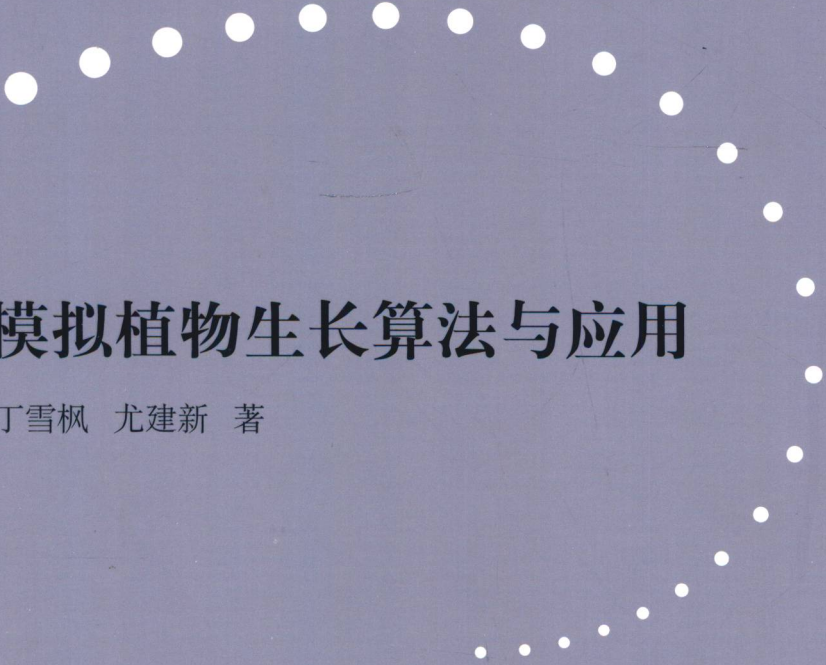




模拟植物生长算法与应用

丁雪枫 尤建新 著

上海
人
民

模拟植物生长算法与应用

丁雪枫 尤建新 著

上海
人
民
出
版
社



图书在版编目 (C I P) 数据

模拟植物生长算法与应用/丁雪枫, 尤建新著.

上海: 上海人民出版社, 2011

ISBN 978 - 7 - 208 - 09819 - 0

I. ①模… II. ①丁…②尤… III. ①最优化算法

IV. ① 0242. 23

中国版本图书馆 CIP 数据核字 (2011) 第 023411 号

责任编辑 曹怡波

封面设计 张志全

模拟植物生长算法与应用

丁雪枫 尤建新 著

世纪出版集团

上海人民出版社出版

(200001 上海福建中路 193 号 www.ewen.cc)

世纪出版集团发行中心发行

上海商务联西印刷有限公司印刷

开本 720 × 1000 1/16 印张 8.25 插页 2 字数 102,000

2011 年 3 月第 1 版 2011 年 3 月第 1 次印刷

ISBN 978 - 7 - 208 - 09819 - 0/T · 4

定价 28.00 元

随着人们的知识进步,管理实践的许多活动方案被不断优化,创造出了丰硕的成果。因此,对优化理论和方法的探索,始终是管理学科领域的一个亮点。在现代信息技术飞速发展的今天,互联网、物联网等新概念和新技术大大地促进了经济全球化的进程,对世界经济的蓬勃发展产生了深远的影响。在这样的发展背景下,优化问题将扮演更加重要的角色。如何有效地选择参数,使得日常运作既能够更好地满足目标的要求,又能够不断地降低成本;如何将有限资源进行更加合理地分配,使得资源配置得以不断的优化……传统的以梯度为基础的优化方法虽然具有可靠性强,计算效率较高等特点,但是处理复杂、困难的非确定性的优化问题(Non-deterministic Polynomial 问题,即多项式复杂程度的非确定性问题,简称 NP 问题),如目标函数或约束条件不连续、不可微、高度非线性等情况时,则能力较差,具有许多的局限性。随着研究的实际系统规模的不断壮大,约束条件增多,系统具有多准则、非线性、不可微、不确定的特点,结构越来越复杂,对于系统的数学描述模型的建立也越发的困难,许多传统的管理理论与方法已经不足以精确地对问题进行求解。这种情况下,探求适用于大规模计算并且具有智能特征的问题求解方法具有十分重要的现实意义。

20 世纪以来,生物界中蚂蚁、鸟儿等动物的一些群居行为引起了研究学者们的关注,这些群居动物可以通过简单的集体协同行为来高效完成觅食、

搬运、防御、建巢、保持和变换队形等复杂的任务和事情,科学家们受动物的行为启示,将这些协同方法采用信息手段在计算机上实现来求解一些实际问题,取得了较为满意的结果。作为计算智能的重要研究内容,目前较为常用的智能优化方法主要有遗传算法、蚁群算法、模拟退火算法、微粒子群算法等等,这些算法大多是建立在生物智能或物理现象基础之上的随机搜索算法,所具有的特点包括:算法的机理简单,易于编程设计与实现,对目标函数和约束条件没有连续性和凸性的限制要求,对计算中数据的不确定性也具有较强的适应能力,多数的情况下都可以在一个可接受的时间范围内获得问题的较优或最优解。智能优化算法凭借其独特简单的算法结构与突出的问题求解能力机制,引发了国内外广大学者的关注和研究热潮,并且在诸多领域得到广泛应用,取得了许多令人满意的成果,展示了强劲的发展势头。

由于大多数的智能优化算法的理论思想来源于自然界中生物群体的群落现象的模拟,因而在理论上还需要不断地进行完善,并提高求解的最优性水平。这时,人们想到了对植物生长的模拟。20世纪30年代,俄罗斯学者Cholodny提出植物的向光性动力机制主要来源于光源引起了一种可促进植物生长的物质向背着光的一侧传递生长,该物质的浓度是随着光源强度的变化而不断变化的。此学说提出不久之后,荷兰学者F. Went用实验的方法证明了Cholodny的学说,这种随着光线强度而不断变化的物质称为形态素。对光面形态素浓度变小,背光面形态素浓度增加,会导致切段两侧形态素浓度分布不均,这种情况下,背光面的细胞纵向生长的速度加快,向光面的细胞的纵向生长的速度则会减慢,使得最终茎干向光弯曲。

20世纪70年代,学者Przemyslaw Prusinkiewicz等人提出了采用计算机模拟植物生长的系统,称为L-系统。根据L-系统的理论,植物的拓扑结构可以表示为植物的根、节点、主枝干、侧枝干,以及植物顶端的选段序列。植物的生长过程是一个规律的分解过程,植物的生成可表示为一个纵向或者横向

的拓扑结构。L-系统通过使用一套自定义的字符串文法和一组对应的替换式产生规则,将这些规则的文法进行分析和替换,经过有限次的迭代之后,就可以描绘出植物发芽、分支、节点生长等整个植物的生长过程,继而生成复杂的图形。L-系统揭示了复杂的自然界现象总可以找到简单的方法来对其进行描述。模拟植物生长算法就是由 L-系统演化而来的。它是通过模拟植物枝干长满整个生长空间的过程,即由植物生长的内在动力及向光性的作用力,建立茎、枝干繁殖生长及凋谢的动力机制,并根据植物学中的形态素浓度理论建立不同光线强度的环境下按照全局最优的方式向着光源快速生长的动力模型。新枝长出后,刚长出新枝的生长点将从生长集合中被删去,新生长出的枝干上所包含的生长点将被加入到生长集合中,反复执行该步骤,直到没有新的枝干生出停止,最后一株完整的植物就生成了。

显然,模拟植物生长算法作为一种新型的智能优化算法,没有参数设置的限制要求,求解大规模的优化问题具有较强的稳定性,对于非线性类的优化问题尤其是整数规划类问题具有一定的普遍性和通用性。可以看出,模拟植物生长算法是一种主动、积极的模拟优化算法,是一个理论上有深化研究意义、实践上有广泛应用价值的模拟优化算法。模拟植物生长算法可以从宏观上帮助人们优化资源配置,提高资源效率和有效性;从微观上帮助某个组织机构(如企业)优化经营管理方案,提升其竞争力和投资回报效率。

本书是以丁雪枫的博士论文为基础,由丁雪枫和尤建新合作为模拟植物生长算法撰写的专著。在这本书中,作者从模拟植物生长算法的思想来源、基本原理出发,介绍了模拟植物生长算法及其在一些典型问题当中的应用,涉及的应用重点还包括了模拟植物生长算法应用于整数规划问题、无线传感器网络最小连通集合划分问题、虚拟企业盟友选择问题、物流中心选址问题、作业车间调度问题的实现技术及应用等内容,而本书的主要贡献是提出了对模拟植物生长算法研究的最新成果和应用技术,为管理实践中的优化问题提

供了新的思路和求解方法。

本书以丁雪枫为主笔,尤建新辅助。全书的体系结构及内容包括五个部分:第一部分是概念,对模拟植物生长算法涉及的研究背景知识及其意义、目标等进行了介绍;第二部分是算法,对模拟植物生长算法的基本原理、算法的动力机制、算法的实现方法、算法的收敛性分析等进行了阐述;第三部分是特点,将模拟植物生长算法与其他几种优化算法进行了对比分析,并讨论了模拟植物生长算法的特点及优势;第四部分是应用,对模拟植物生长算法在一些典型 NP 实际问题中的应用进行了深入的阐述,通过采用模拟植物生长算法及其他优化算法对各个问题分别进行求解,并将计算结果进行比较分析,突出了模拟植物生长算法的性能;第五部分是改进,指出并分析了基本模拟植物生长算法的一些缺点,提出了一些相应的改进策略,通过对实例的计算和进行结果对比,分析验证了改进策略的可行性。

这本专著有助于学习和研究智能优化算法的同仁们加深对模拟植物生长算法的理解,并在应用方面有所支持。这本专著的形成,还得到了上海理工大学博士生导师马良教授的鼎力相助和指导,受益匪浅。我们真诚地希望这是一个持续优化的开端,在优化理论和算法的发展进程中有更多的成果不断涌现,为管理理论和实践的进步添砖加瓦、增添光彩。

丁雪枫 尤建新

2011 年 1 月 11 日于同济大学

目 录

前言	001
----------	-----

第一章 绪论	001
--------------	-----

§ 1.1 优化问题介绍	001
§ 1.2 植物生长的生物学特征	006
§ 1.3 模拟植物生长算法的思想起源	008
§ 1.4 研究模拟植物生长算法的意义	008
§ 1.5 模拟植物生长算法的研究现状	009
§ 1.6 本章小结	010

第二章 模拟植物生长算法模型	011
----------------------	-----

§ 2.1 模拟植物生长算法基本原理	012
§ 2.2 模拟植物生长算法动力学特征	017
§ 2.3 模拟植物生长算法的流程及实现	019
§ 2.4 模拟植物生长算法的收敛性分析	023
§ 2.5 本章小结	027

第三章 模拟植物生长算法与其他优化算法的比较	028
------------------------------	-----

§ 3.1 遗传算法	028
------------------	-----

§ 3.2	蚁群算法	034
§ 3.3	模拟退火算法	038
§ 3.4	微粒子群算法	042
§ 3.5	各算法的对比分析	046
§ 3.6	本章小结	048
第四章	模拟植物生长算法的典型应用	049
§ 4.1	整数规划问题	049
§ 4.2	WSN 最小连通集划分问题	056
§ 4.3	虚拟企业盟友选择问题	067
§ 4.4	中心选址问题	076
§ 4.5	作业车间调度问题	084
§ 4.6	本章小结	090
第五章	模拟植物生长算法改进策略及其应用	091
§ 5.1	基本模拟植物生长算法的缺陷	091
§ 5.2	模拟植物生长算法的几种改进策略	092
§ 5.3	模拟植物生长算法的改进应用	095
§ 5.4	本章小结	109
第六章	展望	110
§ 6.1	算法总结	110
§ 6.2	未来研究工作	113
参考文献		114

§ 1.1 优化问题介绍

在拟阵概念应用到图论理论之前,优化问题最初起始于是对一些实际应用问题的研究,如开关电路的设计、交通运输路线设计、物品装箱设计、广播网的设计等等。当将拟阵概念结合到图论领域当中之后,对于拟阵的理论及其应用研究如 TSP 问题、最短路径问题、最优支撑树问题等,逐渐成为了优化问题研究的目标和热点。

§ 1.1.1 组合优化问题

组合优化(Combinatorial Optimization)问题又称为组合规划,它在管理科学、计算科学、数学等许多工程领域都有着很重要的实际应用,也是运筹学的一个重要分支,其研究基础是针对连续变量求解的函数优化问题^{[1][2]}。组合优化问题是指对有限、离散的数学结构,搜索即可以满足限定条件,又可以使给定的目标函数值达到最优(最大或者最小)的解的研究^{[2][3]}。对于组合优化问题的研究主要包括二次分配问题(Quadratic Assignment Problem)、作业调度问题(Scheduling Problem)、最大/最小树问题(Maximal/Minimal Tree

Problem)、图着色问题(Graph Coloring Problem)、聚类问题(Clustering Problem)、旅行商问题(Traveling Salesman Problem)、装箱问题(Bin Packing Problem)、0-1背包问题(0-1 Knapsack Problem)等等^[4]。目前,随着算法研究的不断深入和算法求解技术日益发展,许多原始的组合优化问题的模型对于求解实际问题已不再适合,因此,这种情况下,不断出现了一些基于原始组合优化问题提出或改进而产生的新的组合优化问题,如最短路径问题(Minimal Path Problem)、最小费用流问题(Minimal Cost Flow Problem)、运输问题(Transportation Problem)等等^{[5][6][7]}。

组合优化问题的目标是在给定有限集的所有可行解子集中,根据某限定目标或条件找出最优子集解。在没有约束条件的情况下,优化问题的数学模型可描述为^[8]:

$$\min f(X) \quad (1.1)$$

通常情况下,这类问题都受到一个或多个约束条件的限制,最优解不但与目标函数有关,还要满足给定的约束条件。因此,最优化问题的数学模型可表示为:

$$\min f(X) \quad (1.2)$$

$$\text{s. t. } g_n(X) \leq 0 \quad (n = 1, 2, \dots, p) \quad (1.3)$$

$$h_m(X) = 0 \quad (m = 1, 2, \dots, q) \quad (1.4)$$

其中, $f(X)$ 为目标优化函数, $X = \{x_1, x_2, \dots, x_k\}$ 为所有可行解构成的解集合, $g_n(X)$ 和 $h_m(X)$ 是约束条件函数。对数学模型的求解实际上就是在可行解空间集合中,对于所有的 $x_i \in X$, 找出最优解 x^* , 使得 $f(x^*) = \min f(x_i)$ 。

§ 1.1.2 算法的计算复杂性

20 世纪 30 年代,对于数论函数的可计算性问题,已有一些学者提出采用可计算函数定义来解决,如 1935 年 Church 提出的 λ -转换演算、1936 年 Turing 提出的抽象计算机模型以及 Godel、Kleene 等人提出的递归函数定义等等^{[9][10]}。20 世纪 60 年代,Hartmanis 和 Stearns 采用对角线法证明了时间递归定理,并首先提出了时间复杂性类的概念,即 $\text{TIME}(f(n))$,这之后逐渐出现越来越多的计算复杂性的研究理论,如近似算法的不可近似性研究、交互式证明系统研究、随机算法的去随机化研究、算法博弈论研究等。可计算性研究主要是针对问题的求解方法,而算法计算复杂性则是从问题求解过程中需要占用的时间和空间等资源出发,继而考虑尽可能减少所占用资源的问题。

算法(Algorithm)是指用来实现某一个任务所需要的方法和具体执行操作步骤。一般情况下都是根据给定的输入信息数据或者数据的初始值,在执行一连串有限的计算步骤后得出期望的输出信息数据或者数据的终止值。算法中往往都会包含一些重复操作以及条件判断,并且在求解同一个问题的时候,采用不同的算法求得的结果所使用的资源(如时间、空间等)也不尽相同,因此,需要确定一些用于衡量一个算法好坏的标准,现在最常用的标准就是算法的计算复杂性。算法计算复杂性(Computational complexity)是指研究计算算法执行过程中需要占用时间和空间等资源的大小,以及考虑尽可能节省这些资源的方法的问题,它是基于可计算性的理论而产生和发展起来的^[11]。

一般来说,设执行某个规模为 n 的工作任务,问题函数表示为 $\text{Task}(n)$,则研究计算复杂性从两方面度量标准进行:一是时间复杂性,即算法执行过程中计算需要的指令数或者步骤数,记为:

$$T(n) = O(\text{Task}(n)) \quad (1.5)$$

随着问题规模 n 的增长,执行算法耗费的的增长速度与 $Task(n)$ 的增长速度成正比。另一方面是空间复杂性,即算法执行过程中需要占用的存储单元数量,其表示方式与时间复杂性类似。

§ 1.1.3 NP 类问题

组合优化问题一般都具有十分精确的数学描述,计算过程很复杂,这类问题所具有的共同特点可归结为对于规模为 n 的问题,其问题空间是 n 个自然数全排列所构成的离散集合^{[12][13]},问题的规模越大,越难求解出最优解。该问题具有多个约束条件,且非凸、不连续、不可微、非线性、高维数,在求解最优值的过程中往往会产生大量的局部最优解。

定义 1.1 假设 A 为问题 P 的一个算法,如果对于求解问题 P 的一个实例时,整个执行过程每个步骤都只有一个选择,那么,称 A 为确定性算法。

定义 1.2 假设 A 为问题 P 的一个算法,如果对于求解问题 P 的一个实例时,算法有如下两阶段组成:

1. 猜测阶段:该阶段任意产生的一串字符串 S , S 可能是也可能根本就不是输入实例的解;
2. 验证阶段:该阶段首先对生成的字串 S 检验是否有适合的形式,如果有适合的形式,则算法接下来检验该字串是否属于实例的一个解,如果是实例的解,则算法停止并返回 YES,否则算法停止并返回 NO。

那么,将这类算法称为非确定性算法。

有定义可见,对于相同的输入数据,确定性算法执行多少次,所得到的结果都是不变的,而非确定性算法所得的结果则是无法预知的。

定义 1.3 多项式时间算法是指可以在输入规模为参变量的某个多项式时间,即算法的运行步数内得到结果的算法。

定义 1.4 可以在多项式时间内求解的判定(是或否)性问题称为 P 类问题。

定义 1.5 在给定正确信息的情况下,可以在多项式时间内验证确定解的问题称为 NP 问题。

由上述定义可得出,P 问题是复杂性为多项式时间的可确定性问题。对于那些很难找到或者根本就不存的多项式时间的问题,如果给出该问题的一个解,则可以在多项式时间内验证这个解是否正确,这一类问题属于 NP 类问题。很明显,所有的 P 类问题都属于 NP 类问题,但是对于 P 类问题是否等于 NP 类问题或者是否不等于 NP 类问题目前还尚未有定论。NP 问题被发现后,逐渐出现了一些特殊的 NP 问题,这类问题就是 NP-Complete 问题。NP-Complete 问题又被称为 NPC 问题或者 NP 完全问题,它是 NP 类问题中最难决定的问题。

定义 1.6 如果 NP 中的每个问题 P, P 的多项式时间都归约到一个判定问题 Q,则称 Q 是 NP-Hard 的。

定义 1.7 如果一个确定性问题属于 NP-Complete 问题,则:

1. 它属于 NP 问题,并且
2. 它属于 NP-Hard 问题

NP-Complete 问题都可以进行归约(Reduction),即如果一个算法适用于一类当中的某个特定的 NP-Complete 问题,那么这个算法也适用于这类的所有问题。从 Stephen A Cook 在 1971 年发现第一个 NP-Complete 问题,目前已经有近 4 000 多个问题被证明是属于 NP-Complete 问题,如哈密尔敦回路问题(Hamiltonian Cycle Problem)、布尔可满足性问题(Boolean Satisfiability Problem)、旅行商问题(TSP)、最小边覆盖问题等等^{[14][15][16][17]}。NP-Complete 问题在工程、管理等很多学科领域都有着广泛的应用,因此对于该类问题的研

究有着重要的理论和实际价值。

关于 P 类, NP 类, NP-Complete 类和 NP-hard 类问题有以下一些结论:

定理 1.1 设 B 是多项式时间算法, 如果 A 可以归约到 B, 则 A 也是多项式时间算法。

定理 1.2 如果 A 是 NP-Complete 问题, 且 A 多项式时间内可以归约到 B, 则 B 是 NP-Complete 问题。

该定理说明, 如果某个 NP-Complete 问题属于多项式时间算法, 则该 NP 类的包含的其他问题也属于多项式时间算法, 如果某个 NP-Complete 问题不属于多项式时间算法, 则该类的其他问题也不属于多项式时间算法, 即全部 NP-Complete 问题都是难度相同的。

定理 1.3 如果 A 是一个 NP-Hard 问题, 并且 A 多项式时间内可归约到 B, 则 B 也是 NP-Hard 的。

该定理说明, 如果 A 多项式时间内可归约到 B, 则 A 与 B 同样不易解。

定理 1.4 $P=NP$ 当且仅当存在一个 NP-Complete 问题, A 是多项式时间算法。

该定理表明, NP 类问题是否具有有效算法与 P 和 NP 是否相等有关, 当 $P=NP$ 时, 具有多项式时间算法; 否则, 不具有多项式时间算法。

§ 1.2 植物生长的生物学特征

植物的拓扑结构可表示为植物的根、节点、主枝干、侧枝干以及植物顶端的选段序列。按照 L-系统的理论, 植物生长是一个规律分解的过程, 植物的生成是一个纵向或者横向的拓扑结构, L-系统通过使用一套子定义的字符串文法和一组对应的替换产生式规则, 将这些规则文法进行分析和替换, 进行

有限次的迭代后,就可以描绘出植物产生、分枝,节间生长等整个植物的生长过程,生成复杂图形。PGSA 根据植物在生长过程中是遵循向光性的动力生长机制,来模拟植物从种子发芽,一直到长满整个生长空间为止这一过程而提出的优化方法。

L-系统揭示了复杂的自然界现象总可以找到简单的方法来对其进行描述。它是一个由字符串开始并对其进行有限次重写的并行系统。促使植物生长的主要动力来自于阳光,阳光对植物的光合作用会导致植物各个生长点的形态素浓度不断变化,背着阳光的部分细胞分裂得更快,生长素浓度也更大,这样背着阳光的地方生长得就快,植物就会向着对着阳光的一面弯曲生长,这就显示了植物的向光性生长的特性,植物的向光性生长是最普遍的植物生长规律。植物在生长的过程中,茎叶的生长情况主要是由一种称作“生长素”的浓度来决定,而生长素浓度的大小则跟茎叶所接受到的阳光资源有关,这种向光性的生长动力不但促进了植物茎叶向各个方向生长,也决定了处于各个方向的茎叶不同的生长状态。形态素浓度值大的生长点较浓度小的生长点具有优先生长机会,而植物形态素浓度的大小主要由植物的向光性决定,同时,植物各个茎和枝干的形态素浓度随着环境位置的变化在生长出新枝后会在各生长点之间重新分配。随着植物的生长素不断累积,面对着阳光照射充足的形态素浓度大的节点会优先获得生长权力,长出新的枝干,那些受阳光照射少而导致形态素浓度过低的节点则在竞争当中慢慢被淘汰。

植物的生长过程首先是从种子开始,生长出第一个主茎,要是植物主茎上的生长点不止一个,那么就根据各个生长点的形态素浓度的大小,来决定哪一个是可以继续生长新枝干的生长点。植物的生长过程其实就是由种子生出茎,茎再生出枝,枝再生出新枝,这样一个反复迭代的过程。利用模拟植物生长算法求解优化问题,就是模拟植物枝干长满整个生长空间的过程,根据植物生长的内在动力及向光性的作用力,建立茎、枝干繁殖生长及凋谢的

动力机制,将植物的整个生长空间看作解的可行域,光源当作全局最优解,植物的动力生长机制由其向光性决定,根据植物学中的形态素浓度理论建立不同光线强度的环境下按照全局最优的方式向着光源快速生长的动力模型。

§ 1.3 模拟植物生长算法的思想起源

目前,随着许多新的组合优化问题的不断出现,一些传统的求解模型和算法已经不再适用,这种情况下,对于原模型或算法的改进方法,以及新方法的研究具有十分重要的现实意义;模拟植物生长算法(Plant Growth Simulation Algorithm, PGSA)是受到生物界中植物或者树木在长大过程中所遵循的一些生长规律的启发,而产生继而被提出一种智能优化搜索算法。1968年,Aristid Lindenmayer 提出了一系列可以对多细胞生命体进行模拟和建模的规则,即 L-系统^{[117][118]}。L-系统是用来表达植物形态结构和生长形态的重要方法,它是描绘虚拟植物,以及模拟植物生长和农作物培育的手段^[119]。L-系统揭示了在复杂的自然现象的背后存在着简单的规律,随着这些规律的深入研究,一些学者将其中的一些规律应用在智能搜索上,并取得了明显的成效。

§ 1.4 研究模拟植物生长算法的意义

模拟植物生长算法是根据 Przemyslaw Prusinkiewicz 等人在 1968 年提出的一个计算机模拟植物生长系统(L-systems)而演化而来的。按照 L-系统的理论,植物生长是一个规律分解的过程,植物的生成是一个纵向或者横向的拓扑结构,L-系统通过使用一套子定义的字符串文法和一组对应的替换产生式规则,将这些规则文法进行分析和替换,进行有限次的迭代后,就可以描绘出植物产生、分枝,节间生长等整个植物的生长过程,生成复杂图形。模拟植