

粘贴与删除系统求解最短有向路的 DNA 计算模型

马芳芳¹, 王淑栋¹, 李涵², 薛圣伟¹

MA Fang-fang¹, WANG Shu-dong¹, LI Han², XUE Sheng-wei¹

1. 山东科技大学 信息科学与工程学院, 山东 青岛 266510

2. 山东科技大学 信息系, 山东 泰安 271209

1. College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong 266510, China

2. College of Information, Shandong University of Science and Technology, Taian, Shandong 271209, China

E-mail: mff007007@sina.com

MA Fang-fang, WANG Shu-dong, LI Han, et al. DNA computing model for shortest directed path problem based on sticker and delete system. *Computer Engineering and Applications*, 2008, 44(25): 40-42.

Abstract: Shortest directed path problem is finding a directed path with minimum weight in two pointed vertex of a directed network. It has extensive application in engineering practice. Sticker system and delete system are two kinds of basic models in DNA computing form model. This paper proposes DNA computing model and biochemical procedures of graph shortest directed path problem using high parallelism of sticker system and delete system.

Key words: DNA computing; sticker system; delete system; shortest directed path problem

摘要: 最短有向路问题是在一个有向网络中的两个指定顶点之间找出一条具有最小权的有向路, 它在工程实践中具有广泛的应用。粘贴系统与删除系统是 DNA 计算形式模型中的两种基本模型。论文利用粘贴与删除系统的巨大并行性给出了求解图最短有向路问题的 DNA 计算模型及其实现算法。

关键词: DNA 计算; 粘贴系统; 删除系统; 最短有向路问题

DOI: 10.3778/j.issn.1002-8331.2008.25.012 文章编号: 1002-8331(2008)25-0040-03 文献标识码: A 中图分类号: O22

21 世纪是生物科学的世纪, 且随着晶体管的宽度逐渐接近极限, DNA 计算已成为人们普遍关注的一种计算模式。图论中的许多 NP-完全问题利用 DNA 计算取得了不错的结果, 如有向图的哈密尔顿路径问题、图的最大团问题、最大独立集问题、图着色问题等^[1-3]。

设 $G=(V, E, W)$ 是一个有向网络, 其中 V, E, W 分别是 G 的顶点集, 边集和权集。最短有向路问题就是在有向网络中的两个指定顶点之间找出一条具有最小权的路。本文涉及的图都是权值为正整数的严格有向图。2004 年 Zuwairie Ibrahim 等人^[4]提出了求解最短有向路问题的一种 DNA 算法。2005 年 Zuwairie Ibrahim 等人^[5]利用 DNA 计算模拟了 Dijkstra 算法求解最短有向路问题。

设 V 是字母表, ρ 是 V 上的对称关系, 即对字母表 V 中的任意元素 x, V 中存在另一个元素 x' , 满足 $\rho(x)=x'$ 且 $\rho(x')=x$ 。 $L_\rho(V), R_\rho(V), LR_\rho(V)$, 分别表示左端为单链右端为双链, 左端为双链右端为单链, 两端为单链中间为双链的不完整双链分子

构成的集合。令 $W_\rho(V)=L_\rho(V)\cup R_\rho(V)\cup LR_\rho(V)$, $WK_\rho(V)=\left[\begin{matrix} V \\ V \end{matrix}\right]_\rho^*$ 。

元素 $\left[\begin{matrix} w_1 \\ w_2 \end{matrix}\right] \in WK_\rho(V)$ 称为双链序列或分子, 其中 $w_1, w_2 \in V^*$, w_1 与 w_2 长度相等且满足 V 上的对称关系 ρ 。

粘贴运算是在不完整分子集合 $W_\rho(V)$ 中, 模拟生物操作中的连接操作与退火操作技术产生语言的运算, 它是粘贴系统的主要运算。 $x=x_1x_2x_3 \in W_\rho(V), y \in W_\rho(V)$ 。 x, y 的粘贴运算如图 1 所示。

粘贴系统具有结构 $\gamma=(V, \rho, A, D)$, 其中 V 是字母表, $\rho \subseteq V \times V$ 是对称关系, $A \subseteq LR_\rho(V), D \subseteq W_\rho(V) \times W_\rho(V)$ 。称 $L_n(\gamma)=$

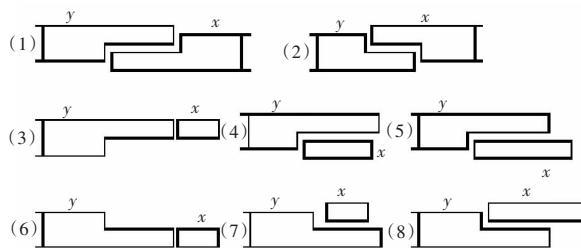


图 1 粘贴运算说明图示

基金项目: 国家自然科学基金(the National Natural Science Foundation of China under Grant No.60503002); 中国博士后科学基金(the China post-doctor Foundation of China under Grant No.20060400344)。

作者简介: 马芳芳(1982-), 女, 助教, 从事 DNA 分子生物计算领域研究; 王淑栋(1973-), 女, 北京大学博士后, 山东科技大学副教授, 从事 DNA 分子生物计算领域研究; 李涵(1980-), 男, 助教, 从事信息管理领域研究; 薛圣伟(1980-), 男, 研究生, 从事 DNA 分子生物计算领域研究。

收稿日期: 2007-12-06 **修回日期:** 2008-02-27

$\left\{w \in V^* \left| \begin{bmatrix} w \\ w \end{bmatrix} \in LM_n(\gamma), w \in V^* \right. \right\}$ 为粘贴系统 γ 产生的语言,其中 $LM_n(\gamma)=\{w \in WK_p(v) \mid x \Rightarrow^* w, x \in A\}, x \Rightarrow y$ 当且仅当对 $(u, v) \in D, y = \mu(u, \mu(x, v)), x, y \in LR_p(V)$ 。 $x \Rightarrow^* w$ 表示从 x 出发,经过 D 上的一对多米诺骨排 (u, v) 作用,生成 $WK_p(V)$ 中的完整双链分子。

删除运算是删除系统的主要运算,通过错配与退火操作技术产生一种语言。删除运算可解释如下: 设 $\alpha=xuvwyz$ 是一条 DNA 单链,其中子链 x, u, w, v, y, z 都是 DNA 单链, α 与另外一条 DNA 单链 \bar{w} 杂交,其中 \bar{u}, \bar{v} 分别是单链 u, v 的补, \bar{u} 与 u 结合, \bar{v} 与 v 结合, 折叠 w 。用内切核酸酶去掉环 w , 加入单链 $\bar{x}, \bar{y}, \bar{z}$ 和聚合酶, 得到完整双链, 加热通过变性使双链变成单链 $xwyz$ 。上述运算删除了 α 中的单链序列 w 。

删除系统 $\gamma=(V, T, A, R)$, 其中 V 是字母表, $T \subseteq V, A \subseteq V, R$ 中元素为三元组 $(u, \alpha/\lambda, v)$, 其中 $u, v \in V^*, \alpha \in (V^* \times \{\lambda\}) \cup (\{\lambda\} \times V^*)$ 。 $(u, \alpha/\lambda, v)$ 表示在上下文 (u, v) 中删除 α , 简记为 $u\alpha v \rightarrow w$ 。 $L(\gamma)=\{\omega \in T^* \mid x \Rightarrow^* \omega, x \in A\}$ 为删除系统 γ 产生的语言, 其中 $x \Rightarrow^* w$ 表示从 x 出发, 经过 R 中删除规则作用, 生成 T^* 中单链分子。

文中未介绍的术语与记号见参考文献[8]。

1 系统的构造

设 $G=(V, E, W)$ 是一个有向网络, 其中 V, E, W 分别是 G 的顶点集, 边集和权集, 其中 $V=\{v_1, v_2, \dots, v_n\}, E=\{e_1, e_2, \dots, e_m\}, W=\{w(e), e \in E\}, w(e)$ 是边 e 上的权值。下面构造能模拟上述有向网络 G 中从顶点 v_1 到顶点 v_n 的最短有向路的生成系统。

1.1 粘贴系统的构造

构造粘贴系统 $\gamma_1=(V_1, \rho, A_1, D)$, 其中 $V_1=V \cup W \cup \{M\}, M$ 是一个符号, 对称关系 $\rho(x)=x, x \in V_1, A_1=\left\{\begin{bmatrix} M \\ M \end{bmatrix} \left(\begin{matrix} v_1 w_{i_1} \\ \lambda \end{matrix} \right) \mid w_{i_1} \in W \right\}$, $D=\left\{\left(\begin{matrix} M v_n M \\ \lambda \end{matrix} \right), \left(\begin{matrix} \lambda \\ v_n M \end{matrix} \right), \left(\begin{matrix} M v_i w_{i_{i_1}} \\ \lambda \end{matrix} \right), \left(\begin{matrix} \lambda \\ v_i w_{i_{i_1}} M \end{matrix} \right) \mid v_i \in V, w_{i_{i_1}} \in W \right\}$, $L_n(\gamma_1)=\{M v_1 w_{i_1} \dots M v_i w_{i_{i_1}} \dots M v_n M \mid v_i \in V, w_{i_{i_1}} \in W\}$ 。

1.2 删除系统的构造

构造删除系统 $\gamma_2=(V_2, T, A_2, R)$, 其中 $V_2=V_1, R=\{w_{i_{i_1}}, M v_{i_{i_1}}/\lambda, w_{i_{i_1} i_{i_2}}, A_2=L_n(\gamma_1), L_n(\gamma_2)=\{M v_1 w_{i_1} w_{i_{i_1} i_{i_2}} \dots w_{i_n} M v_n M \mid w_{i_1}, w_{i_{i_1} i_{i_2}}, w_{i_n} \in W\}$ 。

2 最短有向路 DNA 算法

2.1 粘贴与删除系统求解最短有向路问题算法

步骤 1 利用粘贴系统 γ_1 生成形如 $M v_1 w_{i_1} \dots M v_i w_{i_{i_1}} \dots M v_n M$ 的单链分子。

步骤 2 利用删除系统 γ_2 生成形如 $M v_1 w_{i_1} w_{i_{i_1} i_{i_2}} \dots w_{i_n} M v_n M$ 的单链分子。

步骤 3 读出 $M v_1 w_{i_1} w_{i_{i_1} i_{i_2}} \dots w_{i_n} M v_n M$ 长度, 最短的就是所求最短有向路。

2.2 求解最短有向路问题的 DNA 算法的实现

对粘贴系统 γ_1 中的 A_1 与 D 中元素进行编码。用长度为 l 的 DNA 序列编码 $v_i(i=1, 2, \dots, n)$ 与 $M(l \in Z^+)$ 。 w_{ij} 的编码长度

取决于 w_{ij} 的值, 如 $w_{ij}=5$, 用长度为 5 的 DNA 序列编码 w_{ij} 。

2.2.2 生物操作步骤

步骤 1 将编码好的 D 中单链分子放入含有不完整双链分子 $\begin{bmatrix} M \\ M \end{bmatrix} \left(\begin{matrix} v_1 w_{i_1} \\ \lambda \end{matrix} \right)$ 的试管中, 加入缓冲液和连接酶, 不完整分子向右进行延伸操作, 生成初始数据池。利用探针分离出带有粘性末端的不完整双链分子, 加热试管生成形如 $M v_1 w_{i_1} M \dots v_i w_{i_{i_1}} M \dots M v_i w_{i_n} M v_n M$ 的单链分子。

步骤 2 在包含 $M v_1 w_{i_1} M \dots v_i w_{i_{i_1}} M \dots M v_i w_{i_n} M v_n M$ 的试管中加入每条边权 w_{ij} 的补 $\bar{w}_{ij}(i, j=1, 2, \dots, n)$ 与内切核酸酶、连接酶去除 $M v_1$ 与 $v_n M$ 之间的顶点得到形如 $M v_1 w_{i_1} w_{i_{i_1} i_{i_2}} \dots w_{i_n} M v_n M$ 的单链分子。

步骤 3 将这些单链分子通过琼脂糖凝胶电泳, 找出长度最短的 DNA 分子, 然后进行测序, 读出顶点顺序, 这个顶点顺序即为最短有向路路径。

3 实例分析

3.1 实例

为了说明上述算法的可行性, 以图 2 为例, 求解图 2 中以 v_1 为始点, v_5 为终点的最短有向路。

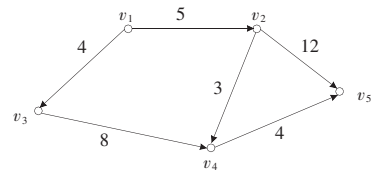


图 2 简单赋权有向图

3.2 编码与生物操作

3.2.1 编码

对粘贴系统 γ_1 中的 M 与图 2 中 5 个顶点、每条边上的权进行编码, 编码如下:

- M : ATGCCA v_1 : ATCGGA v_2 : GTCTGA
- v_3 : AGTGAC v_4 : CATCGT v_5 : GACTGA
- w_{12} : ATGAC w_{13} : TTCA w_{24} : GAA
- w_{25} : TTCCGATCGCTC w_{34} : GTCCGCCA w_{45} : CTAC

利用 DNA 合成器合成 A_1 中不完整双链分子 $\begin{bmatrix} M \\ M \end{bmatrix} \left(\begin{matrix} v_1 w_{12} \\ \lambda \end{matrix} \right)$,

$\begin{bmatrix} M \\ M \end{bmatrix} \left(\begin{matrix} v_1 w_{13} \\ \lambda \end{matrix} \right)$ 与 D 中单链分子 $\left(\begin{matrix} \lambda \\ v_1 w_{12} M \end{matrix} \right), \left(\begin{matrix} \lambda \\ v_1 w_{13} M \end{matrix} \right), \left(\begin{matrix} M v_2 w_{24} \\ \lambda \end{matrix} \right), \left(\begin{matrix} M v_2 w_{25} \\ \lambda \end{matrix} \right), \left(\begin{matrix} M v_3 w_{34} \\ \lambda \end{matrix} \right), \left(\begin{matrix} M v_4 w_{45} \\ \lambda \end{matrix} \right), \left(\begin{matrix} M v_5 M \\ \lambda \end{matrix} \right), \left(\begin{matrix} \lambda \\ v_5 M \end{matrix} \right), \left(\begin{matrix} \lambda \\ v_2 w_{24} M \end{matrix} \right), \left(\begin{matrix} \lambda \\ v_2 w_{25} M \end{matrix} \right), \left(\begin{matrix} \lambda \\ v_3 w_{34} M \end{matrix} \right), \left(\begin{matrix} \lambda \\ v_4 w_{45} M \end{matrix} \right)$ 。

3.2.2 生物操作

下面分 3 步简单介绍求解图 2 中从顶点 v_1 到 v_5 的最短有向路问题的 DNA 生物操作过程。

步骤 1 将编码好的单链分子放入含有不完整双链分子的试管中, 加入缓冲液, 聚合酶, 单链分子与不完整双链分子退火, 在聚合酶的作用下进行延伸, 生成初始数据池。在初始数据池中, 除了经过粘贴运算形成的完整分子外, 还存在一定数量的、未完成计算的不完整分子, 可以使用探针将这些不完整分

子从数据池中删除。利用 M 对应的 DNA 序列的补序列 \bar{M} 作为模板设计出分离探针,并将其固定在试管壁上,将含有初始数据池的溶液置入含有分离探针的试管中,使得这些探针与残存的粘性末端充分杂交,然后将杂交后的溶液置入一试管,则这个新试管里的所有分子都是完整分子。加热变性(通常加热温度在 $85^{\circ}\text{C}\sim 95^{\circ}\text{C}$)生成单链分子 $Mv_1w_{i_1}\cdots Mv_iw_{i_{i_1}}\cdots Mv_5M$ 。

步骤 2 在上一步得到包含 $Mv_1w_{i_1}\cdots Mv_iw_{i_{i_1}}\cdots Mv_5M$ 的试管中加入每条边权 w_{ij} 的补 \bar{w}_{ij} 及内切核酸酶与 DNA 连接酶(通常反应在室温 $<30^{\circ}\text{C}$ 下进行),内切核酸酶可以切去不完整双链分子中的单链部分,即去除 Mv_1 与 Mv_5M 之间的 Mv_i ,采用聚丙烯酰胺凝胶电泳鉴定酶解产物,利用 DNA 连接酶得到单链分子 $Mv_1w_{i_1}w_{i_2}\cdots w_{i_5}Mv_5M$ 。

步骤 3 通过琼脂糖凝胶电泳,找出长度最短的 DNA 分子。为了进一步知道每个分子代表的路径情况,必须对 DNA 链进行序列测定,采用 Sanger 末端终止法进行测序,从凝胶的放射自显影片上直接读出 DNA 的核苷酸顺序,测出的最短有向路路径为 $v_1v_2v_3v_5$ 。

4 讨论

本文利用 DNA 计算形式模型中的粘帖与删除系统与电泳、酶切反应、探针等分子生物技术手段解决了最短有向路问题,下面对最短有向路问题的边权给出两点说明:

(1) 当图的边权为小数时,可以采取放大若干倍使其变成整数的方法来解决。

(2) 当图的边权差异很小时,这是最困难的情况,需要寻找更加合理的编码方法来解决。

文中只解决了有向网络的最短有向路问题,当所研究图为无向图时,将图中指定顶点间的无向边用端点相同方向相反的两条有向边来代替,这样无向图的最短路问题就可转化为有向图的最短有向路问题,利用同样的编码与生物操作可以解决无

向图的最短路问题。

本文的算法是将数学问题的求解同生物技术密切结合起来,在具体的生物实现过程中仍有不足之处,例如内切核酸酶的切除可能会不完全。由此点可能导致非法解的出现,但随着现代分子生物学技术的发展,我们相信这个问题会逐步得到克服,使得差错率控制在一定的范围内。

文中讨论了有向网络的最短有向路问题的 DNA 算法,将粘帖与删除系统与分子生物操作结合起来求解最短有向路,对 n 个顶点、 m 条边的有向网络,编码共需 $n+m+1$ 条 DNA 单链分子,合成 $2m+2$ 条不完整双链分子或单链分子,为解决最短路问题提供了一种新的方法。

参考文献:

- [1] Adleman L. Molecular computation of solution to combinatorial problems[J]. Science, 1994, 66: 1021-1024.
- [2] Ouyang Q, Kaplan P D, Liu Shu mao, et al. DNA solution of the maximal clique problem[J]. Science, 1997, 278: 446-449.
- [3] Liu Q H, Guo Z, Condon A E, et al. DNA computing on surfaces[J]. Nature, 2000, 403: 175-179.
- [4] Head T, Kaolan P D, Bladergroen R R, et al. Computing with DNA by Operating on Plasmids[J]. Biosystem, 2000, 57: 87-93.
- [5] Liu WB, Xu J A. DNA Algorithm for the graph coloring problem[J]. Journal of Chemical Information and Computers, 2002, 42: 1176-1178.
- [6] Paun G, Rozenberg G, Salomaa A. DNA computing: New computing paradigms[M]. Berlin Heidelberg: Springer Verlag, 1998: 30-48.
- [7] Ibrahim Z, Tsuboi Y. Molecular computation approach to compete Dijkstra's algorithm[C]//5th Asian Control Conference, 2004: 22-28.
- [8] Ibrahim Z, Tsuboi Y. A study on lower bound of direct proportional length-based DNA computing for shortest path problem[C]//LNCS 3314: CIS 2004. Berlin Heidelberg: Springer-Verlag, 2004.

(上接 8 页)

通道和 cache,以加快数据传输速度等。经优化,系统的运行速度可达 $11\text{ f/s}\sim 14\text{ f/s}$,基本满足实时监控的需要。图 6 是系统工作的截图。



图 6 监控平台截图

5 结论及展望

智能监控是一个具有挑战性的研究领域,既有理论研究意义,也有很强的实用价值。但由于使用背景、具体环境等因素的复杂性,目前仍需依赖人工进行不间断分析,还未能找到“通

用”的解决算法。本文主要针对室内监控这一特定背景环境,对运动目标的跟踪监控进行了研究,提出了自适应更新的 Mean-shift 方法,并设计搭建了测试平台。实验表明,该平台可以有效地检测并跟踪室内运动目标,实现了部分分析任务的自动处理。如何分类并识别运动物体将是下一步研究的重点。

参考文献:

- [1] 方帅. 计算机智能视频监控系统关键技术研究[D]. 沈阳: 东北大学, 2005.
- [2] 王亮, 胡卫明, 谭铁牛. 人运动的视觉分析综述[J]. 计算机学报, 2002, 25(3): 225-237.
- [3] Texas Instrument Co. TMS320DM642: Video imaging fixed-point digital signal processor[EB/OL]. (2007-01). <http://www.ti.com/lit/gpn/tms320dm642>.
- [4] Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking[J]. IEEE transactions on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564-577.
- [5] Aach T, Kaup A, Mester R. Statistical model-based change detection in moving video[J]. Signal Processing, 1993, 31(2): 165-180.