

DNA 计算机中图的深度优先搜索遍历算法

魏国辉^{1,2}, 杨春德¹, 谭军²

(1. 重庆邮电大学计算机学院, 重庆 400065; 2. 重庆邮电大学生物信息学研究所, 重庆 400065)

摘要:提出 DNA 计算机中图数据结构的一种设计方法, 给出具体的存储结构以及深度优先搜索遍历的算法。该算法实现了在 DNA 计算机下图元素的遍历。为证明其可行性, 给出一个具体的算法实例, 描述了 DNA 计算机上的运行机制。依据分子生物学的理论, 证明算法是有效且可行的。

关键词: DNA 计算机; 图; 深度优先搜索; 遍历

Depth_first Search Traversing Algorithm of Graphs in DNA Computer

WEI Guo-hui^{1,2}, YANG Chun-de¹, TAN Jun²

(1. College of Computer Science, Chongqing University of Posts and Telecommunications, Chongqing 400065;

2. Institute of Bioinformatics, Chongqing University of Posts and Telecommunications, Chongqing 400065)

【Abstract】 The design of a graph data structure, the memory structure and the depth_first search traversing algorithm are proposed. This algorithm can make the graph implement traversing in DNA computer. In order to prove the feasibility of the method, an actual example is given out, which describes the implementation of the algorithm. On the basis of the experimental theory of bio-molecular, the algorithm is an effective and feasible method.

【Key words】 DNA computer; graph; breadth_first search; traversing

1 概述

DNA 计算是近年来计算机研究领域的一个热点方向, 在近几年分子生物计算机的研究中倍受学者们的关注^[1-3]。基于分子生物技术的 DNA 计算是一种模拟生物分子 DNA 的结构并借助于生化反应作为计算工具的超大规模并行计算, 而且 DNA 的双螺旋结构具有巨大的信息存储容量, 这也正是人们对 DNA 计算机产生极大兴趣之所在。自从 1994 年 Adleman^[4] 开创 DNA 计算领域以来, 1995 年 Lipton 等人^[5] 在 Adleman 的试验基础上, 解决了更有趣的 NP-完全问题(可满足问题)。1997 年, Ouyang 等人^[6] 利用 DNA 计算解决了另一个 NP-完全问题——图的最大团问题。2000 年, Wisconsin 大学的 Liu 等人^[7] 介绍了一种基于表面的 SAT 问题的解决方法。2004 年, Shapiro 等人^[8] 又提出了一种用于控制基因表达的分子自动计算机并应用于疾病诊断治疗。这为 DNA 计算向 DNA 计算机的发展提供了进一步的理论依据。

DNA 计算机的最终目的就是能够进行数据的处理, 这需要类似与电子计算机一样的数据结构来进行数据的组织。在东华大学丁永生教授^[9]、李汪根博士^[10] 对堆栈以及队列数据结构的研究基础上, 本文深入研究了 DNA 计算机中图元素的编码结构以及对图中元素的遍历操作, 并给出了算法实例。

2 基于质粒 DNA 计算的图存储结构的设计

2.1 图的概念^[11]

图的概念为: 图 G 是一个三元组, 记作 $G = \langle V(G), E(G), \varphi(G) \rangle$ 。其中,

(1) $V(G) = \{v_1, v_2, \dots, v_n\}, V(G) \neq \emptyset$, 称为图 G 的结点集合。

(2) $E(G) = \{e_1, e_2, \dots, e_m\}$ 是 G 的边集合, 其中, 若 e_i 为 $\{v_j,$

$v_i\}$, 则称 e_i 是以 v_j 和 v_i 为端点的无向边; 若 e_i 为 $\langle v_j, v_i \rangle$, 则称 e_i 是以 v_j 为起点, v_i 为终点的有向边。

(3) $\varphi(G) : E \rightarrow V \times V$ 称为关联函数。

图是一种比较复杂的数据结构。在图形结构中, 结点之间的关系可以是任意的, 图中任意 2 个数据元素之间都可能相关。图分为有向图与无向图, 在本文中所说的图都是指无向图, 如图 1 所示。

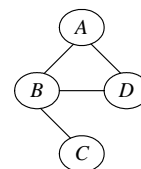


图 1 无向图

图的遍历是指从图中某一顶点出发访问其余顶点, 且使每一个顶点仅被访问一次。通常图的遍历有 2 种: 深度优先搜索遍历和广度优先搜索遍历。笔者以深度优先搜索为例来说明 DNA 计算机中的图遍历的实现。如图 1 所示, 深度优先搜索过程为: 从图中某个顶点 v 出发, 访问此顶点, 然后依次从 v 的未被访问的邻接点出发深度优先遍历图, 直至图中所有和 v 有路径连通的顶点都被访问到。若此时图中尚有顶点未被访问到, 则另选图中一个未曾被访问的顶点作起始点, 重复上述过程, 直到所有点都被访问过。图的遍历算法是求解图的连通性问题、拓扑排序和关键路径等算法的基础。

2.2 质粒 DNA^[12]计算的图的存储结构

由于图的复杂性, 直接导致了它的存储结构的复杂。图没有顺序映像的存储结构, 其存储结构包括数组表示法和链

基金项目: 国家自然科学基金资助项目(30600329)

作者简介: 魏国辉(1983 -), 男, 硕士研究生, 主研方向: 网络图论, 生物信息学, DNA 计算; 杨春德, 副教授、硕士; 谭军, 教授、在职博士研究生

收稿日期: 2008-01-15 **E-mail:** weigh0501@163.com

表表示法。其中，链表又分为邻接表、十字链表、邻接多重表等。本文设计了一种类邻接表的存储结构，用来表示数据元素之间的关系。

使用质粒 DNA 存储图的信息时必须考虑图的各个顶点之间的关系。在此采用类邻接表的结构存储，将某一顶点元素作为头结点与其所有的邻接顶点放在同一条 DNA 链中，每一个顶点对应一条链，如图 1 则可以设计 4 条链来存储相应的信息。然后将 4 条链导入 4 个质粒中，以备操作。

3 DNA 计算机编码

3.1 图元素编码

对图 1 中的每个顶点元素进行编码，每个元素用 20 bp~30 bp 的核苷酸片段编码，这主要是为了考虑 DNA 链的特异性，同时每个元素的 DNA 链的两边都设计了粘性末端以便在存储结构中进行插入与删除。4 个元素的粘性末端分别对应 Pst I, Sal I, EcoR I, Kpn I 4 个限制性内切酶切后的序列，如图 2 所示，代表元素编码，方框左右序列为粘性末端。

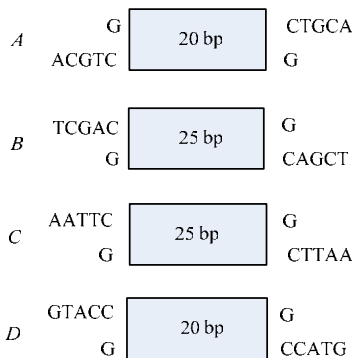


图 2 ABCD 4 元素编码

3.2 存储结构编码

对图的元素进行编码之后，需要根据上面提到的存储结构以及图元素之间的关系，在 DNA 计算机中进行 DNA 链的编码来存储图的信息，以便进行相应的操作。如图 3 所示，B 的存储结构包括以元素 B 为头结点，B 的邻接点组成的 DNA 链。为了便于进行 DNA 链的剪切操作，在各个元素的两端设计了酶切序列位点。对应的限制性内切酶为 Sal I, Pst I, EcoR I, Kpn I。图 4 依次为对应的限制性内切酶 Sal I, Pst I, EcoR I, Kpn I 的识别序列，箭头表示酶切位点。合成后分别以 ABCD 为头结点，与其邻接点组成的 4 条 DNA 链导入 4 个质粒，分开顺次保存在标为 a, b, c, d 的 4 个试管中，然后分别转入大肠杆菌扩增这样的质粒以便遍历时进行操作。除了规定位置，质粒其他地方不能出现限制性内切酶识别序列。这样就在 DNA 计算机中存储了图 1 的无向图的信息。

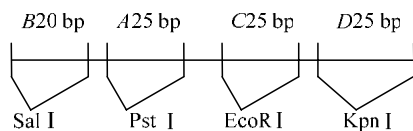


图 3 B 与其邻接点的 DNA 编码

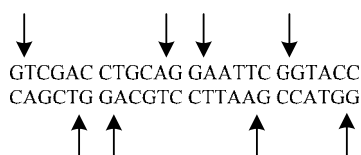


图 4 限制性内切酶的识别序列

3.3 Detector 分子设计编码

另一方面，为每个元素设计了一个唯一的 Detector 分子，可以从当前被连接上的 Detector 分子得知当前已被访问图元素的值，用来表示该元素已被访问过。Detector 分子右端是一个 Primer，在实验中可以进行扩增，达到实验要求。将这些 Detector 分子添加到 T 试管中。如图 5 所示。

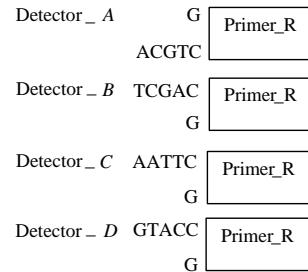


图 5 Detector 分子

4 深度优先搜索算法实例

以图 1 为例来描述深度优先搜索算法的操作。

(1)在 a 中加入内切酶 Pst I，将 A 切下，访问 A，然后通过凝胶电泳技术分离出 A，转入 T 中。质粒重新环化。同时，在 T 中使用连接酶 T4 将 A 与其 Detector 分子连接。A 的邻接点是 B, D，同样的方法，使用 Sal I 将 B 切下，然后访问 B，分离出去转入 T 中。质粒重新环化。在 T 中使用连接酶 T4 将 B 与其 Detector 分子连接标记已被访问。

(2)B 的邻接点在 b 中，首先将 B 切下，B 已被标记。则将 B 分离出去。质粒重新环化。再将邻接点 A 切下，检查 T 中 A 已被标记，则将 A 分离出去。质粒重新环化。然后切下 C，检查 T 中未被标记，访问 C，然后将 C 转入 T 中，质粒重新环化。在 T 中使用连接酶 T4 将 B 与其 Detector 分子连接标记已被访问。

(3)C 的邻接点在 c 中，首先将 C 切下，C 已被标记。则将 C 分离出去。质粒重新环化。再将 B 切下，检查 T 中 B 已被标记，则将 B 分离出去。质粒重新环化。

(4)检查 abcd 试管中的尚有的元素。在 a 中使用 Kpn I 将 D 切下，检查 T 中 D 未被标记，访问 D，然后将 D 转入 T 中，质粒重新环化。在 T 中使用连接酶 T4 将 D 与其 Detector 分子连接标记已被访问。

(5)D 的邻接点在 d 中，首先将 D 切下，D 已被标记。则将 D 分离出去。质粒重新环化。再将邻接点 A 切下，检查 T 中 A 已被标记，将 A 分离出去。质粒重新环化。然后将 B 切下，检查 T 中 B 已被标记，则将 B 分离出去。质粒重新环化。

(6)检查在 b 中，还有 D，由于在 T 中已标记访问，则将 D 分离出去。质粒重新环化。

这样，最后图 1 中的元素全部被访问过，从而实现了元素的深度优先搜索遍历。

5 结束语

本文研究了 DNA 计算机中图元素的编码结构以及对图中元素的遍历操作，并通过生物学操作的例子，模拟 DNA 计算机中深度优先搜索算法，由分子生物学理论证明这是可以实现的。

从上面的算法例子也可以看出，电子计算机中能够实现的算法在 DNA 计算机中同样可以实现，而且在电子计算机中使用的递归方法在 DNA 计算机中同样适用。如在上文的

(下转第 258 页)