

一种基于分治的三维匹配问题 DNA 计算算法

周 旭¹,李肯立²,乐光学^{1,2},杨志邦²

(1. 嘉兴学院数学与信息工程学院,浙江嘉兴 314001; 2. 湖南大学计算机间与通信学院,湖南长沙 410082)

摘 要: 本文基于 Aldeman-Lipton 模型的生物操作与粘贴模型的解空间,提出一种三维匹配问题的 DNA 计算新模型;同时基于此模型和传统计算机中分治策略,提出一种求解三维匹配问题的 DNA 计算新算法.将提出的算法与已有文献结论的对比分析表明:本算法将穷举算法中的 DNA 链数从 $O(2^n)$ 减少至 $O(2^{n/2}) \approx O(1.414^n)$,同时生物操作数由 $O(n^2)$ 减少至 $O(15n + 30q)$,测试试管数由所需的 $O(n)$ 减少至 $O(1)$,最大链长由 $O(15n + 45q)$ 减少至 $O(15n/2 + 45q)$.因此,本算法理论上在试管级生化反应条件下能将求解三维匹配问题的规模从 $67(2^{67} \approx 10^{22})$ 提高到 $134(67 \times 2 = 134)$.同时,与传统的穷举搜索算法相比,该算法具有高效的空間利用率及容错技术的优点.

关键词: DNA 计算; 三维匹配问题; 分治策略; NP 完全问题

中图分类号: T18 **文献标识码:** A **文章编号:** 0372-2112 (2010) 08-1831-06

A New DNA Computing's Algorithm for 3-Dimensional Matching Problem Based on Divide and Conquer Strategy

ZHOU Xu¹, LI Ken-li², YUE Guang-xue^{1,2}, YANG Zhi-bang²

(1. College of Mathematics and Information Engineering, Jiaxing University, Jiaxing, Zhejiang 314001, China;

2. School of Computer and Communications, Hunan University, Changsha, Hunan 410082, China)

Abstract: At present, it is how to decrease the DNA volume that plays an important role in the development of DNA computing. In this paper, for the objective to reduce the DNA volume of 3-Dimensional Matching Problem which is a famous NP-complete problem, an improved DNA computing model based on the operations in Adleman-Lipton's model and the solution space of the sticker-based model is put forward. Then, the divide and conquer is introduced into the DNA computing and a new DNA computing's algorithm is proposed. In a computer simulation, the DNA strands of maximum number required was $O(1.414^n)$, the time complexity was $O(15 \times n + 30 \times q)$, the test tube complexity was $O(1)$ and the longest DNA strand was $O(15n/2 + 45q)$. Hence, the scale of 3-Dimensional Matching Problem may be enlarged from 67 to 134. This new algorithm is highly space-efficient and error-tolerant compared to the conventional brute-force searching, and can be scaled-up to solve large and hard 3-Dimensional Matching Problem. By the approach, we can also show DNA computing's vast potentials for resolving NP problems.

Key words: DNA computing; 3-dimensional matching problem; divide and conquer; NP-complete problem

1 引言

1994 年 Adleman^[1]开创性地提出了 DNA 计算模型之后, DNA 计算与 DNA 计算模型的研究形成了理论计算机科学、数学和生物学领域的一个新的研究热点,吸引了众多计算机科学家和生物学家的研究兴趣,用 DNA 计算解决的数学问题的种类也迅速增长^[2~10].

DNA 计算以其具有的海量存储和并行运算能力从理论上可克服电子计算机存储量小与运算速度慢的不足^[11]. 目前为止,一个测试试管已可产生 10^{22} 条 DNA 链,它可使 10^{22} 位数据以数据并行的方式并行运行^[12].

因此, DNA 计算可提供相当于 10^{22} 个处理单元的并行性和 $O(10^{22})$ 的存储空间^[13,14].

三维匹配问题(3-Dimensional Matching Problem, 简称 3-D Matching Problem)^[15] 作为六个经典的 NP 完全问题之一,在分配,调度,交通和网络流等领域有着广泛的应用.然而现有求解大规模三维匹配问题的方法在传统计算机上的计算时间都不能令人满意.

文献[16]提出的三维匹配问题的 DNA 计算算法基于穷举算法,算法所需 DNA 分子链数为 $O(2^n)$.从公开出版的文献看,目前尚无解决三维匹配问题既具较低 DNA 空间复杂性又具生化实验可行性的 DNA 计算算

法.为此,在文献[16]工作的基础上,本文对三维匹配问题的 DNA 计算算法中的可扩展性问题进行了较深入的探索,主要工作可概述为:

(1)为了提高求解三维匹配问题的 DNA 计算算法的可扩展性,本文基于 Aldeman-Lipton 模型的操作与粘贴模型的解空间,提出一种求解三维匹配问题的 DNA 计算新模型.

(2)借鉴传统计算中分治策略,提出一种三维匹配问题的 DNA 计算新算法,设计了解空间生成器,三维匹配预搜索器及三维匹配搜索器.

该算法在生物操作数由 $O(n^2)$ 减少至 $O(15n + 30q)$,测试试管数由所需的 $O(n)$ 减少至 $O(1)$,最大链长由 $O(15n + 45q)$ 减少至 $O(15n/2 + 45q)$,DNA 链数由 $O(2^n)$ 减小至 $O(1.414^n)$.

2 求解三维匹配问题的 DNA 计算模型

2.1 问题描述^[15]

给定具有 q 个元素互不相交的集合 X, Y, Z , 其中 $X = \{x_1, \dots, x_q\}$, $Y = \{y_1, \dots, y_q\}$, $Z = \{z_1, \dots, z_q\}$. 集合 $W \subseteq X \times Y \times Z$, $W = \{(x_k, y_l, z_m) \mid x_k \in X, y_l \in Y, z_m \in Z, \text{其中 } 1 \leq k, l, m \leq q\}$, 定义 $|W| = n$ 代表集合 W 中元素个数 ($1 \leq n \leq q$). 我们从 W 中找出一个匹配 $W^1 \subseteq W$, 匹配 $|W^1| \leq n$, 并且匹配 W^1 中没有任意两个三元组的符号在同一坐标是相同的. 三维匹配问题就是找出集合 W 的最小三维匹配.

给定集合 $X = \{1, 2\}$, $Y = \{3, 4\}$, $Z = \{5, 6\}$, $W = \{(1, 3, 5), (2, 4, 6), (2, 3, 6), (1, 4, 5)\}$. 得到集合 W 三维匹配的解为 $W^1 = \{(1, 3, 5), (2, 4, 6)\}$ 或 $\{(2, 3, 6), (1, 4, 5)\}$.

2.2 DNA 计算新模型

本文提出了一种基于文献[3]中提出的 DNA 计算模型的解决三维匹配问题的新模型. 该模型基于粘贴模型 (Sticker model)^[17] 的解空间, 同时对 Adleman-Lipton 模型^[2~5] 中的生物操作进行了适当的扩展, 兼具 Adleman-Lipton 模型和粘贴模型 (Sticker model) 的优点, 具有随机存储能力, 杂交后误码率较低, 生化试验可行性高等优点.

2.2.1 DNA 计算新模型解空间

本文 DNA 计算新模型基于粘贴模型的解空间. 如式 1, 集合 X, Y, Z 分别用 q 位的二进制串 $x_1^i x_2^i \dots x_q^i, y_1^i y_2^i \dots y_q^i, z_1^i z_2^i \dots z_q^i$ 标示, 其中 x_k^i, y_l^i, z_m^i 标示集合 X, Y, Z 中的元素 x_k, y_l, z_m 是否在可能的三维匹配中 ($i \in \{0, 1\}, 1 \leq k, l, m \leq q$). x_k^1, y_l^1, z_m^1 代表元素 x_k, y_l, z_m 在可能的三维匹配中, x_k^0, y_l^0, z_m^0 代表元素 x_k, y_l, z_m 不在可能的三维匹配中.

对于任一给定的集合: $W = \{W_1, W_2, \dots, W_n\}$, 则可

用一个 n 位二进制串 $w_1^i w_2^i \dots w_n^i$ 表示一种可能的三维匹配方案, w_b 标示 W 中第 b 个元素 ($1 \leq b \leq n$), i 标示元素 w_b 是否在三维匹配中 ($i \in \{0, 1\}$). 令 w_b^1 代表元素 W_b 在可能的三维匹配 W^1 中, w_b^0 代表元素 W_b 不在可能的三维匹配 W^1 中. 匹配 $W^1 = \{w_1, w_2, w_3\}$ 可用二进制串 " $w_1^1 w_2^1 w_3^1$ " 表示. 因此, 三维匹配问题可以转化为求满足条件且含 1 最少的二进制串.

$$\begin{cases} X \rightarrow x_1^i x_2^i \dots x_q^i; Y \rightarrow y_1^i y_2^i \dots y_q^i \\ Z \rightarrow z_1^i z_2^i \dots z_q^i; W \rightarrow w_1^i w_2^i \dots w_n^i \end{cases}, i \in \{0, 1\} \quad (1)$$

2.2.2 DNA 计算新模型生物操作

将三维匹配问题的输入编码成 DNA 碱基序列并生成解空间后, 本文的基本生化操作可描述如下:

■添加 (Append): 分为添加一特定 DNA 链与添加特定长度的 DNA 片段两种添加操作, 即 AppendStrand (T, Z) 和 Append-Length (T, k).

■荧光标记 (Color): 操作 Color (T, C) 中给定试管 T , 对试管中的所有 DNA 链添加颜色为 C 的荧光素.

■抽取 (Extract): 分为按照特定 DNA 片段进行抽取与按照链长进行抽, 即 Extract-Strand (T, S) 和 Extract-Length (T).

■Adleman-Lipton 模型中生物操作: 合并 (Merge), 检测 (Detect), 复制 (Amplify), 读取 (Read) 与丢弃 (Discard).

3 三维匹配问题的 DNA 计算算法

3.1 三维匹配问题的 DNA 计算算法

基于分治策略, 根据 DNA 操作的处理特性^[18, 19], 求解三维匹配问题的算法用 DNA 分子操作实现的基本过程描述如下:

Algorithm 3.1: 三维匹配问题 DNA 计算算法框架

1. 将集合 $W = (w_1, w_2, \dots, w_n)$ 平均分到集合 $W_1 = (w_1, w_2, \dots, w_{n/2})$ 与集合 $W_2 = (w_{n/2+1}, w_{n/2+2}, \dots, w_n)$ 中. 在试管 T_{01}, T_{02} 中用 DNA 链表示 W_1 与 W_2 的所有子集.
2. 利用预搜索器基于三维匹配问题满足的条件对试管 T_{01} 与 T_{02} 中的解空间进行预搜索, 得到三维匹配问题的可满足解空间.
3. 对试管 T_{01} 中每个不同的 DNA 链的前 $n/2$ 个不同的子段分别连接上红颜色的荧光素, 在 T_{02} 中每个不同的 DNA 链的前 $n/2$ 个不同的子段分别连接上绿颜色的荧光素.
4. 三维匹配问题解的搜索过程将相应的 DNA 链转换成等长的链长, 使用荧光标记和凝胶电泳技术, 利用并行搜索器对三维匹配问题进行搜索, 找出三维匹配问题的解.

本算法将分治策略应用于 DNA 计算算法中, 为实现解搜索的并行性, 采用荧光标记方法, 在表示 W_1 所有子集的试管 T_{01} 中每个不同的 DNA 链的前 $n/2$ 个不同的子段分别连接上红色的荧光素, 而在表示 W_2 所有子集的 T_{02} 中每个不同的 DNA 链的前 $n/2$ 个不同的子段分别连接上的绿色的荧光素. 以下介绍算法 3.1 中 DNA 子算法的详细设计.

3.2 三维匹配问题解空间生成器

首先生成三维匹配问题的解空间,设计了下述解空间生成器。

Algorithm 3.2: 三维匹配问题解空间 DNA 计算算法

```

Procedure Init_Solution_Generator( $T_0, n/2$ )
输入:空的试管  $T_0$ ,集合  $W$  中元素个数  $n/2$ 
输出:包含集合  $W$  中  $n/2$  个元素解空间的试管  $T_0$ 
1: For  $i = 1$  down to  $n/2$ 
2:   Amplify( $T_0, T_1, T_2$ )
3:   Append-Strand( $T_1, w_i^1$ )
4:   Append-Strand( $T_2, w_i^0$ )
5:    $T_0 := \cup(T_1, T_2)$ .
6: EndFor
EndProcedure

```

算法 Init_Solution_Generator($T_0, n/2$)需要执行 $n/2$ 次复制操作, n 次添加操作, $n/2$ 次合并操作,假定序列每一位的颜色用长度为 15 的 DNA 链来表示,那么实现算法 Init_Solution_Generator($T_0, n/2$)后 DNA 链长为 $15 \times n/2$,需要试管数为 3。

3.3 三维匹配问题并行搜索器

为了达到并行搜索出解链的目的,设计了三维匹配问题预搜索器与三维匹配问题搜索器。

3.3.1 三维匹配问题预搜索器

根据三维匹配问题满足的条件设计的三维匹配问题预搜索器如下:

Algorithm 3.3.1: 三维匹配预搜索器的 DNA 计算算法

```

Procedure 3-DMatching_Resolvent_Searcher( $T_0, U$ )
输入:包含三维匹配问题解空间的试管  $T_0$ ,用于对解空间进行搜索的集合  $U$ .
输出:三维匹配问题的可满足解.
1: For  $i = 1$  down to  $n/2$ 
2:   Extract-Strand( $T_0, w_i^1$ )
3:    $T_1 := (T_0, w_i^1) +$  and  $T_2 := (T_0, w_i^0) -$ .
4:   If(Detect( $T_1$ ) = 'yes') then
5:     Assume that  $W_i = (x_k, y_l, z_m), 1 \leq k, l, m \leq q$ 
6:     Extract-Strand( $T_1, x_k^1$ )
7:      $T_{OFF} := (T_1, x_k^1) +$  and  $T_{ON} := (T_1, x_k^1) -$ .
8:     Discard( $T_{OFF}$ )
9:     Append-Strand( $T_{ON}, x_k^1$ )
10:    Extract-Strand( $T_{ON}, y_l^1$ )
11:     $T_{OFF} := (T_{ON}, y_l^1) +$  and  $T_{ON} := (T_{ON}, y_l^1) -$ .
12:    Discard( $T_{OFF}$ )
13:    Append-Strand( $T_{ON}, y_l^1$ )
14:    Extract-Strand( $T_{ON}, z_m^1$ ) +
15:     $T_{OFF} := (T_{ON}, z_m^1) +$  and  $T_{ON} := (T_{ON}, z_m^1) -$ .
16:    Discard( $T_{OFF}$ )
17:    Append-Strand( $T_{ON}, z_m^1$ )
18:    EndIf
19:     $T_0 := \cup(T_{ON}, T_2)$ .
20: EndFor
EndProcedure

```

算法 3-DMatching_Resolvent_Searcher(T_0, U)需要执行 $2n$ 次抽取操作, $3n/2$ 次添加操作, $3n/2$ 次丢弃操

作, $n/2$ 次合并操作,算法 3.4 完成后得到的 DNA 链长为 $15 \times (n/2 + 3q)$,需要的试管数为 4。

3.3.2 三维匹配问题搜索器

为了搜索三维匹配问题的解,设计三维匹配问题搜索器如下:

Algorithm 3.3.2: 三维匹配搜索器的 DNA 计算算法

```

Procedure 3-D_Matching_Searcher( $T_{01}, T_{02}, U$ )
输入:包含三维匹配问题解空间的试管  $T_{01}, T_{02}$ ,用于对解空间进行搜索的集合  $U = X \cup Y \cup Z$ .
输出:三维匹配问题的合法解.
1: For  $i = 1$  to  $3 \times q$ 
2:   Extract-Strand( $T_{01}, u_{n/3+i}^1$ )
3:    $T_1 := (T_{01}, u_{n/3+i}^1) +$  and  $T_2 := (T_{01}, u_{n/3+i}^1) -$ .
4:   Extract-Strand( $T_{02}, u_{n/3+i}^1$ )
5:    $T_3 := (T_{02}, u_{n/3+i}^1) +$  and  $T_4 := (T_{02}, u_{n/3+i}^1) -$ .
6:    $k = 2 \times i$ 
7:   IF (Detect( $T_1$ ) = 'yes') then
8:     IF(Detect( $T_4$ ) = 'yes') then
9:       Append-Length( $T_1, 2^{k-1}$ ).
10:      Append-Length( $T_4, 2^{k-1}$ ).
11:     EndIF
12:   EndIF
13:   IF(Detect( $T_2$ ) = 'yes') then
14:     IF(Detect( $T_3$ ) = 'yes') then
15:       Append-Length( $T_2, 2^k$ ).
16:       Append-Length( $T_3, 2^k$ ).
17:     EndIF
18:   EndIF
19:    $T_{01} := \cup(T_1, T_2)$ 
20:    $T_{02} := \cup(T_3, T_4)$ 
21: EndFor
22:  $T_{01} := \cup(T_{01}, T_{02})$ .
EndProcedure

```

算法 3-D_Matching_Searcher(T_{01}, T_{02}, U)需要执行 $6q$ 次抽取操作, $6q$ 次复制操作, $12q$ 次添加操作, $6q + 1$ 次合并操作, DNA 链长度为 $O(45q + 15n/2)$,需要的试管数为 6。

3.4 三维匹配问题的 DNA 计算算法

将实现算法 3.1 中步 1-4 的子算法与前叙各子算法组成一个整体,得到的三维匹配问题 DNA 计算算法。

Algorithm 3.4: 三维匹配问题的 DNA 计算算法

```

Procedure 3-D Matching( $T_{01}, T_{02}, n$ )
输入:空试管  $T_{01}, T_{02}, T_{03}$ .
输出:三维匹配问题的解.
1: Init_Solution_Generator( $T_{01}, n/2$ ) and
   Init_Solution_Generator( $T_{02}, n/2$ ),
2: 3-D Matching_Resolvent_Searcher( $T_{01}, X, Y, Z$ ) and
   3-D Matching_Resolvent_Searcher( $T_{02}, X, Y, Z$ )
3: Color( $T_{01}, \text{Red}$ ) and Color( $T_{02}, \text{Green}$ )
4: 3-Dimensional_Matching_Searcher( $T_{01}, T_{02}, U$ ).
5: If(Detect( $T_{01}$ ) = 'yes') then
6:   Extract-Length( $T_{01}$ ).
7:   Read( $T_{01}$ ).
8: EndIf
EndProcedure

```

定理 3.4 算法 3-D Matching(T_{01}, T_{02}, n)可并行求

解三维匹配问题.

证明 算法 3.4 第 1 步执行 $\text{Init_Solution_Generator}(T_{01}, n/2)$ 与 $\text{Init_Solution_Generator}(T_{02}, n/2)$ 将集合 W_1 和 W_2 对应的子集用 DNA 链表示, 并通过两种不同的荧光素表示来自不同集合的子集. 第 2 步 3-D matching_Resolvent_Searcher(T_{01}, X, Y, Z), 3-D Matching_Resolvent_Searcher(T_{02}, X, Y, Z), 执行预搜索操作, 分别搜索出试管 T_{01}, T_{02} 中合法的 DNA 链. 第 3 步执行 $\text{Color}(T_{01}, \text{red})$ 与 $\text{Color}(T_{02}, \text{green})$ 操作对试管 T_{01} 中每个不同的 DNA 链的前 $n/2$ 个不同的子段分别连接上红色的荧光素, 在 T_{02} 中的 DNA 链的前 $n/2$ 个不同的子段分别连接上的绿色的荧光素. 第 4 步 3-Dimensional_Matching_Searcher(T_{01}, T_{02}, U). 遍历集合 U 中所有的元素, 对于试管 T_{01}, T_{02} 中相应的 DNA 链末尾添加一段等长的 DNA 片段. 最后把所有可满足的 DNA 链合并到试管 T_{01} 中. 第 5-8 步完成最后解链的搜索过程. 第 5 步对试管 T_{01} , 使用检测操作, 若返回 'yes' 执行第 6-7 步. 第 6 步对试管执行抽取操作, 抽取具有相同链长的 DNA 链. 第 7 步使用读取操作, 读取链长相同并且具有两种颜色的 DNA 链信息, 最终得到三维匹配问题的解.

3.5 性能分析与比较

定理 3.5 算法 3.4 求解 $X = \{x_1, \dots, x_q\}, Y = \{y_1, \dots, y_q\}, Z = \{z_1, \dots, z_q\}$, 集合 $W \subseteq X \times Y \times Z, W = \{(x_k, y_l, z_m) \mid x_k \in X, y_l \in Y, z_m \in Z, \text{其中 } 1 \leq k, l, m \leq q\}, |W| = n$ 的三维匹配问题所需要的生物操作数为 $O(15n + 30q)$, 测试试管数为 $O(1)$, 最大链长为 $O(45q + 15n/2)$, DNA 链数为 $O(2^{n/2}) \approx O(1.414^n)$.

证明 算法 3.4 由 8 步组成, 与文献[2~5]的相应算法类似. 在第 1 步需要执行 n 次复制操作, $2n$ 次添加操作, n 次合并操作. 第 2 步需要执行需要执行 $4n$ 次抽取操作, $3n$ 次添加操作, $3n$ 次丢弃操作, n 次合并操作. 第 3 步执行 2 次着色操作. 第 4 步需要需要执行 $6q$ 次抽取操作, $6q$ 次复制操作, $12q$ 次添加操作, $6q + 1$ 次合并操作. 第 5 步需要需要执行 1 次检测操作, 第 6 步需要需要执行 1 次按 DNA 链长度抽取操作, 第 8 步需要需要执行 1 次读取操作. 因此, 如式 2, 该算法总的生物操作数为 $15n + 30q + 6$, 因此算法总的生物操作数为 $O(15n + 30q)$.

$$\begin{aligned} \text{Operation_Num} &= (n + 2 \times n + n) + (4 \times n + 3 \times n + 3 \times n + n) \\ &\quad + 2 + (6 \times q + 6 \times q + 12 \times q + 6 \times q + 1) + 1 + 1 \\ &= 15 \times n + 30 \times q + 6 \end{aligned} \quad (2)$$

由文献[2~5]中的算法分析可知, 算法 3.2 需要试管数为 3, 算法 3.3.1 需要试管数为 4, 算法 3.3.2 需要试管数为 6, 因此, 本该算法需要的试管数为 $3 \times 3 + 4 \times 3 + 6 = 27$ 即 $O(1)$ 由算法 3.3.1 与引理 3.3.2 的结论,

算法中所有 DNA 链的最大链长为 $O(45q + 15n/2)$.

本算法借鉴了分治策略将包含 n 个元素的集合平均分到两个各含 $n/2$ 个元素的集合 W_1 与 W_2 中, 在子算法 $\text{Init_Solution_Generator}(T_{01}, n/2)$, $\text{Init_Solution_Generator}(T_{02}, n/2)$, 中各生成 $2^{n/2}$ 个 DNA 链. 此后算法执行过程中没有再生成新的 DNA 链. 因此, 总的 DNA 链数为 $O(2^{n/2})$ 即 $O(2^{n/2}) \approx O(1.414^n)$.

文献[16]中的算法依然使用穷举求解问题, 算法的空间复杂度为 $O(2^n)$, 需要的试管数为 $O(n)$, 最大链长为 $O(15n + 45q)$, 生物操作数为 $O(n^2)$. 以现有生化实验水平, 每个试管中生化药物的浓度能达到每根试管中包含 10^{22} 个 DNA 链^[12], 因此, 上述算法理论上均仅能求解 $n \leq 67$ 的三维匹配问题 ($10^{22} \approx 2^{67}$).

上述算法的性能和本文算法性能的比较结果见表 1. 从表 1 可见, 本文算法在生物操作数由 $O(n^2)$ 减少至 $O(15n + 30q)$, 测试试管数由所需的 $O(n)$ 减少至 $O(1)$, 最大链长由 $O(15n + 45q)$ 减少至 $O(15n/2 + 45q)$, DNA 链数由 (2^n) 减小至 $O(1.414^n)$. 因此, 该算法不仅理论上可求解三维匹配问题的规模将可达到 $134 (2^{134/2} \approx 2^{67} \approx 10^{22})$, 而且在空间复杂性和实验可行性等综合性能上的比较优势也是明显的^[2~5,8].

表 1 三维匹配问题 DNA 计算算法的性能对比

算法	DNA 链数	DNA 链长	操作数	试管数
文献[14]	$O(2^n)$	$O(15n + 45q)$	$O(n^2)$	$O(n)$
本文算法	$O(1.414^n)$	$O(15n/2 + 45q)$	$O(15n + 30q)$	$O(1)$

4 模拟实验结果

为验证算法的有效性, 本文采用文献[2~5,8]中的实验方法. 以 $X = \{1, 2\}, Y = \{3, 4\}, Z = \{5, 6\}, W = \{(1, 3, 5), (2, 4, 6), (2, 3, 6), (1, 4, 5)\}$ 为三维匹配问题的实例, 给出使用该算法对此实例的模拟求解过程.

4.1 DNA 编码^[2~5,9]

表 2 标示求解图 1 中三维匹配问题所需的 DNA 序列. 其中 $w_i (1 \leq i \leq 2)$ 为子集的标号, 用于标示集合 W_i 中的第 i 各元素是否在可能的三维匹配中; x_j, y_j 与 $z_j (1 \leq j \leq 2)$ 的值用于标示集合 X, Y, Z 中的第 j 个元素是否在可能的三维匹配中.

4.2 算法求解过程

当编码完全产生后, 即可对算法进行实现. 首先, 把集合 $W = \{(1, 3, 5), (2, 4, 6), (2, 3, 6), (1, 4, 5)\}$ 平均分到两个集合 W_1 与 W_2 中. 得到 $W_1 = \{(1, 3, 5), (2, 4, 6)\}, W_2 = \{(2, 3, 6), (1, 4, 5)\}$

表 2 图 1 中三维匹配问题所需的 DNA 序列

位	5'→3' DNA 序列	焓(H)	熵(S)	自由能(G)
w_0^0	AATTCACAAACAATT	114.4	299.4	25.0
w_1^1	CCTTATCATCCAAATC	112.8	284.9	27.8
w_2^0	AATTCGCCATTCCTTA	108.5	273.0	27.0
w_2^1	TCTCTCTCTAATCAT	105.2	270.5	24.4
w_3^0	CTTCTCCACTATACT	111.1	288.3	25.0
w_3^1	CCTTTCTAACCTTCA	103.8	272.6	22.3
w_4^0	AAACTCTACATACAC	109.9	285.5	24.5
w_4^1	AATTAACAATCATCT	104.3	273.0	22.7
x_1^0	TTACTCTTAACATCT	112.1	282.8	27.8
x_1^1	TTAATCAAATCCCTA	102.1	266.0	22.7
x_2^0	ATTCTAACTCTACCT	105.2	277.1	22.4
x_2^1	TCTAATATAATTACT	104.8	283.7	19.9
y_1^0	AAAACCTCACCTCCT	113.7	288.7	27.5
y_1^1	TAATTCCATAACCTA	110.2	288.2	24.0
y_2^0	CCAATTCCAATAATC	113.0	290.6	26.2
y_2^1	ACCATCTCCAATTC	111.1	280.0	27.6
z_1^0	TTTCCAACCTCTTCA	108.2	276.9	25.6
z_1^1	CTACTACAATCCAC	101.1	253.6	25.3
z_2^0	TATCTACTAAACCAA	114.4	299.4	25.0
z_2^1	TAATACCTAATTACC	112.8	284.9	27.8

算法的执行过程如下:

(1) 执行算法 3.4 第(1)步得到包含三维匹配问题的解空间的试管 T_{01} 与 T_{02} . 其中:

试管 T_{01} 中 DNA 链为: $\{w_1^0 w_2^0, w_1^0 w_2^1, w_1^1 w_2^0, w_1^1 w_2^1\}$;

试管 T_{02} 中 DNA 链为: $\{w_1^0 w_2^0, w_1^0 w_2^1, w_1^1 w_2^0, w_1^1 w_2^1\}$.

(2) 解空间生成后执行算法 3.4 第 2 步后得到:

试管 T_{01} 中的 DNA 链为:

$$\{w_1^0 w_2^0 x_1^0 x_2^0 y_1^0 y_2^0 z_1^0 z_2^0, w_1^0 w_2^1 x_1^0 x_2^1 y_1^0 y_2^1 z_1^0 z_2^1, w_1^1 w_2^0 x_1^1 x_2^0 y_1^0 y_2^0 z_1^0 z_2^0, w_1^1 w_2^1 x_1^1 x_2^1 y_1^1 y_2^1 z_1^1 z_2^1\}.$$

试管 T_{02} 中的 DNA 链为:

$$\{w_1^0 w_2^0 x_1^0 x_2^0 y_1^0 y_2^0 z_1^0 z_2^0, w_1^0 w_2^1 x_1^0 x_2^1 y_1^0 y_2^1 z_1^0 z_2^1, w_1^1 w_2^0 x_1^1 x_2^0 y_1^0 y_2^0 z_1^0 z_2^0, w_1^1 w_2^1 x_1^1 x_2^1 y_1^1 y_2^1 z_1^1 z_2^1\}.$$

(3) 执行算法 3.4 第 3 步操作后

试管 T_{01} 中的 DNA 链为:

$$\{Red-w_1^0 w_2^0 x_1^0 x_2^0 y_1^0 y_2^0 z_1^0 z_2^0, red-w_1^0 w_2^1 x_1^0 x_2^1 y_1^0 y_2^1 z_1^0 z_2^1, Red-w_1^1 w_2^0 x_1^1 x_2^0 y_1^0 y_2^0 z_1^0 z_2^0, Red-w_1^1 w_2^1 x_1^1 x_2^1 y_1^1 y_2^1 z_1^1 z_2^1\}.$$

试管 T_{02} 中的 DNA 链为:

$$\{Green-w_1^0 w_2^0 x_1^0 x_2^0 y_1^0 y_2^0 z_1^0 z_2^0, Green-w_1^0 w_2^1 x_1^0 x_2^1 y_1^0 y_2^1 z_1^0 z_2^1, Green-w_1^1 w_2^0 x_1^1 x_2^0 y_1^0 y_2^0 z_1^0 z_2^0, Green-w_1^1 w_2^1 x_1^1 x_2^1 y_1^1 y_2^1 z_1^1 z_2^1\}.$$

(4) 执行算法 3.4 第 4 步得到:

试管 T_{01} 中的 DNA 链为:

$$\{Red-w_1^0 w_2^0 x_1^0 x_2^0 y_1^0 y_2^0 z_1^0 z_2^0 + (2^1 + 2^3 + 2^5 + 2^7 + 2^9 + 2^{11}) \text{ bpDNA 片段}, Red-w_1^0 w_2^1 x_1^0 x_2^1 y_1^0 y_2^1 z_1^0 z_2^1 + (2^0 + 2^3 + 2^5 + 2^6 + 2^8 + 2^{11}) \text{ bpDNA 片段}, Red-w_1^1 w_2^0 x_1^1 x_2^0 y_1^0 y_2^0 z_1^0 z_2^0 + (2^1 + 2^2 + 2^4 + 2^7 + 2^9 + 2^{10}) \text{ bpDNA 片段}, Red-w_1^1 w_2^1 x_1^1 x_2^1 y_1^1 y_2^1 z_1^1 z_2^1 + (2^0 + 2^2 + 2^4 + 2^6 + 2^8 + 2^{10}) \text{ bpDNA 片段}\}.$$

$+ (2^0 + 2^2 + 2^4 + 2^6 + 2^8 + 2^{10})$ bpDNA 片段, $Green-w_1^0 w_2^0 x_1^0 x_2^0 y_1^0 y_2^0 z_1^0 z_2^0 + (2^0 + 2^2 + 2^4 + 2^6 + 2^8 + 2^{10})$ bpDNA 片段, $Green-w_1^0 w_2^1 x_1^0 x_2^1 y_1^0 y_2^1 z_1^0 z_2^1 + (2^0 + 2^3 + 2^4 + 2^7 + 2^8 + 2^{11})$ bpDNA 片段, $Green-w_1^1 w_2^0 x_1^1 x_2^0 y_1^0 y_2^0 z_1^0 z_2^0 + (2^1 + 2^2 + 2^5 + 2^6 + 2^9 + 2^{10})$ bpDNA 片段, $Green-w_1^1 w_2^1 x_1^1 x_2^1 y_1^1 y_2^1 z_1^1 z_2^1 + (2^1 + 2^3 + 2^5 + 2^7 + 2^9 + 2^{11})$ bpDNA 片段}.

(5) 执行检测操作, 试管 T_{01} 非空. 按照试管中的 DNA 链长进行分离. 得到链长相等且具有两种颜色的 DNA 链为:

$$\{Red-w_1^0 w_2^0 x_1^0 x_2^0 y_1^0 y_2^0 z_1^0 z_2^0 + (2^1 + 2^3 + 2^5 + 2^7 + 2^9 + 2^{11}) \text{ bpDNA 片段}, Green-w_1^1 w_2^1 x_1^1 x_2^1 y_1^1 y_2^1 z_1^1 z_2^1 + (2^1 + 2^3 + 2^5 + 2^7 + 2^9 + 2^{11}) \text{ bpDNA 片段}\};$$

$$\{Red-w_1^1 w_2^1 x_1^1 x_2^1 y_1^1 y_2^1 z_1^1 z_2^1 + (2^0 + 2^2 + 2^4 + 2^6 + 2^8 + 2^{10}) \text{ bpDNA 片段}, Green-w_1^0 w_2^0 x_1^0 x_2^0 y_1^0 y_2^0 z_1^0 z_2^0 + (2^0 + 2^2 + 2^4 + 2^6 + 2^8 + 2^{10}) \text{ bpDNA 片段}\}.$$

由此得到三维匹配问题的解即为 $\{(1, 3, 5), (2, 4, 6)\}$ 或 $\{(2, 3, 6), (1, 4, 5)\}$.

5 结论

在文献[2~5, 19~22]工作的基础上, 本文对三维匹配问题 DNA 计算模型的可扩展性问题进行了较深入的探索; 根据现有 DNA 计算模型的生物操作特性^[2~5, 9]和 NP 完全的三维匹配问题的并行求解需求, 基于文献[2~5]的 DNA 计算模型, 借鉴传统计算中的分治策略设计了一种求解三维匹配问题的 DNA 计算新算法. 本文算法丰富了 DNA 计算的研究内容, 为解决 DNA 计算指数爆炸问题提供一种新的思路. 下一步将在文献[18~22]的基础上对图着色问题及最大团问题做进一步的研究, 以提出针对以上两个问题的亚指数 DNA 计算算法.

参考文献:

- [1] Adleman L. Molecular computation of solutions to combinatorial problems[J]. Science, 1994, 266(5187): 1021 - 1024.
- [2] Michael Ho, Chang W L, Guo M. Fast parallel solution for set-packing and clique problems by DNA-based computing[J]. IEICE Transactions on Information and System, 2004, E87-D(7): 1782 - 1788.
- [3] Chang W L, Guo M, Michael H. Fast parallel molecular algorithms for DNA-based computation[J]. IEEE Transactions on Nanobioscience, 2005, 4(2): 133 - 163.
- [4] Chang W L, et al. Molecular solutions for the subset-sum problem on DNA-based supercomputing[J]. BioSystems, 2004, 73: 117 - 130.
- [5] Chang W L, et al. Fast parallel molecular solutions for DNA-based supercomputing: Factoring Integers[J]. IEEE Transactions on Nanobioscience, 2005, 4(2): 149 - 163.

- [6] 许进, 强小利, 方刚, 等. 一种图顶点着色 DNA 计算机模型[J]. 科学通报, 2006, 51(4): 480 - 487.
Xu Jin, Qiang Xiao Li, Fang Gang, et al. A DNA computer model for solving vertex coloring problem[J]. Chinese Science Bulletin, 2006, 51(20): 2541 - 2549. (in Chinese)
- [7] 李源, 方辰, 欧阳颀. 最大团问题的 DNA 计算机进化算法[J]. 科学通报, 2004, 49(5): 439 - 443.
Li Yuan, Fang Chen, Ou Yang Qi. Genetic algorithm in DNA Computing: A solution to the maximal clique problem[J]. Chinese Science Bulletin, 2004, 49(9): 967 - 971. (in Chinese)
- [8] Wang X L, Bao Z M, Hu J J, et al. Solving the SAT problem using a DNA computing algorithm based on ligase chain reaction[J]. BioSystems, 2008, 91: 17 - 125.
- [9] 李肯立, 周旭, 许进. 图 3-着色问题的 $O(2^n)$ 链数 DNA 计算机算法[J]. 电子学报, 2008, 36(11): 2096 - 2101.
Li Ken Li, Zhou Xu, Xu Jin. An $O(2^n)$ Volume Molecular Solutions for the 3-Colorable Problem on DNA-Based Supercomputing[J]. Acta Electronica Sinica, 2008, 36(11): 2096 - 2101. (in Chinese)
- [10] 刘文斌, 朱翔鸥, 王向红, 等. DNA 计算的研究进展[J]. 电子学报, 2006, 34(11): 2053 - 2057.
Liu Wen Bin, Zhu Xiang Ou, Wang Xiang Hong, et al. Progress in the research of DNA computing[J]. Acta Electronica Sinica, 2006, 34(11): 2053 - 2057. (in Chinese)
- [11] Shin S Y, Lee I H, Kim D M, et al. Multiobjective evolutionary optimization of DNA sequences for reliable DNA computing[J]. IEEE Transactions on Evolutionary Computation, 2005, 9(2): 143 - 158.
- [12] Qiu Z F, Lu M. A new approach to advance the DNA computing[J]. Applied Soft Computing, 2003, (3): 177 - 189.
- [13] Lipton R J. DNA solution of hard computational problems[J]. Science, 1995, 268(28): 542 - 545.
- [14] 卢圣栋. 现代分子生物学实验技术[M]. 北京: 中国协和医科大学出版社, 1999. 23 - 59.
- [15] Garey M R, Johnson D S. Computer and Intractability: A guide to the theory of NP-completeness[M]. San Francisco: W. H. Freeman and Company, 1979.
- [16] Michael Ho, Chang W L, Cao J N. Using sticker to solve the 3-dimensional matching problem in molecular super-computers[J]. High Performance Computing and Networking, 2004, 1: 128 - 139.
- [17] 许进, 李三平, 董亚非. 粘贴 DNA 计算机模型(I): 理论[J]. 科学通报, 2004, 49(3): 205 - 212.
Xu Jin, Li Shan Ping, Dong Ya Fei. Sticker DNA computer model-Part I: Theory[J]. Chinese Science Bulletin, 2004, 49(3): 205 - 212. (in Chinese)
- [18] 李汪根, 丁永生. DNA 计算机中队列数据结构的设计及实现[J]. 计算机学报, 2007, 6(30): 993 - 998.
Li Wang Gen, Ding Yong Sheng. Design and implementation of queue data structure in DNA computer[J]. Chinese Journal of Computers, 2007, 6(30): 993 - 998. (in Chinese)
- [19] 黄育潜. DNA 计算的基本代数原理(一)[J]. 计算机学报, 2008, 3(31): 353 - 371.
Huang Yu Qian. The basic algorithm principles of DNA computing (I)[J]. Chinese Journal of Computers, 2008, 3(31): 353 - 371. (in Chinese)
- [20] Li Kenli, Zhou Shouting, Xujin. Fast parallel molecular algorithms for DNA-Based computation: solving the elliptic curve discrete logarithm problem over $GF(2^n)$ [J]. Journal of Biomedicine and Biotechnology, 2008, 1: 1 - 10.
- [21] Li D F, Li X R, et al. Scalability of the surface-based DNA algorithm for 3-SAT[J]. BioSystems, 2006, 85: 95 - 98.
- [22] 李肯立, 周旭, 邹舒婷. 一种改进的最大团问题的 DNA 计算机算法[J]. 计算机学报, 2008, 12: 2173 - 2181.
Li Ken Li, Zhou Xu, Zou Shu Ting. Improved volume molecular solutions for the maximum clique problem on DNA-based supercomputing[J]. Chinese Journal of Computers, 2008, 12: 2173 - 2181. (in Chinese)

作者简介:



周旭女, 1983 年生于江苏宿迁, 助教, 研究方向为并行计算、DNA 计算。
E-mail: happypanda2006@126.com



李肯立男, 1971 年生于湖南娄底, 教授, 中国计算机学会高级会员, 研究方向为并行处理、生物计算、网格计算等。
E-mail: lkl510@263.net

乐光学男, 1963 年生于贵州, 嘉兴学院教授, 博士生, 主要研究方向为计算机网络与网络安全、DNA 计算等。

杨志邦男, 1984 年生于湖南, 博士生, 主要研究方向为嵌入式并行计算、图像识别、DNA 计算等。