

基于三链 DNA 结构的 0-1 整数规划改进研究*

任晓玲, 白雪, 刘希玉

(山东师范大学 管理科学与工程学院, 济南 250014)

摘要: 为实现 DNA 计算中对解的有效筛选, 防止探针与探针之间的错配、发夹结构等, 以及便于检测最终解, 提出了改进的三链 DNA 模型求解 0-1 规划的设计。该方法编码 n 个变量的每种组合的所有排列情况。此编码方式不仅使计算所需有效分子量从 $O((2n)!)$ 下降到 $O(2^n n!)$, 并使对可行解的筛选更加有效。利用寡聚脱氧核苷酸(ODN)在 RecA 蛋白介导下与同源的双链 DNA 匹配成三螺旋 DNA 的特点, 可推广到更多以双链 DNA 分子为计算模型的解的检测中。

关键词: 三链 DNA; 0-1 整数规划; DNA 计算

中图分类号: TP301.6 文献标志码: A 文章编号: 1001-3695(2013)01-0056-04

doi: 10.3969/j.issn.1001-3695.2013.01.013

Improvement of 0-1 integer programming based on triple-stranded DNA structure

REN Xiao-ling, BAI Xue, LIU Xi-yu

(School of Management Science & Engineering, Shandong Normal University, Jinan 250014, China)

Abstract: To realize the effective screening of solutions in DNA computing and to prevent the mismatch between probes, the formation of the hairpin structure etc, this paper presented an improved triple-stranded DNA model to solve 0-1 integer programming problem. This method made a full array of all combinations of n 0-1 variables. Compared with the original one, the amount of DNA strands needed dropped from $O((2n)!)$ to $O(2^n n!)$ and the improved one made a better selection of feasible solutions. Triple-helix structure could be constructed by the homologous double-helix DNA with the oligodeoxyribonucleotides (ODN) in the mediated of RecA protein. It could make use of its special structure to promote it to the selection of solutions with the double-helix computing model.

Key words: triple-stranded DNA; 0-1 integer programming; DNA computing

0 引言

1957年, Feisenfeld 等人提出了三链核酸的概念。由于双链 DNA 中两条相互缠绕且方向相反的链是依靠互补碱基对之间的氢键相连, 碱基对又偏向螺旋轴一侧, 所以在两条链之间形成了大小两沟。三链 DNA 正是在双链 DNA 的基础上, 通过含有多聚嘌呤的那条链, 以 Hoogsteen 和反式 Hoogsteen 型氢键与新加入的第三条链相连接形成的, 且第三条链位于双链 DNA 的大沟中, 主要有嘧啶-嘌呤-嘧啶和嘌呤-嘌呤-嘧啶两种类型。

2004年, Shigemori 等人发现寡聚脱氧核苷酸能在 RecA 蛋白及 ATP _{γ} S 的存在下与双链 DNA 形成稳定的三链结构。由于三链 DNA 在形成过程中不会发生错配和形成发夹结构, 而且反应后得到的解为双链形式, 比较稳定, 不会像单链 DNA 因为过长而形成发夹结构或断裂。因此, 基于三链结构 DNA 计算的错误率和编码复杂度会降低。

目前, 关于三链 DNA 的研究侧重于生物、医学领域, 利用

三链 DNA 结构实现 DNA 计算的实例不多, 主要解决了 0-1 整数规划问题^[1, 2]、可满足性问题^[3, 4]、顶点着色问题^[5]、邮递员问题以及工序安排问题^[6]。2008年, 杨静等人利用同源的存在抗蛋白质的脱氧核苷酸(ODN)在双螺旋 DNA 链中形成的三链 DNA 结构解决了简单的 0-1 整数规划问题, 通过对三链 DNA 的构造和筛选完成了可行解的获取。文献[3, 4]在阐述三链 DNA 的形成过程以及可行性实验研究的基础上, 提出将三链 DNA 计算模型应用于求解可满足性问题。将 n 个变量的不同取值对应为 $2n$ 种不同的 RecA 蛋白-ODN 复合体探针, 与初始解空间在适当的反应液中哺育, 通过探针是否与库链形成稳定的三链 DNA 结构判断并筛选满足符合每个子句的解序列。在求解图顶点着色问题时, 利用探针与库链中某些 DNA 双链形成的三链 DNA 结构过滤掉相邻节点着同一颜色的非可行解, 进而筛选出相邻节点着不同颜色的可行解。文献[6]利用类似的方案求解了供需安排问题。

1 0-1 规划问题

作为整数规划的特殊形式, 0-1 规划是运筹学中的一个重

收稿日期: 2012-05-22; 修回日期: 2012-06-25 基金项目: 国家自然科学基金资助项目(61170038); 山东省自然科学基金资助项目(ZR2011FM001); 国家教育部人文社会科学研究项目(12YJA630152); 山东省社会科学基金资助项目(11CGLJ22); 山东省高等学校科技计划资助项目(J12LN22, J12LN65)

作者简介: 任晓玲(1988-), 女, 山东济南人, 硕士研究生, 主要研究方向为 DNA 计算、聚类研究(851171287@qq.com); 白雪(1989-), 女, 山东济南人, 硕士研究生, 主要研究方向为 DNA 计算、信息管理; 刘希玉(1964-), 男, 泰山学者, 教授, 博士, 主要研究方向为信息管理与电子商务、计算智能、非线性分析、计算机辅助创新设计等。

要问题。2003 年,殷志祥等人^[7,8]设计出两种基于表面的 DNA 计算模型求解一般形式的 0-1 规划问题: a) 通过固定在表面上且被荧光分子标记的 ODN 在被其互补链杂交时的荧光猝灭现象,排除不满足约束条件的序列; b) 通过 ODN 与标有不同荧光的探针杂交后产生的荧光数目判断满足约束条件的可行解。2005 年, Zhou 等人^[9]以分子信标技术为基础,通过电泳实验直接找到 0-1 规划问题的最优解,在一定程度上减少了实验步骤,提高了实验效率。同年, Wang 等人^[10]通过将一般整数规划转换为特殊形式的 0-1 规划,利用荧光标记技术对问题进行求解。在近几年的研究中, Zhang 等人^[11]和罗海波^[12]将基于溶液和基于表面的 DNA 计算模型结合起来,提出基于溶液与表面结合的 DNA 计算模型,充分利用了在溶液中计算的巨大并行性和高存储性,以及在表面上计算方便解的读取、易于自动化的优点。许进等人^[13]将接入和删除实验结合起来,对闭环 DNA 分子存储的数据进行置换,解决了更广泛范围 0-1 规划问题。

在杨静利用三链 DNA 求解 0-1 整数规划的问题中,对于 $2n$ 个变量生成 $(2n!)$ 种长度为 $(30 \times 2n)$ bp 的 DNA 单链,再根据碱基互补配对原则生成 $(2n!)$ 种双链 DNA,构成初始数据池。随后利用表示每个变量的 DNA 单链生成核蛋白细状体,通过把满足第一个语句的所有组合的核蛋白细状体制成探针与初始数据池中的双链 DNA 进行反应,生成满足第一个约束条件的三链 DNA 可能组合。通过对反应产物的纯化和提取,获得满足第一个约束条件的双链 DNA。对于具有 m 个约束条件的 0-1 整数规划问题,对剩下的各约束条件进行相同的反应,最后得到的双链 DNA 即为满足 m 个约束条件的解。

在可行解的读取方面,仅提到通过 PCR 扩增技术读出可行解,笔者认为此处存在一定缺陷,因为长为 $(30 \times 2n)$ bp 的 DNA 双链中包括各变量所有可能的取值,无法直接判断并读出符合条件的取值;另外,对于所有可行解,使用比较目标函数值获得最优解的方法可以进一步优化。针对以上两个问题,本文提出了一种新的初始数据池构造方法,通过构造 $(2^n \times n!)$ 种长度为 $(20 \times n)$ bp 的 DNA 双链形成初始数据池,实现可行解的读取,并利用 DNA 纳米金探针与产物发生杂交反应时发出的不同荧光来选择最优解。具体方法将在下文给出。

2 0-1 整数规划的形式转换

本文主要利用三链 DNA 模型对杨静提出的解决 0-1 整数规划问题的 DNA 算法进行改进,故首先给出 0-1 整数规划的一般形式,在此假设目标函数的价值系数为整数:

$$\begin{aligned} \max(\min) z &= c_1x_1 + c_2x_2 + \dots + c_nx_n \\ \text{s. t.} \quad &\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \geq (\leq, =) b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \geq (\leq, =) b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \geq (\leq, =) b_m \\ x_i = 0, 1 \quad 1 \leq i \leq n \\ a_{ij}, b_i \text{ 为整数} \quad i=1, \dots, m; j=1, \dots, n \end{cases} \end{aligned} \quad (1)$$

为了实现 DNA 计算 0-1 整数规划问题,要将式 (1) 的形式进行转换,将其转换为系数为 0、1 的整数规划的形式。具体操作步骤如下:

a) $i = 1 \rightarrow m$, 检查不等式 i 的右端项是否非负。若 $b_i < 0$, 则两端同乘 -1 , 得到变化后的整数规划形式为

$$\max(\min) z = c_1x_1 + c_2x_2 + \dots + c_nx_n$$

$$\text{s. t.} \quad \begin{cases} \sum_{j=1}^n a_{ij}^{(1)} x_j \geq (\leq, =) b_i^{(1)} \quad i=1, \dots, m \\ x_j = 0, 1 \quad 1 \leq j \leq n \\ a_{ij}^{(1)}, b_i^{(1)} \text{ 为整数且 } b_i^{(1)} \geq 0, i=1, \dots, m; j=1, \dots, n \end{cases} \quad (2)$$

b) $j = 1 \rightarrow n$, 令

$t_j = \max(|a_{1j}^{(1)}|, |a_{2j}^{(1)}|, \dots, |a_{mj}^{(1)}|, |c_j|)$, 将 x_j 扩展成 t_j 个与之相等的变量: $x_{j1} = x_{j2} = \dots = x_{jt_j}$, 并将其作为新添加的约束条件。

c) $i = 1 \rightarrow m, j = 1 \rightarrow n$, 令

$a_{ij}^{(1)} x_j = x_{j1} + x_{j2} + \dots + x_{jt_j} | a_{ij}^{(1)} |$ 。经过前三步,将整数规划问题约束条件的系数变成了只包含 0、-1、1 三种形式的系数。

d) $i = 1 \rightarrow m, j = 1 \rightarrow n$, i th 不等式中系数为 -1 的变量 x_j (或 x_{jq} ($q = 1, 2, \dots, t_j$)) 用 $1 - \bar{x}_j$ 代替,即 $-x_j = \bar{x}_j - 1$, 同时将常数项 -1 移至不等式右端,不等式右端仍然保持非负。

e) $j = 1 \rightarrow n$, 改变 $x_{j1}, x_{j2}, \dots, x_{jt_j}$ 的下标,分别取成不同的下标,步骤 b) 中的约束条件也要进行相应改变。于是,最终得到的整数规划形式为目标函数中的价值系数和约束条件的技术系数为 0 或 1 的整数规划,其中包含 $2n$ 个取值为 0、1 的整数变量 $(x_i, \bar{x}_i, i = 1, \dots, n)$ 。

通过以上步骤,就得到了自变量系数仅为 0、1 的最简单的 0-1 整数规划,而本文设计的 DNA 算法正是基于这种最简单的结构。另外,文献 [10] 给出将一般的整数规划问题转换为 0-1 线性规划问题。于是通过转换,本文算法实际上可以解决一般的整数规划问题。

3 算法改进及相关证明

3.1 算法改进

在将一般的整数规划问题转换为简单的 0-1 线性规划问题的基础上,基于已成熟的生物操作实验,给出基于三链 DNA 结构求解该问题的基本算法、生物算法及其操作步骤如下:

a) 基本算法

- (a) 生成给定问题的变量的全部 0、1 组合的各种排列;
- (b) 保留满足第一个约束条件的解;
- (c) 重复 (b), 直到保留满足最后一个约束条件的解;
- (d) 根据求最大或最小值筛选可行解, 读出最优解。

b) 生物算法

假设通过转换之后得到的 0-1 规划包含 $2n$ 个变量、 m 个约束方程。

(a) 构造表示所有 $2n$ 个变量的单链 DNA, 并形成其相应补链。反应后得到 n 个变量的每种组合的全部排列情况(注意此时每个位置 i 上只能出现 x_k, \bar{x}_k 两个变量的其中之一)。

(b) 将表示满足第一个约束条件的变量组合对应的补链的 DNA 单链与抗原蛋白质混合在一定条件下生成核蛋白细状体。

(c) 利用满足第一个约束条件的蛋白细状体作为探针,生成满足第一个约束条件的三链 DNA。

(d) 利用磁珠吸附链(霉)亲和素的高度亲和力,将所有满足第一个约束条件的三链 DNA 分离出来,再形成双链 DNA, 此时的数据池都是满足第一个约束条件的双链分子。

(e) 重复 (b) ~ (d), 排除所有不满足条件的解,最后得到满足所有条件的双链 DNA。

(f) 利用目标函数中的变量作为探针筛选最优解,使用

PCR 扩增, 读出最优解。

c) 生物操作步骤

(a) 对给定的变量 $x_i, \bar{x}_i (i = 1, 2, \dots, n)$, 合成 $2n$ 种短的 ODN, 同时根据碱基互补配对原则确定变量的补链。以 Braich 等人^[14]提供的方法在 DNA 自动合成仪上合成全部的 $(2^n n!)$ 解空间序列, 或者利用文献[15]在解决分箱问题时产生初始数据池的方法, 产生所需的 $(2^n n!)$ 条解空间序列, 然后经 PCR 扩增成双链 DNA, 将这些双链称为初始数据池, 同时, 将目标函数中变量的补链采取不同大小的纳米粒子设计成探针。

(b) 在表面上制作满足第一个约束条件的所有变量组合的补链, 形成所有满足第一个约束条件的变量组合的单链探针 (可任意设置其中各变量在单链中的位置)。在这些单链探针的 5' 端添加一个聚酯纤维 A, 再标记上生物素, 然后将这些探针与 RecA 蛋白、ATP_γS 混合, 在适宜的条件下哺育, 使 RecA 蛋白包裹探针形成核蛋白细状体。

(c) 将处理好的探针与数据池混合, 在适当反应液中混合哺育, 这些探针在 RecA 蛋白介导下与库链的某些区域就形成了稳定的三链 DNA 结构。

(d) 将包裹有链(霉)亲和素的磁珠与上一步得到的产物在适合的条件下混合。由于生物素与链(霉)亲和素的高度亲和性, 形成的三链 DNA 结构与探针一起被吸附在磁珠上, 并与其他库链分开。经过洗脱及脱蛋白处理, 得到符合第一个约束条件的解序列。

(e) 对剩下的约束条件中变量取值重复 (b) ~ (d) 的操作, 每次重复操作都要与 (d) 得到的解序列混合反应。在 m 个约束条件都进行了筛选之后, 得到了满足约束条件的解序列。

(f) 提取最终的解序列, 使用相应的引物进行 PCR 扩增, 将扩增后的产物与目标函数中变量的 DNA 纳米金探针在合适条件下于玻片上发生杂交反应。因为不同的变量其相应探针的纳米粒子尺寸不同, 利用荧光分析法^[16] 采用光纤照明器照亮以识别不同尺寸纳米颗粒散射出的颜色。若求最大值, 就找颜色最多的单链; 若求最小值, 就找颜色最少的单链。最后, PCR 扩增相应单链, 读取最优解。

3.2 相关证明

命题 1 第 3 章中给出的转换前 0-1 规划的可行解与转换后的可行解等价。

证明 经过扩充相等变量, 添加等价变量和约束条件等之后得到的转换后的 0-1 整数规划为

$$\begin{aligned} \max(\min) z' &= \sum_{i=1}^k d_i x_i \\ \text{s.t.} \quad &\begin{cases} \sum_{j=1}^k s_{ij} x_j \geq (\leq, =) b_i & i = 1, \dots, l \\ d_j s_{ij} = 0, 1, x_j = 0, 1 & i = 1, \dots, l; j = 1, \dots, k \end{cases} \end{aligned} \quad (3)$$

1) 证明式 (1) 的可行解对应的变化之后的解为式 (3) 的可行解。

对任意式 (1) 的可行解, 不失一般性, 假设 $(1, \rho, \dots, 1)$ 为 n 维解向量。现只证明最简单的一种情况, 即只有一个变量需要进行等价扩增的情形。由于剩余情形都是进行等价的线性转换, 故剩余情况可理解为在现有情况下有限次累加的结果。不妨假设式 (1) 中只有 i th 约束条件的变量 x_i 需要进行扩增, 且 $t_i = 2$, 于是将 x_i 扩增为两个相等变量 x_0 和 x_1 。若 i th 原条件为 $2x_i + x_2 + \dots + x_n \geq (\leq, =) b_i$, 其转换结果为该约束条件变为。

$$x_0 + x_1 + x_2 + \dots + x_n \geq (\leq, =) b_i$$

同时, 为了构造方便, 只添加约束条件 $x_0 + \bar{x}_1 = 1$ (即此时只添加新变量 x_0 和 \bar{x}_1)。由 $(1, \rho, \dots, 1)$ 对应得到 $n+2$ 维解向量为

$$\begin{pmatrix} 1, & 1, & 0, & \dots, & 1 \\ (x_0 & x_1 & \bar{x}_1 & & x_n) \end{pmatrix}$$

显然满足式 (3) 中除了 i th 约束条件之外的约束条件。而对于 i th 约束条件, 带入后与原结果一致, 故该解向量为式 (3) 的可行解向量。

2) 对于证明变换后的式 (3) 的可行解为式 (1) 的可行解。

同样, 只证明式 (3) 是经过式 (1) 进行了一步等价转换得到, 不妨设与上一步证明一致的转换。其可行解为 $(x_0, x_1, \bar{x}_1, \dots, x_n) = (1, 1, 0, \dots, 1)$, 对应到原问题中同样满足 i th 约束条件, 目标函数值同样保持一致。

命题 2 通过所设计的筛选满足约束条件的操作, 可在所制备的 $(2^n n!)$ 条链找到全部可行解。

证明 $i = 1$ 时, 将满足第一个约束条件的探针放入数据池, 由于 $(2^n n!)$ 条链表示了所有变量组合的全部排列, 故可以按所制备的探针上变量的顺序筛选出满足第一个约束条件的 DNA 链。

设现在按照第 i 个约束条件筛选, 即此时在前 $i-1$ 个约束条件已满足的数据池中进行筛选。不失一般性, 假设 i th 约束条件为 $x_1 + x_2 + \bar{x}_3 \geq 2$, 满足约束条件的组合为 $x_1 x_2, x_1 \bar{x}_3, x_2 \bar{x}_3$, 制备相应 $x_1 x_2, x_1 \bar{x}_3$ 和 $x_2 \bar{x}_3$ 的探针, 而要找的就是至少满足三种情况之一的 DNA 链, 不妨设满足 $x_1 = 1, \bar{x}_3 = 1$ 的 DNA 链在筛选了前 $i-1$ 次后仍然存在于数据池中。

假设 $x_1 \bar{x}_3$ 无法与当前数据池中的 DNA 链结合, 说明变量 x_1 和 \bar{x}_3 前后相邻且满足前 $i-1$ 个约束条件的 DNA 链不存在, 即 $\dots x_1 \bar{x}_3 \dots$ 这种排列情况不存在, 这与 $(2^n n!)$ 包含所有变量组合的各种排列相矛盾, 该假设不成立。于是 $x_1 \bar{x}_3$ 一定能与当前数据池中的 DNA 链结合形成三链结构, 否则 $x_1 = 1, \bar{x}_3 = 1$ 这种情况至少不满足前 $i-1$ 个条件之一。故在第 i 步一定可以筛选出满足前 i 个约束条件的 DNA 链, 否则可行解集为空集。当 $i = 2 \rightarrow m$, 就可以筛选出满足问题约束条件的可行解, 或者可行解集为空集。

命题 3 $2^n n! \leq (2n)!$

证明 由于 $n \geq 1$, 得到 $2^n n! > 1, (2n)! > 1$, 于是比较 $\log 2^n n!$ 和 $\log (2n)!$ 。

$$\frac{\log 2^n n!}{\log (2n)!} = \frac{\sum_{i=1}^n \log i + n \log 2}{\sum_{i=1}^n \log i + \sum_{j=n+1}^{2n} \log j}$$

当 $n = 1$ 时, 可验证 $2^n n! = (2n)!$;

当 $n > 1$ 时, $n \log 2 < \sum_{j=n+1}^{2n} \log j$, 故 $\frac{\log 2^n n!}{\log (2n)!} < 1$ 。

综上, 有 $2^n n! \leq (2n)!$

因此, 根据命题 3 可见, 该模型的计算空间复杂度为 $O(2^n n!)$, 要少于文献[1]计算模型中的 $O((2n)!)$, 并且在解的筛选上更加方便。

4 改进算法的实例

4.1 问题描述

现实生活中许多案例可以转换为 0-1 整数规划问题。例

如某校某学院准备开展新一年研究生招生复试及录取工作,经开会讨论决定,由院中甲、乙、丙、丁、戊五名教师于近期着手部署并落实复试考生报到、体检、面试、审核四项主要工作。具体日程安排如下:

- a) 迎接复试考生到本院报到,至少需要两名值班教师,经询问,五名教师中甲、乙和丁有时间到场组织安排。
- b) 组织本院考生到校医院参加体检,该工作至多指定一名教师随行,且只有乙、丙和戊三名教师有时间参加。
- c) 确定考生面试时间、地点,进行研究生复试面试,至少需要两名值班教师,且五名教师中除甲外均可以到场负责。
- d) 审核复试考生成绩及其他相关信息,公布拟录取考生名单,该工作最多需要一名教师,只能从甲和丁两名教师中选择指派该任务。

试问如何将四项工作指派给相应的教师,在研究生复试及录取工作顺利完成的条件下,使参与负责的教师人数最少?

该问题属于指派问题中的 0-1 指派,设甲、乙、丙、丁、戊五位教师为 x_1, x_2, x_3, x_4, x_5 , 将其转换为如下形式的 0-1 整数规划问题:

$$\begin{aligned} \min z &= x_1 + x_2 + x_3 + x_4 + x_5 \\ \text{s. t. } &\begin{cases} x_1 + x_2 + x_4 \geq 2 \\ x_2 + x_3 + x_5 \leq 1 \\ x_2 + x_3 + x_4 + x_5 \geq 2 \\ x_1 + x_4 \leq 1 \\ x_1, x_2, x_3, x_4, x_5 = 0, 1 \end{cases} \end{aligned}$$

4.2 求解过程

对于给定的案例,首先构造 10 种短的 ODN,其中五种为 x_1, x_2, x_3, x_4, x_5 表示 $x_i (i = 1, 2, 3, 4, 5)$ 取 1,另外五种为 $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \bar{x}_5$ 表示 x_i 取 0,而且 x_1, x_2, x_3, x_4, x_5 和 $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \bar{x}_5$ 的补链也随之确定。

具体编码如下:

- $x_1 = 5' - \text{CAACCTTGAAACTTGGGATA} - 3'$
- $x_2 = 5' - \text{AAAGGAGTTCAGGAAGAGA} - 3'$
- $x_3 = 5' - \text{CCTCTAAGGGAGGAGTTGAC} - 3'$
- $x_4 = 5' - \text{GCCTAGAGGAAAGAAGGAGT} - 3'$
- $x_5 = 5' - \text{AAGAGAAAAGGGATATTCTCT} - 3'$
- $\bar{x}_1 = 5' - \text{GTTGGAACCTTTGAACCTTAT} - 3'$
- $\bar{x}_2 = 5' - \text{TTTCTCAAGGTCCTTCTCT} - 3'$
- $\bar{x}_3 = 5' - \text{GGAGATTCCTCTCAACTG} - 3'$
- $\bar{x}_4 = 5' - \text{CGGATCTCCTTCTTCTCA} - 3'$
- $\bar{x}_5 = 5' - \text{TTCTCTTTCCCTATAAGGA} - 3'$

利用以上 10 种 ODN 生成 $(2^5)!$ (3840) 种 DNA 单链,再经过 PCR 扩增成双链 DNA 构成初始数据池。同时,将 x_1, x_2, x_3, x_4, x_5 的补链用不同大小的纳米粒子设计成探针。

对于第一个约束条件 x_1, x_2, x_4 中有两个取 1 就可以满足条件,为此制作 $x_1x_2, x_1x_4, x_2x_4, x_1x_2x_4$ 对应的四种单链探针如图 1 所示,进而将这些探针构造成为核蛋白细状物。

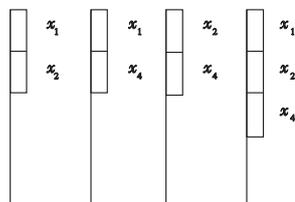


图1 四组探针

将处理好的探针与初始数据池混合反应生成稳定的三链 DNA 结构,通过过滤和提纯可获得满足第一个约束条件的解序列。

对以下三个约束条件进行相同的操作,得到满足全部约束条件的解序列。通过 PCR 扩增,并将扩增后的产物与目标函数中变量的 DNA 纳米金探针反应散射出的颜色判断该问题的最优解 $\min z = 2$, 并且读取最优解为 $x_1 = 0, x_2 = 1, x_3 = 0, x_4 = 1, x_5 = 0$, 即由乙和丁两位教师负责该院研究生招生工作,在保证其顺利进行的基础上使安排的教师人数最少。

该方法充分利用了三链 DNA 结构的稳定性,上述问题按照改进之后的编码方式需要生成 3 840 种 DNA 分子,而采用文献 [1] 中的编码方式,其所需分子种类为 $(2 \times 5)! = 3\ 628\ 800$ 种,大大改善了原有编码方法的空间复杂度。同时,改进的方法可以保证第 n 次筛选出来的 DNA 链都是满足前 n 个约束条件的双链,便于最优解的检测。

5 结束语

本文介绍了基于三链 DNA 结构的 0-1 整数规划求解算法,通过给出由一般的 0-1 整数规划向系数为 0-1 的整数规划转换的过程,实际上解决了一般的整数规划问题。该算法改进了原算法在可行解的读取和最优解的获得两方面的缺陷和不足,通过对 0-1 规划中 n 个变量进行全排列,改变了变量的编码顺序,将 $(2n)!$ 种解序列改进为 $2^n n!$ 种解序列,并通过 DNA 纳米金探针与产物发生杂交反应时散射出的不同颜色获得最优解,实现了最优解的快速获取。模型中,由于初始数据池为稳定的双链结构,同时第三条链由于形成了蛋白质细状物,故不会出现单链的发夹、错配等结构,从而使计算过程中反应更加充分、效率更高,同时错配率也随之降低。然而,该算法仍存在一定缺陷,如果目标函数的价值系数较大,在将其转换为简单的 0-1 整数规划的过程中需要新增的变量数目也较大。另外,如何进一步降低初始数据池中的分子数量,值得在理论和实验上进一步探索。

参考文献:

- [1] 杨静,殷志祥. 基于抗原中介三链 DNA 结构的 0-1 整数规划[J]. 计算机工程与应用 2008, 44(2): 76-79.
- [2] 杨静. 基于三链 DNA 结构的 0-1 整数规划[D]. 淮南: 安徽理工大学 2008.
- [3] 方刚,张社民,朱岩,等. 基于三链核酸的 DNA 计算[J]. 生物信息学 2009, 7(3): 181-185.
- [4] FANG Gang, XU Jin, ZHANG She-min. Performing DNA computation with RecA-mediated triple-stranded DNA structure[J]. Natural Computing 2009, 8(1): 149-156.
- [5] FANG Gang, ZHANG She-min, XU Jin et al. A novel DNA computing model based on RecA-mediated triple-stranded DNA structure[J]. Progress in Nature Science 2007, 17(6): 708-711.
- [6] YANG Jing, YIN Zhi-xiang, HUANG Kai-feng. The working operation problem on triple-stranded DNA structure model[C]//Proc of WRI Global Congress on Intelligent Systems. Washington DC: IEEE Computer Society 2009: 295-299.
- [7] 殷志祥,张凤月,许进. 0-1 规划问题的 DNA 计算[J]. 电子与信息学报 2003, 25(1): 62-66.
- [8] YIN Zhi-xiang, ZHANG Feng-yue, XU Jin. The general form of 0-1 programming problem based on DNA computing[J]. Biosystems, 2003, 70(1): 73-78.

(下转第 77 页)

度量指标,提出了基于动机倾向性的标签推荐模型(OMDM),并设计了基于该模型的算法。在两个不同的数据集上的实验结果表明,基于动机倾向性的社会标签推荐模型能获得更准确的备选标签。

本文在标签推荐研究上提供了一个新的研究视角,但研究本身还存在很多局限性。本文仅从五个观察指标对倾向性进行了度量,是否还存在其他不同质的度量指标,或者说从心理学、认知科学的理论角度有没有其他的指标来测量倾向性还有待研究。另外,本文提出的模型和算法并没有考虑所推荐的标签的新颖性、召回率,这也是笔者准备进行扩展的研究内容。

参考文献:

- [1] SIGURBJÖRNSSON B, Van ZWOL R. Flickr tag recommendation based on collective knowledge [C]//Proc of the 17th International Conference on WWW. New York: ACM Press 2008:327-336.
- [2] HOTH O A, JÄSCHKE R, SCHMITZ C, et al. Information retrieval in folksonomies: search and ranking [C]//Proc of the 3rd European Conference on the Semantic Web. Berlin: Springer 2006:411-426.
- [3] SYMEONIDIS P, NANOPOULOS A, MANOLOPOULOS Y. A unified framework for providing recommendations in social tagging systems based on ternary semantic analysis [J]. *IEEE Trans on Knowledge and Data Engineering* 2010, 22(2): 179-192.
- [4] HARVEY M, BAILLIE M, RUTHVEN I, et al. Tripartite hidden topic models for personalised tag suggestion [C]//Proc of the 32nd European Conference on IR Research. Berlin Springer: 2010: 432-443.
- [5] Motivation [EB/OL]. [2012-02-20] <http://en.wikipedia.org/wiki/Motivation> cite_note-0.
- [6] TOM C. Two cultures of fauxnomies collide [EB/OL]. [2012-02-22]. http://www.plasticbag.org/archives/2005/06/two_cultures_of_fauxnomies_collide.
- [7] HAMMOND T, HANNAY T, LUND B, et al. Social bookmarking tools (1): a general review [J]. *D-Lib Magazine* 2005: 11(4).
- [8] HECKNER M, HEILEMANN M, WOLFF C. Personal information management vs. resource sharing: towards a model of information behaviour in social tagging systems [C]//Proc of International AAAI Conference on Weblogs and Social Media. 2009: 42-49.
- [9] WASH R, RADER E. Public bookmarks and private benefits: an analysis of incentives in social computing [J]. *Journal of the American Society for Information Science and Technology* 2007, 44(1): 1-13.
- [10] AMES M, NAAMAN M. Why we tag: motivations for annotation in mobile and online media [C]//Proc of SIGCHI Conference on Human Factors on Computing Systems. New York: ACM Press: 2007: 971-980.
- [11] NOV O, NAAMAN M, CHEN Ye. Motivational, Structural and tenure factors that impact online community photo sharing [C]//Proc of the 3rd International AAAI Conference on Weblogs and Social Media. 2009: 138-145.
- [12] SINHA R. A cognitive analysis of tagging [EB/OL]. [2012-02-24]. <http://rashmisisinha.com/2005/09/27/a-cognitive-analysis-of-tagging>.
- [13] MARLOW C, NAAMAN M, BOYD D, et al. HT06, tagging paper, taxonomy, Flickr, academic article, to read [C]//Proc of the 17th Conference on Hypertext and Hypermedia. New York: ACM Press, 2006: 31-40.
- [14] XU Zhi-chen, FU Yun, MAO Jian-chang, et al. Towards the semantic Web: collaborative tag suggestions [C]//Proc of Collaborative Web Tagging Workshop at the WWW. New York: ACM Press 2006.
- [15] SEN S, LAM S K, RASHID A M, et al. Tagging communities, vocabulary evolution [C]//Proc of the 20th ANNIVERSARY Conference on Computer Supported Cooperative Work. Berlin: Springer, 2006: 181-190.
- [16] STROHMAIER M, KÖRNER C, KERN R. Why do users tag? detecting users' motivation for tagging in social tagging systems [C]//Proc of ICWSM. 2010: 339-342.
- [17] JIANG J J, CONRATH D W. Semantic similarity based on corpus statistics and lexical taxonomy [C]//Proc of International Conference on Research Computational Linguistics. 1997: 19-33.
- [18] FU W T, KANNAMPALLIL T G, KANG Ruo-gu, et al. Semantic Imitation in social tagging [J]. *ACM Trans on Computer-Human Interaction* 2010, 17(3): 1-37
- [19] 范克新. 社会学定量方法 [M]. 南京: 南京大学出版社, 2004: 346-358.
- [20] LANDIS J R, KOCH G G. The measurement of observer agreement for categorical data [J]. *Biometrics*, 1977, 33(1): 159-174.
- [21] SALTON G, MCGILL M J. Introduction to modern information retrieval [M]. [S. l.]: McGraw-Hill, 1983.
- [22] CILIBRASI R, VITANYI P M B. The Google similarity distance [J]. *IEEE Trans on Knowledge and Data Engineering* 2007, 19(3): 370-383.
- [23] 靳延安, 李玉华, 刘行军. 不同粒度标签推荐算法的比较研究 [J]. *计算机应用研究* 2012, 29(2): 504-509.
- [24] 窦玉萌, 赵丹群. 协作标注系统研究综述 [J]. *现代图书情报技术* 2009, 3(2): 9-17.

(上接第59页)

- [9] ZHOU Kang, GAO Zun-hai, XU Jin. An algorithm of DNA computing on 0-1 planning problem [J]. *Advances in Systems Science and Applications* 2005, 5(4): 587-593.
- [10] WANG Shi-ying, YANG Ai-ming. DNA solution of integer linear programming [J]. *Applied Mathematics and Computation* 2005, 170(1): 626-632.
- [11] ZHANG Feng-yue, YIN Zhi-xiang, LIU Bo, et al. DNA computation model to solve 0-1 programming problem [J]. *Biosystems* 2004, 74(1-3): 9-14.
- [12] 罗海波. 基于 0-1 规划的 DNA 计算模型的设计与实现 [D]. 沈阳: 东北大学, 2008.
- [13] 许进, 周康, 覃磊, 等. 0-1 规划问题的闭环 DNA 算法 [J]. *系统工程与电子技术* 2009, 31(4): 947-951.
- [14] BRAICH R S, CHELYAPOV N, JHONSON C, et al. Solution of a 20-variable 3-SAT problem on a DNA computer [J]. *Science* 2002, 296(4): 499-502.
- [15] SANCHES C A A, SOMA N Y. A polynomial-time DNA computing solution for the bin-packing problem [J]. *Applied Mathematics and Computation* 2009, 215(6): 2055-2062.
- [16] 孙伟, 尤加宇, 江宏, 等. 纳米粒子标记 DNA 探针的制备与检测应用 [J]. *中国卫生检验杂志*, 2005, 15(8): 1008-1010.