有待进一步研究,再由于作者学识水平所限,书中内容难免存在不足和错误之处,恳请读者给予批评指正。

作者真诚地感谢 Thelliez 教授和 Hamam 教授给予的极为宝贵的学术建议;真诚地感谢清华大学贾培发教授和机械部北京自动化所机器人中心主任王金友高级工程师的热情帮助;真挚地感谢清华大学计算机系孙增圻教授在百忙之中为本书撰写的序言。

作者热诚地感谢山东大学出版社对本书的出版工作给予的大力支持。

作者十分感谢杨月清女士在百忙之中为本书做了大量的文字工作。

作者

1995.6



作者简介

孟庆春，男，
生于1958年。1982
年毕业于山东大学
电子系，1986年于
吉林工大电子系硕
士毕业；1993年获
法国巴黎第十二大
学博士学位。大学
毕业以后，一直从
事高教工作，现工
作于烟台大学。

作者主要从事
电子技术、微机应
用和自动控制方面
的研究。在国外学
习期间及归国后，
主要从事以机器人
技术为主的智能控
制技术和智能仪表
方面的研究。

目 录

序言

前言

第一章 基因算法的理论和原理	(1)
1.1 基因算法的起源简介	(1)
1.2 基因算法的步骤	(3)
1.2.1 基因算法的原理	(3)
1.2.2 基因算法的基本步骤	(4)
1.3 一个简单的基因算法及其特点	(6)
1.3.1 一个简单的基因算法实例	(6)
1.3.2 基因算法的特点	(8)
1.4 基因算法领域中有待进一步研究的问题	(9)
1.4.1 改进群体成员的再生产的方法	(10)
1.4.2 改进个体成员间的基因交换技术	(11)
1.4.3 编码技术的改进	(12)
1.4.4 改进算法的总体思想	(12)
1.4.5 基因算法与其它算法的结合	(13)
1.4.6 在基因算法中引入并行算法技术	(13)
1.5 基因算法的应用简介	(13)
1.5.1 基因算法在函数优化方面的应用	(14)
1.5.2 基因算法在机器人控制方面的应用	(15)
1.5.3 基因算法在控制系统中的应用	(15)
1.5.4 基因算法在工程优化设计方面的应用	(16)
1.5.5 基因算法在机器学习系统中的应用	(16)
第二章 带有门限变换函数的基因算法	(18)
2.1 基因算法中的分享思想和聚集思想	(18)
2.1.1 分享思想	(18)
2.1.2 聚集思想	(21)
2.2 带有门限变换函数的基因算法的原理	(21)
2.2.1 门限变换函数的定义	(21)
2.2.2 带有门限变换函数的基因算法	(23)
2.3 带有门限变换函数的基因算法的应用	(25)
2.4 带有门限变换函数的基因算法的门限值选择方法	(28)
2.4.1 门限值 GH 的选取方法	(28)

2.4.2	模拟实验及结果分析	(29)
2.5	关于基因算法中的编码理论	(31)
第三章	基因算法的主要应用	(38)
3.1	基因算法的初始应用	(38)
3.1.1	Bagley 和自适应式操作游戏方法	(38)
3.1.2	Rosenberg 及其生物细胞模拟	(39)
3.1.3	Cavicchio 及其模式识别	(39)
3.1.4	Hollstien 和函数优化	(41)
3.2	基因算法在工程优化设计中的应用	(42)
3.2.1	管道系统能耗的优化	(42)
3.2.2	结构优化	(45)
3.3	基因算法在控制系统中的应用	(47)
3.3.1	基因算法在控制策略学习系统中的应用	(47)
3.3.2	基因算法用于动态系统中的控制规则归纳	(53)
3.4	基因算法与机器学习	(56)
3.5	基因算法与神经网络理论	(60)
3.5.1	Miller 的神经网络	(61)
3.5.2	Miller 的基因运算	(62)
3.5.3	个体成员的质量计算	(62)
3.5.4	实验方法	(63)
3.5.5	实验结果	(63)
3.5.6	其它方向的研究	(64)
第四章	基因算法在移动式机器人的动力学控制和路线规划方面的应用	(66)
4.1	机器人问题简介	(66)
4.1.1	机器人的定义和分类	(66)
4.1.2	机器人领域的主要研究方向	(67)
4.2	移动式机器人的动力学模型	(68)
4.2.1	双轮驱动式机器人的动力学模型	(69)
4.3	机器人的工作环境模型	(72)
4.4	移动式机器人的动力学优化控制函数的构成	(75)
4.4.1	罚函数原理	(75)
4.4.2	移动式机器人路线规划的约束及其动力学优化控制评价函数的构成形式	(77)
4.5	带有对称编码的基因算法	(81)
4.5.1	对称码理论	(81)
4.5.2	带有对称编码的基因算法思想	(84)
4.5.3	关于对称编码性质的证明	(86)
4.6	带有对称编码的基因算法在机器人优化控制中的应用	(88)

第五章 基因算法在机器人优化控制应用中的模拟实验结果	(92)
5.1 实验设计介绍	(92)
5.1.1 实验阶段一	(93)
5.1.2 实验阶段二	(97)
5.1.3 实验阶段三	(97)
5.1.4 实验阶段四	(101)
5.2 算法评价和展望	(108)
附录	(110)
附录 I	(110)
附录 II	(117)
后 记	(126)
参考文献	(130)

第一章 基因算法的理论和原理

在人类的社会生活和科学研究活动中,每当人们遇到问题时,都是首先力图找到问题的解答,然后再努力改进所找到的解,以至达到所得解的最优程度。寻找一个问题的最优解是人们努力追求的目标。因此,求解问题的最优解的优化技术就随之而产生,并且已经引起了人们的高度重视,在许多生产、科研领域中得到了迅速的推广和应用。

追求问题的最优解,具有广泛的社会效益,可以大大地提高生产率,节省人力、物力和财力,促进社会生产的发展。优化技术正是在这种推动力下诞生和发展起来的。当今优化技术已成为一个重要的科学分支。它涉及到了社会生产、科学研究的许多领域,如工程优化设计、系统控制、计算机工程、人工智能等等。优化技术在某一学科中得到应用时,往往在该学科中就会产生出一个新的学科分支来。这充分表现了优化技术这一新兴科学的活力。特别由于工程应用和快速大容量计算机的发展以及算法的改进,优化技术在最近 20 年中发展十分迅速。

正是在人们追逐社会生活中许多问题的最优解的过程中,优化、搜索技术不断地发展、完善^[29,111,119]。到今天,已经有许多发展成熟的优化技术,例如线性规划、非线性规划、动态规划(1961)等等^[123]。另有一类建立在统计理论基础上的组合优化技术,如模拟淬火法^[1,46,47,49,63]、树型淬火法^[6,9,36]和基因算法^[18,29,32,41]等。拓扑学的网络理论中也涉及到许多优化技术问题^[114,122]。神经网络理论^[111]同样可应用于这一领域中。可以说优化技术是当今世界上发展最快和最具有活力的学科之一。特别随着计算机技术和人工智能技术的飞速发展,人们对各种优化技术的需求更加强烈了。

基因算法是一类基于统计理论基础上的组合优化、搜索技术。它们的发展较晚,直至 80 年代中后期,才真正引起了广大学者的重视。这一算法的创始人,美国密西根大学的 Holland 教授是在认真分析了生物社会进化发展的规律后,才提出和发展了这一算法的理论,并得到了人们的广泛承认。下面我们将介绍这一技术的发展过程和基本原理。

1. 1 基因算法的起源简介

人们已经对传统的优化和搜索技术有了较多的了解,这些技术也较为成熟,并得到了广泛地应用,很好地解决了许多实际问题。但是传统的方法也存在有相当的弱点。Goldberg 在其论著^[29]中首先分析了传统方法的某些缺点,从而为基因算法的提出开辟了道路。

Goldberg 将传统的基于计算的优化、搜索技术分为两类:间接方法和直接方法。传统的方法往往要求所求解函数的连续性。但在实际中,我们常常会遇到一些不连续的且具有大的解空间和具有多个解、以及在解空间中存在有严重噪声(干扰)的问题。这样

的问题往往具有很大的算法复杂度，不适合于传统方法。对这类问题，传统的方法很难达到所期望的目标。因为在许多复杂的问题中，传统方法所求得解往往只是局部的，不能够反映出问题的整个解的形式。只有对它们进行相应的改进后方可能提高其求解问题的性能，从而求得高质量的解。

另一方面，传统方法所得到的解的鲁棒性 (Robustness) 往往不好。Goldberg 列举了一个具有两个峰值函数和一个不连续具有噪声的函数，以说明传统方法往往不知道“爬哪个山峰”好^[29]。但是要求一个优化算法所得到的解具有鲁棒性是十分有意义的。对于许多人工系统 (Artificial Systems) 来说，解的鲁棒性往往会带来很多收益。如果一个人工系统具有较大的鲁棒性的话，那么费用昂贵的再设计和改造工作就可以减少或避免。如果具有更高的自适应能力，那么现存的系统就可以更好地、更长时间地执行和完成其功能。无论是在工程系统、计算机系统还是在商业系统中，人工智能的设计者一定要追求系统的鲁棒性、有效性和柔软性。这就促使设计者要去求助于生物系统所具有的上述要求。生物系统的自我改造 (Self repair)、自我导向 (Self guidance) 和再繁殖 (Reproduction) 的特点，为它们在更复杂系统中的应用奠定了基础。

因此，我们得出了一个令人鼓舞的结论：在要求具有鲁棒性的场合，生物界的进化规律更具有吸引力。通过对生物界的研究，自适应和自生存的秘密就可以被很好地掌握。面对许多复杂的问题和传统优化搜索技术的缺陷，在认真研究了生物界的进化特点后，人们就提出了基因算法 (Genetic Algorithms : GAs) 这一理论。从而形成了一类新的算法技术，开辟了一个新的算法领域。

然而基因算法并非只是一个漂亮的措辞。在人们的研究中，已经从理论上和实践上证明了，基因算法 (Genetic Algorithms) 能为求解更复杂的问题提供具有鲁棒性的优化、搜索技术。在这方面第一部专著是密西根大学的 Holland^[41] 教授于 1975 年的 “Adaptation in Natural and Artificial Systems” 一书。随后的一些博士论文和文章也为这一算法的进一步发展提供了后续工作和成果。特别在函数优化和系统控制应用方面，基因算法得到了广泛、深入的研究，并且已经有了扎实的理论基础。现在这一领域中的主要工作是进一步推广其在商业、科学和工程中的应用。应该指出 Holland 教授的学生也对这一算法的发展和完善，做出了许多重要贡献^[10,18,29,103,104,105,109]。

所以，正是由于人们意识到了基于计算 (Calculus based) 和数值式的优化技术的弱点后，才去努力研究和发展统计启发式的搜索技术。这类技术在许多方面虽然还不能说好于传统的方法，但是也已经表现出了非凡的能力，解决了一些用传统方法难以解决的复杂问题，并得到了满意的结果，这就是为什么人们对它们产生了越来越浓厚的兴趣。因此有人断言，这类优化技术大有超过传统方法的趋势。

我们知道，世界上任何生物，在自然界中都受到两种因素的影响，即不利于其生存的环境条件和有利于其生存发展的环境条件。在生物的进化过程中，它们不断地自我改变，以求能适应自己周围的生存环境，将不利条件转变为有利条件。有利条件和不利条件作为一对矛盾，总是存在于生物界。这就要求生物自身不断地进化、发展。在大自然中，能够很好地改变自己 and 能够适应生存环境的物种将生存延续下去。相反，不能的，则要消亡。基因算法就是在将这一生物进化原理应用于优化、搜索技术的过程中，形成了

自己的算法特点。

生物在进化过程中,优良的个体能够得到更多、更快地繁殖,而劣质的个体则易消亡或得不到较好地繁殖。一个物种中的个体成员间进行一些基因交换,形成杂交优势,可以产生出更优良的个体成员。通过淘汰劣质个体和个体间的基因交换,使整个物种的质量得以提高。一种生物要延续下去和进行生物基因交换,还需要一个具有一定规模的群体 (Population)。基因算法正是在生物界的这一进化原理基础上发展起来的。

与其它的优化技术不同,基因算法应用概率的转换规则来引导其搜索过程,而用确定性方法来组织相似的成员。这并不奇怪,概率理论的应用并不表明这一方法仅仅采用随机的搜索技术。基因算法使用随机选样法作为一个工具来引导算法向着搜索空间中具有解存在的方向发展。

基因算法具有一些新的思想,而这些思想从未被人们在优化、搜索技术领域中的应用过。Holland 教授在全面研究了自然界中生物进化的过程后,才提出了这一理论。后来这一理论引起了人们广泛的兴趣。特别在 80 年代中后期,人们将基因算法的研究推到了高潮。到了 90 年代,基因算法的理论已基本成熟。人们的研究方向主要集中于基因算法在许多领域中的应用和与其它算法的结合两方面。其应用的范围相当广泛,从社会科学到工程优化设计、计算机科学、控制工程等领域。所求解的问题的复杂性也越来越大。基因算法与其它算法的结合,主要是为了使不同的算法之间能够相互取长补短,以便形成更好的算法。

可以简单地讲,基因算法与传统方法相比有四点不同之处:直接使用变量的编码形式,用一个变量群体搜索问题的解,采用以目标函数值为主的信息和采用统计随机理论。这些不同之处导致了该算法相对于其它常用算法的多种优势。下面我们来介绍关于基因算法的组成、原理和特点。

1. 2 基因算法 (GAs) 的步骤

经过以上介绍,我们已知基因算法是一种基于在生物界进化过程中,优胜劣汰的原理基础上发展起来的优化、搜索技术。是一种采用了统计启发式搜索技术的组合优化算法。可以说与模拟淬火法属于同一类。但是,模拟淬火法 (Simulated Annealing: SA) 是在蒙特-卡罗 (Monte Carlo) 算法原理^[64,67]基础上发展起来的,其中采用了统计启发技术。这一算法的创始人 Kirpatrick 受到了统计物理中分子降温过程的启发,故被称之为模拟淬火法^[24,46,47,76]。人们曾用 Boltzmann 机器理论对这种算法的机理进行了分析^[1],并用非平稳马尔科夫链理论对其收敛性进行了研究^[27,58]。所以说这种算法在原理上与基因算法有所区别,只是两种算法都引用了统计理论而已。但都属于组合优化算法。从 80 年代至今,模拟淬火法也十分盛行^[63,76,88],在许多实际问题中得到了应用,它是 80 年代出现的主要算法之一。

1. 2. 1 基因算法的原理

在将生物界的进化规律引入到优化求解技术中来时,我们所遇到的第一个问题就是

必须思考如何将所要求解的问题中的变量参数变成象带有生物基因一样的个体成员 (Members), 或者讲染色体 (Chromosomes)。这些基因会影响生物个体的进化发展, 通过基因改变和互换基因, 可以改变生物个体的特性, 产生出品质不同的个体。同样在问题求解过程中也有类似的仿照。我们可以设想将问题的变量参数进行编码 (Coding), 编码后产生出多个不同的变量个体成员或称编码串 (String), 每个个体成员由多个单元或称元素表示 (Elements)。这些元素好象生物体的基因, 编码串好比生物体内的染色体 (Chromosome)。通俗地讲, 这些元素就是编码串中的位 (Bits)。每个个体成员所反映出来的特征不一样, 有的适应度 (Fitness) 好, 接近问题的解; 有的适应度差, 远离问题的解。那么适应度好的个体成员应该得到发展和“繁殖”。而适应度差的个体成员则不应该再发展下去。那么第二个问题就是应该有一个评价个体成员优劣的标准, 以决定每个个体的前途。

第三个问题是如何进行个体成员之间的基因交换问题。我们知道生物界的基因交换是通过“婚配”来进行的。“婚配”后, 两个个体成员将产生出下一代个体成员, 即后代 (Offsprings)。所产生的后代, 既带有其父辈双亲的特点, 也就是说继承了先辈的基因, 又有与其父辈双亲不同的特点, 也就是说发生了演化或变异。应该讲, 继承和演化是一对共存的矛盾。继承是为了演化, 演化又是为了新的继承。生物群体在演化和继承的矛盾中不断地发展和进化, 使其能够适应自身的生存环境, 克服自身原存的弱点, 一代一代地发展、改良下去, 以完善自身。对于基因算法来讲, 我们还需要确定一个在编码串之间进行元素交换的方式。

以上三个问题是基因算法的基本组成。必须很好地解决这三个问题, 才能体现出这类算法的特点。从近年来的许多研究来看, 基因算法的发展和完善也是从解决这三个问题入手的。针对这三个问题, 人们提出了许多改进措施, 形成了许多改进的各有特色的算法技术。这是合情合理的, 因为我们不能指望生物界的众多物种以同样的方式进化、繁殖。在广大的生物界, 各自的生存环境有异, 因而自己进化、繁衍的方式也不相同。同理, 在优化搜索技术领域, 人们有许多要求解的问题, 这些问题具有各自的特点, 有连续的、非连续的; 有多解的、单解的; 有逻辑型的、模拟型的; 有等值解的、非等值解的等等, 情况复杂多变, 我们同样不能期望用一种技术来完全有效地求解所有的问题。因此, 根据所求解问题的不同特点, 发展不同的算法技术是理所当然的和完全必要的。也正是由于这一点, 优化搜索技术才能得到不断的发展, 才有其发展的动力。

1. 2. 2 基因算法的基本步骤

根据以上提出的三个问题, 从生物界的进化原理出发, 一个简单的基因算法从形式上看并不复杂, 它应该包含如下四个基本运算:

- * 变量编码 (Coding)
- * 群体成员的再生产或繁殖 (Reproduction)
- * 个体成员之间的元素 (基因) 交换 (Crossover)
- * 个体成员的变异 (Mutation)

在基因算法中每次迭代运算, 即完成再生产、基因交换和个体变异后, 所产生出来

的新的群体成员被称为新一代成员，或称下一代群体 (Offspring)。每次迭代运算称之为代运算 (Generation Operation)，可以表示为 $G(t)$ 。有时 $G(t)$ 也可表示第 t 代的群体成员，但一般用 $P(t)$ 表示一代群体成员。

以上每一步运算的具体含义为：

- * 变量编码是以随机形式产生出一个反映不同变量参数的群体成员，也就是第一代群体成员 $P(1)$ 。这些个体成员是问题的解的候选成员。形象地讲，有时也称个体成员为编码串 (Strings)。编码串可以是二进制形式的，也可以是十进制形式的。一个群体中应含有一定数量的个体成员。

- * 再生产 (或选出优良成员 (Selection)) 过程是一个复制优良成员的过程。

所谓评价优良成员的标准，一般是根据一个个体成员所反映的评价函数的值或在群体中占有的概率值或适应度 (Fitness) 来评价的。一个个体成员具有较大的适应度，那么它就有更多的机会进入下一代的群体中，有更多的机会得到发展。反之，不良的个体成员就不易进入下一代的运算中，就有可能死亡。

- * 个体成员之间的元素 (Elements) 或基因交换，分两步完成。第一步首先随机性地选择配对成员，即基因交换的伙伴和交换基因的位置；第二步在所选中的伙伴间和基因交换位置上，交换个体成员间的一部分元素。交换方式多种多样，可以从交换位置上开始交换其前的每一个元素，也可以只交换一定比例的元素等等。例如，有两个个体成员被配对为基因交换伙伴，即 S_1 和 S_2 ：

$$S_1 = (0\ 1\ 1\ 0\ 1), \quad S_2 = (1\ 1\ 0\ 0\ 0)$$

同时随机地选中了位置 2 (即第二位) 为基因交换位置 (从右边向左数)，两个个体成员中从第二位以前的所有各位全部交换，用 “—” 符号表示基因交换位置，则有

$$S_1 = (0\ 1\ 1\text{—}0\ 1)$$

$$S_2 = (1\ 1\ 0\text{—}0\ 0)$$

基因交换后则产生出两个新的个体成员 S_1' 和 S_2' ：

$$S_1' = (0\ 1\ 1\ 0\ 0)$$

$$S_2' = (1\ 1\ 0\ 0\ 1)$$

通过基因交换改变了原来的两个个体成员，而得到了两个新的个体成员。

- * 个体变异的作用主要是为了避免在随机性的基因交换中发生恶性变化的情况。或者说是为了防止一些个体成员丢失重要的基因 (或元素)。在基因算法中，这是一个补救性的措施。也是为了进一步防止算法的不成熟情况 (Prematurity) 发生。

应该指出，编码过程、个体基因交换、群体成员的再生产和个体变异都是按照统计和随机理论进行的。随机选择的方式是多种多样的，因算法设计的不同而变。一般地讲有两个参数比较重要，即基因交换概率 P_c 和个体变异概率 P_m 。它们分别决定群体中参加上述运算的个体数量。

基因算法中经常使用如下符号：

- * $G(t)$ ：第 t 代运算
- * $P(t)$ ：第 t 代的群体成员

- * P_c : 个体成员间的基因交换概率
- * P_m : 群体成员的个体变异概率
- * SP : 问题的解空间, 且有 $P(t) \in SP$

基因交换概率 P_c 的计算是参加个体基因交换的数量和整个群体成员的数量之比。个体变异概率 P_m 是参加变异的元素与整个群体中含有的元素之比。

设一个群体 $P(t)$ 含有 m 个体成员, 每个个体成员有 n 个元素, 每一代运算中参加个体基因交换的成员数量为 m_c , 参加个体变异的元素为 n_m , 那么有:

$$P_c = \frac{m_c}{m} \quad P_m = \frac{n_m}{n \cdot m}$$

有时在基因算法中, 根据需要经常开辟一些子群体成员 (Subpopulation), 以 SSP 表示。这些子群体成员分别对应于部分解的子空间。下面通过一个实例来详细介绍一个简单的基因算法的全过程。

1. 3 一个简单的基因算法及其特点

可以说, 一个简单的基因算 (A Simple Genetic Algorithm: SGA) 法令人惊讶的简单。简要地讲, 就是根据一定的评价标准, 反复地复制群体成员中的部分个体成员和除去部分个体成员, 直至算法结束为止。那么这一算法怎样工作呢? 我们不妨通过一个手工的例子来看一下基因算法的步骤及其产生的效果。

1. 3. 1 一个简单的基因算法实例

设有一个二次函数

$$f(x) = x^2 \quad x \in [0, 31]$$

要求以整数 1 的精度求

$$\max f(x)$$

通过第一步随机编码后, 将产生一个初始群体 $P(1)$, 设在初始群体中有四个个体成员, 每个个体成员的长度为五位:

(0 1 1 0 1)

(1 1 0 0 0)

(0 1 0 0 0)

(1 0 0 1 1)

从理论上讲, 群体中个体成员的长度是根据所求解问题的解空间的大小和所要求的解的精度来确定的。问题的解的空间大或者所要求的解的精度高, 那么群体中的编码串的长度就要长, 即位数要多。反之, 位数就少。

这四个成员各自代表的目标函数的值 f 不同或者讲各自的适应度 (fitness) 不同 (适应度一词来源于生物界)。我们自然地要用 f 的值来衡量每个个体成员的优劣, 以决定它们未来的命运。那么我们将这四个取样成员及其适应度和各自在群体中所占的比重列于表 1. 3. 1 中:

表 1.3.1 个体成员所占比重的计算

编 号 №	个体成员 (Strings)	适应度 (Fitness)	占适应度的百分比 (% of Total)
1	01101	169	14.4
2	11000	576	49.2
3	01000	64	55.0
4	10011	361	30.9
合计		1170	100.0

Goldberg 将基因算法中各成员所占的比重形象地表示在一个取样概率圆上 (Sample Wheel), 以每个个体成员在该圆上所占的面积来表示其百分比。

下面我们以手工方式模拟一个简单的基因算法, 并将所得结果列于表 1.3.2 中。

按照基因算法的基本步骤有:

1. 对函数的变量 x 以二进制形式进行编码, 产生出四个初始个体成员 (见表 1.3.2 中的 (1) 项)。

2. 计算 x 和 $f(x)$ 的值, 以便决定传递到下一代的成员数量 M_p (Passing Members), 见表 1.3.2 中的 (2) 项。

3. 计算每个个体成员所具有的适应度和比重 (见表 1.3.2 中的 (3) 项)。

4. 根据各个个体成员所占的比重选出传递到下一代的个体成员及其要复制的数量 (见表 1.3.2 中的 (4) 项)。

5. 随机地选择进行个体成员间基因交换的伙伴和基因交换的位置 (见表 1.3.2 中的 (5) 项)。

6. 完成基因交换, 生成新一代群体成员 (见表 1.3.2 中的 (6) 项)。

7. 如果算法进行到了第 GN 代 (即代数 GN), 那么算法停止; 否则返回到第 2 步中去。

表 1.3.2 简单基因算法的手工模拟

成员编号 №	(1) 初始群体	(2) x_i	(2) $f_i(x_i)$	(3) $f_i / \sum f_i$	(3) $f_i / \bar{f}$	(4) M_p
1	01101	13	169	0.14	0.58	1
2	11000	24	576	0.49	1.97	2
3	01000	8	64	0.06	0.22	0
4	10011	19	361	0.31	1.23	1
总 和:			1170	1.00	4.00	4
平均值:			239	0.25	1.00	1
最大值:			576	0.49	1.97	2