

DNA 芯片在 0-1 规划问题中的应用*

张凤月^{1) **} 殷志祥^{1, 2)} 许进¹⁾

(¹华中科技大学控制科学与工程系, 武汉 430074; ²淮南工业学院数理系, 淮南 232001)

摘要 生物芯片技术和 DNA 计算分别是近年来生命科学与信息科学的新兴研究领域, 对信息高度并行的获取与处理是二者本质特性。而 0-1 规划问题作为运筹学中一个重要的问题, 到目前为止还没有好的算法。在 DNA 计算和 DNA 芯片基础上, 提出了基于 DNA 芯片解决 0-1 规划问题的 DNA 计算新模型, 与以往 DNA 计算模型相比, 该模型具有高信息量和操作易自动化的优点。同时指出 DNA 芯片技术有望作为新型生物计算的芯片。

关键词 DNA 芯片, DNA 计算, 0-1 规划问题

学科分类号 TP301.6, Q78

利用 DNA 分子解决计算问题基于两个原则: DNA 分子的 Watson-Crick 互补原则和 DNA 链生化反应的巨大并行性。前者可以转换为计算机的通用语言^[1], 而后者使利用穷举法解决困难类和不可计算类问题成为可能。1994 年 Adleman^[2]以 DNA 序列作为信息载体, 利用分子生物学试验技术解决了一个 NP 完全问题——Hamilton path problem 问题, 并指出 DNA 分子计算具有高信息度、高并行度的特点, 从而允许 DNA 计算机在线性时间内求解象 NP 完全问题之类更困难的问题(与此相反的是图灵机求解这类问题所要求的时间是指数递增的)。之后, 许多学者提出了不同的 DNA 计算模型, 并相继解决了图的最大团问题、“骑士”问题、最大独立集问题和 SAT 问题 (the satisfiability problem) 等^[3~6]。其中, Liu 等^[7]提出在载体表面解决 SAT 问题的模型, 提示 DNA 计算从理论模型论证到应用计算芯片研究的可能性。Wu^[8]对此方法又做了进一步的改进。但 DNA 计算理论模型也存在算法所需要的核苷酸分子数量呈指数增加, 误差较大, 常规生化技术耗时费力等不足。

DNA 芯片 (DNA chip) 是由 Forder 等^[9]提出并在 1996 年研制成功的。其实质是含有按特定排列方式固定于一定位置可寻址的基因探针的硅片、玻片、塑料片等, 在一定条件下进行生化反应, 反应结果经标记后通过荧光扫描仪器扫描及计算机分析, 即可获得样品中大量基因序列表达的信息。与常规生物技术相比, DNA 芯片最大的特点在于其对生物信息高通量并行性的获取和处理。目前 DNA 芯片技术已在基因表达谱分析、新基因发现、基因突变及多态性分析、基因组文库作图、疾病诊

断和预测、药物筛选等研究领域表现出强大的生命力和良好的应用前景^[10]。

本文结合目前先进的 DNA 芯片技术和 DNA 计算理论, 给出 0-1 规划问题的 DNA 计算模型以及基于 DNA 芯片的 0-1 规划问题 DNA 计算新模型, 该模型的提出使利用 DNA 计算解决规划问题成为可能, 并且由于 DNA 芯片技术的应用, 在一定程度上弥补了以往 DNA 计算模型的不足, 在利用生物大分子解决计算问题的研究领域中有巨大的应用潜能。

1 0-1 规划问题的数学模型

0-1 规划问题是整数规划的特殊情形, 它的变量 x_i 仅取值 0 或 1, 这时 x_i 称为 0-1 变量, 或者二进制变量。它的一般形式为:

$$\begin{aligned} \max(\min) z = & c_1x_1 + c_2x_2 + \dots + c_nx_n \\ \left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \leqslant (=, \geqslant) b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \leqslant (=, \geqslant) b_2 \\ \dots \dots \dots \dots \dots \dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \leqslant (=, \geqslant) b_m \end{array} \right. \\ x_1, x_2, \dots, x_n = & 0, 1 \end{aligned}$$

0-1 规划问题是运筹学中的一个重要问题, 其应用非常广泛。关于它的算法很多, 如穷举法, 隐枚举法等, 但到目前为止还没有好的算法。本文利用 DNA 计算解决如下一种特殊的 0-1 规划问题, 它是指派问题的推广。

* 国家自然科学基金资助项目 (60274026, 60174047)。

** 通讯联系人。

Tel: 027-87543563, E-mail: zfywuhan@mail.hust.edu.cn

收稿日期: 2002-11-19, 接受日期: 2003-02-27

$$\begin{aligned}
 & \max(\min) z = c_1x_1 + c_2x_2 + \dots + c_nx_n \\
 a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n & \leqslant (=, \geqslant) b_1 \\
 a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n & \leqslant (=, \geqslant) b_2 \\
 & \dots \\
 a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n & \leqslant (=, \geqslant) b_m \quad (\star) \\
 x_1, x_2, \dots, x_n, a_{ij} & = 0, 1; \\
 c_i, b_j & \text{为非负整数} \\
 (i &= 1, 2, \dots, m; j = 1, 2, \dots, n)
 \end{aligned}$$

下面讨论 (\star) 所对应的 0-1 规划问题的算法。步骤 1：生成给定问题的变量取值为 0 或 1 的所有可能组合；步骤 2：利用每一约束条件剔除非可行解；步骤 3：保留生成剩余的解；步骤 4：重复进行步骤 2, 3，排除所有的非解，从而得到问题的所有可行解；步骤 5：比较各可行解对应的目标函数值，进而得到最优解。

2 0-1 规划问题的 DNA 计算模型

对应于上述算法的 DNA 计算模型和算法简单描述为 6 个步骤。步骤 1：制作代表给定问题的变量的 DNA 链（寡聚核苷酸链）及其互补链；步骤 2：根据 Watson-Crick 互补原则，将 DNA 链进行杂交，从而获得代表一个给定问题的变量取值为 0 或 1 的所有可能组合的 DNA 数据库；步骤 3：用一定的生物方法对结果标记，使其可以检测到判断满足或不满足约束条件的 DNA 组合；步骤 4：根据约束方程排除所有的不满足该方程条件的组合，保留满足的组合；步骤 5：重复进行步骤 3，继续排除不满足条件的组合（对于步骤 3 已经判断为不满足的不再考虑），这样可排除所有不满足条件的组合，从而也就判断了给定问题的可行解；步骤 6：对各可行解计算出目标函数的值，进而判断出最优解。

对于 a_{ij} 为正整数时，只须将 $a_{ij}x_j$ 表示为 $\sum_j^{a_{ij}} x_j$ 即可。

本文的其他术语和记号参见文献 [11]。

3 0-1 规划问题的 DNA 芯片模型

对于一个含有 n 个变量 x_1, x_2, \dots, x_n 和 m 个方程的方程组。对应算法的步骤 1，可分为两步。第一步首先合成 $3n$ 种共 3 组 DNA 片段。第一组的 n 种 DNA 片段分别表示变量 x_1, x_2, \dots, x_n ；第二组的 n 种 DNA 片段分别是变量 x_1, x_2, \dots, x_n 的互补链表示为 $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$ ，该组 DNA 片段经过荧光素标记；第三组的 n 种 DNA

片段同样是变量 x_1, x_2, \dots, x_n 的互补链，并分别记为 $\bar{x}'_1, \bar{x}'_2, \dots, \bar{x}'_n$ ，该组 DNA 片段未经过荧光素标记。构造 DNA 链要遵循以下原则：避免不同 DNA 片段的错误杂交，即代表变量的寡聚核苷酸链应具有较大差异；尽量保证代表变量的寡聚核苷酸链的 GC 含量为 50%，这样能够保证在之后的反应过程机会均等。（注意这里 x_i 和 \bar{x}'_i 对应的寡聚核苷酸杂交表示变量 x_i 取值为 1，而 x_i 和 \bar{x}'_i 所对应的寡聚核苷酸杂交表示变量 x_i 取值为 0）。第二步是将第一步第一组合成的 DNA 链固定在芯片的表面上，以包含全部 n 个变量的 DNA 链为一组，重复排列为 DNA 矩阵，组数大于 2^n （必须指出，对于含 n 个变量的方程，其可能解为 2^n ，因此要求该模型的点样组数大于 2^n ，以保证下一步杂交可以产生所有可能解。不可否认，也有产生重解的可能，尽管重解不会影响最终计算结果）。

对应算法的步骤 2，将一定量的第二组和第三组合成的 DNA 链和上步表面的 DNA 矩阵进行杂交。在严格条件下杂交，之后在适当条件下用缓冲液清洗掉未杂交上的 DNA 链。这一步骤杂交将会形成包含该问题所有可能解的 DNA 数据库。

对应算法的步骤 3，因为在第一步已经对第二组 DNA 链进行了荧光标记，所以当某一个变量取值 1 时，对应为呈现一定荧光颜色，当其值取 0 时，对应为无色。这样就可以在下一步将正确解找出。

对应算法的步骤 4，利用激光共聚焦显微镜或 DNA 芯片扫描仪获取反应后 DNA 矩阵的图像信息并进行分析，保留满足给定方程的可能解，排除不满足的解。

对应算法的步骤 5，重复步骤 4（没有必要再一次重复获取图像资料，也就是说对芯片信息的一次获取即可进行多个约束方程的分析），对其他方程进行可行解的筛选。

对应算法的步骤 6，对各可行解计算出目标函数的值，进而判断出最优解。下面具体讨论一个简单 0-1 规划问题。

$$\begin{aligned}
 \min u &= 2x + y + 3z \\
 \begin{cases} x + z \leqslant 1 \\ x + y + z \geqslant 2 \\ y + z \leqslant 1 \\ x, y, z = 0, 1 \end{cases}
 \end{aligned}$$

步骤 1：首先构造 3 种寡聚核苷酸链，序列为 AACCTGGT, ACGATAGC 和 AGAGTCTC，分别

代表变量 x, y, z , 同时构造这 3 种寡聚核苷酸链的互补链, 序列为 TTGGACCA, TGCTATCG 和 TCTCAGAG, 对应分别代表 $\bar{x}, \bar{y}, \bar{z}$ 及 $\bar{x}', \bar{y}', \bar{z}'$. 然后将代表 x, y, z 变量 DNA 的 5' 端进行巯基修饰, 将代表 $\bar{x}, \bar{y}, \bar{z}$ 的 3 种寡聚核苷酸 5' 端进行荧光标记, 标记物为荧光黄 (FAM).

步骤 2: 将 5' 端进行巯基修饰了的 DNA 片段 (代表 x, y, z 变量), 按一定排列、一定量点样于事先已处理好的芯片上, 点样矩阵大于 8, 图 1 示意芯片点样矩阵 (点样矩阵是可寻址的). 在实际应用中, 必须设置阳性和阴性对照, 来控制反应条件.

	1	2	3	4	5	6	7	8
x	(x)							
y	(y)							
z	(z)							

Fig. 1 DNA array on chip

步骤 3: 将一定量代表 x, y, z 和 $\bar{x}, \bar{y}, \bar{z}$ 的 DNA 链进行均匀混合, 混合物在合适条件下对芯片矩阵进行杂交反应 (如 42°C, 6 h), 反应后用一系列缓冲液在严格条件下冲洗芯片, 使没有杂交上的核苷酸被清洗掉. 该杂交将产生所有可能解 (图 2).

	1	2	3	4	5	6	7	8
x	(x̄)	(x̄)	(x̄)	(x̄)	(x̄̄)	(x̄̄)	(x̄̄)	(x̄̄)
y	(ȳ̄)	(ȳ̄)	(ȳ̄)	(ȳ̄)	(ȳ̄)	(ȳ̄̄)	(ȳ̄̄)	(ȳ̄̄)
z	(z̄̄)	(z̄̄)	(z̄̄)	(z̄̄)	(z̄̄̄)	(z̄̄̄)	(z̄̄̄)	(z̄̄̄)

Fig. 2 DNA data library after hybridization

The value of variable x, y, z were: 1 (1, 1, 1), 2 (1, 1, 0), 3 (1, 0, 1), 4 (1, 0, 0), 5 (0, 1, 1), 6 (0, 1, 0), 7 (0, 0, 1) and 8 (0, 0, 0) respectively.

步骤 4: 利用激光共聚焦显微镜对芯片进行观察, 记录并分析结果. DNA 杂交具有阳性结果 (荧光亮点, 图 2 中用黑色半圆表示) 表示该变量取值为 1, 无色为 0. 对于此约束方程组的第一个约束方程 $x + z \leq 1$, 解 1, 3 列中代表变量的 x ,

z 的 DNA 杂交片段都呈现荧光, 表示 x, z 同时取值为 1, 该组解不满足约束方程. 同理, 可以得到满足该约束方程的可能解对应为图 2 中 2, 4, 5, 6, 7, 8 列.

步骤 5: 在上一步的基础上, 对约束方程组的第二个约束方程 $x + y + z \geq 2$ 进行类似分析, 其 DNA 链对应的编码变量满足该约束方程的可能解是 2, 5 列 (图 2). 同理分析第三个约束方程得到可满足解为第 2 列 (图 2).

步骤 6: 这样得到了满足约束方程组的变量 (可行解), 比较它们的目标函数值即可求出最优解. 对于本问题可行解仅为第 2 列, 因此第 2 列对应的编码变量值 (1, 1, 0) 也是最优解, 对应的目标函数最小值为 3.

4 讨 论

本文对运筹学中一个重要问题即 0-1 规划问题给出了 DNA 计算模型, 并且提出了基于 DNA 芯片解决 0-1 规划问题的 DNA 计算模型. 该模型将问题的变量映射为 DNA 链, 在芯片上制备可寻址的 DNA 变量矩阵, 通过加入代表变量 DNA 链的互补链 (其中代表变量取值为 1 的 DNA 互补链以荧光标记, 取值为 0 的不标记) 后, DNA 分子就会按照 Watson-Crick 互补原则进行杂交, 从而并行地生成问题的所有可能解. 之后通过对杂交后 DNA 芯片图象进行分析即可得到问题的最优解. 和以往的 DNA 计算模型相比, 该模型由于运用了 DNA 芯片技术, 具有高密度样品矩阵, 有可能解决更多个变量的 0-1 规划问题. 对于每个约束方程对应的生物学操作, 如杂交、冲洗、检测和结果记录与分析, 均可以实现自动化, 这样就可以在一张芯片上同时解决多个约束方程的生物学操作, 这也意味着有可能用一张芯片来解决简单的 0-1 规划问题. 该模型的提出对 DNA 计算有一定的促进作用, 使用的技术是在 DNA 芯片技术中常用和比较成熟的, 因而是可以实现的. 随着分子生物学及生物工程, 特别是生物芯片相关技术进一步发展, 该方法的进一步扩展有可能解决一般的规划问题. 本文也力图通过对简单问题的求解来体现 DNA 芯片在 DNA 计算芯片研究中的潜在价值. 我们相信, DNA 计算机新理论与生物芯片新技术的结合, 必将对生命科学与计算机科学产生积极的影响.

参考文献

- 1 Freund R, Paun G, Rozenberg G, et al. Watson-Crick finite automata. In: Rubin H, Wood D H, eds. DNA based Computers III, 48. Philadelphia: American Mathematical Society, 1999. 305~317
- 2 Adleman L M. Molecular computation of solutions to combinatorial problems. Science, 1994, 266 (5187): 1021~1023
- 3 Lipton R J. DNA solution of hard computation problem. Science, 1995, 268 (5210): 542~545
- 4 Ouyang Q, Kaplan P D, Liu S M, et al. DNA solution of the maximal clique problem. Science, 1997, 278 (17): 446~449
- 5 Head T, Rozenberg G, Bladergroen R B, et al. Computing with DNA by operating on plasmids. Biosystems, 2000, 57 (2): 87~93
- 6 Sakamoto K, Gouzu H, Komiya K, et al. Molecular computation by DNA hairpin formation. Science, 2000, 288 (5): 1223~1226
- 7 Liu Q H, Wang L M, Fruto A G, et al. DNA computing on surfaces. Nature, 2000, 403 (13): 175~179
- 8 Wu H Y. An improved surface-based method for DNA computation. Biosystems, 2001, 59 (1): 1~5
- 9 Forder S P, Read J L, Pirrung M C, et al. Light-Directed, spatially addressable parallel chemical synthesis. Science, 1991, 251 (4995): 767~773
- 10 马立人, 蒋中华. 生物芯片. 北京: 化学工业出版社, 2000. 1~15
- 11 Ma L R, Jiang Z H. Biochip. Beijing: Chemical Industry Press, 2000. 1~15
- 12 Bondy J A, Murty U S R. Graph Theory With Applications. New York: The Macmillan Press TD, 1976. 28~89

Application of DNA Chip on 0-1 Planning Problem*

ZHANG Feng-Yue^{1)***}, YIN Zhi-Xiang^{1,2)}, XU Jin¹⁾⁽¹⁾ Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China;²⁾ Department of Math. and Phy., Huainan Technology Institute, Huainan 232001, China)

Abstract DNA chip and DNA computing are new research areas in biology science and information science separately. The essential characteristic of both is the massive parallel of obtaining and managing information. The 0-1 programming problem is an important problem in opsearch and has very widespread application. But up to now, there does not exist any good algorithm yet. A new DNA computing model is provided to solve a 0-1 planning problem based on DNA chip. The method has some significant advantages and the result suggests the potential of DNA chip used as a DNA computer chip.

Key words DNA chip, DNA computing, 0-1 planning problem

* This work was supported by a grant from The National Natural Sciences Foundation of China (60274026, 60174047).

** Corresponding author. Tel: 86-27-87543563, E-mail: zfywuhan@mail.hust.edu.cn

Received: November 19, 2002 Accepted: February 27, 2003