广义分子计算模型在0-1背包问题中的应用

杨 震[1] 马天宝[1] 余 文[2] 李艳梅[1]

(北京理工大学机电学院爆炸科学与技术国家重点实验室 北京 100081)[1]
(北京邮电大学智能通信软件多媒体北京市重点实验室 北京 100876)[2]

摘 要 生物分子计算在实现上有很多局限性。借鉴了广义图灵模型(Generalized Turing Model,GTM)[1]。该模型是由分子计算粘贴模型与图灵机相结合而得到的,并且已证明可以在多项式时间内准确获得0-1整数规划、集合覆盖等多个NP完全问题的全体可行解集。在此基础上将GTM应用于求解0-1背包问题,仿真展现了该模型的优点。

关键词 广义分子计算,图灵机 ,0-1背包问题

中图法分类号 TP301 **文献标识码** A

Application of Generalized Molecular Computation Model in 0-1 Knapsack Problem

YANG Zhen[1] MA Tian-bao[1] YU Wen[2] LI Yan-mei[1]

(State Key Laboratory of Explosive Science and Technology, Beijing Institute of Technology, Beijing 100081, China)[1]
(Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing 100876, China)[2]

Abstract Biomolecular computing has many limitations in implementation. In the literature, molecular computing sticker model and turing machine were combined to generate a generalized turing model GTM. The model has been proved to be accurate to get all the feasible solution set of mulitiple NP complete problems like 0-1 integer programming and set covering. In this paper, on this basis, GTM was applied to solve the 0-1 knapsack problem. The simulation shows the advantages of the model and further validation of the extensive application of the model.

Keywords Generalized molecular computation, Turing model, 0-1 knapsack problem

工程科学与生命科学的相互交叉、相互渗透和相互促进是现代科学技术发展的显著特点之一。现代科学的理论研究与科学实践中存在着大量的与优化、自适应相关的问题,对这些难题的解答一直是科学界的一个难题。人们在对生物的研究中发现自然界的生物可以通过学习、模仿去适应大自然,人们通过启发将自身进化这一思想引入到工程科学领域,典型的方法就是分子计算方法。

1994年,美国科学家L. Adleman在Science杂志上发表了他的研究成果[2],成为第一个在实验室中使用分子计算的方法解决数学问题的科学家,从此分子计算引起了世界各国相关专家的注意。分子计算研究领域主要包括分子计算机理、分子计算编码、分子计算模型、分子计算的实验方法与技术等方面。分子计算模型由于拥有很广泛的实际应用前景,并具有并行性高、低功耗等优点,因此被广泛应用于求解各类NP完全问题[3-9]。

分子计算经过近20多年的发展,已有多种模型被提出,如剪接模型、粘贴模型等。这些模型的作用都被证明与图灵机是等价的,且具有计算完备性。近些年分子计算在很多方面都取得了不错的进展,如序列编码、计算模型、检测手段等。例如Lipton在1995年对SAT问题给出了一种分子计算模型[10];2003年,Dirk等人在实验室里实现了Lipton的上述设计[11];2002年,Braich等人应用粘贴DNA计算模型对具有20个变量的可满足性问题成功地进行了求解,这是当时应用分子计算模型从规模上所得到的最好的结果。同时我国分子计算的研究也越来越引起科学家的关注,如2006年,许进等人基于粘贴模型构建了DNA计算机,并用该DNA计算机成功地对5阶顶点的顶点着色问题的DNA算法进行了实现;2008年,周康等人结合粘贴模型和DNA芯片模型的优势提出了八皇后问题的粘贴DNA芯片算法[12]。

上述模型大多具有较强的生物工程背景,在实现上有很多局限性。本文借鉴的广义图灵模型(GTM),不依赖某种特定技术。它由一台单带图灵机、一条单向只写带及只写头和一条工作带及读写网络组成。在只写带和工作带之间存在一个特殊的映射函数。通过一个特定GTM的形式描述,显示GTM具有指数的并行读写能力和指数的加速性能,可以在多项式时间精确地求解满足性问题的全体解集。该模型具有编码简单、错误率低、并行性高、低功耗等优点,已在多个NP完全问题中得到应用[12],如整数规划问题[13]和集合覆盖问题等,说明了该模型的应用的广泛性。本文将GTM应用于求解0-1背包问题,仿真展现了该模型的优点。

本文受国家自然科学基金资助项目(11272066)资助。

杨 震(1988—),男,硕士生,主要研究方向为广义分子计算,E-mail:yangzhen90259@163.com;**马天宝**(1982—),男,副教授,主要研究方向为爆炸力学的计算以及仿真;**余 文**(1964—),男,教授,主要研究方向为分子计算;**李艳梅**(1981—),女,博士,主要研究方向为分子计算。

1　广义图灵模型(GTM)

1.1　GTM 的结构

如图 1 所示,GTM 由 3 部分组成:1)一个基本图灵机,包括一个控制器 Q、一条单向无限延伸的基本带和一个读写头;2)一个有限的只写带及读写头;3)一个与只写相关的有限的工作带及读写网络,简称读写器。为了方便,基本带、只写带和工作带分别被编号为…,−2,−1,0,1,2,…、1,2,3,…,n 和 0,1,2,…,N,其中 $N=2^n-1$。

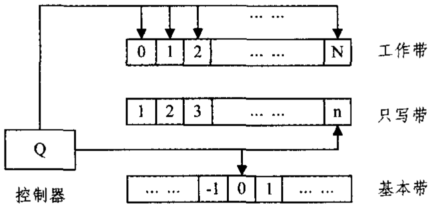

图 1　GTM 的结构

GTM 把分子计算中的 4 个碱基分别用 0、1、∗、Ø 4 个四值逻辑符号来代替。其中 0 表示没有被选中,1 表示被选中,∗ 表示 0 和 1 的叠加态,Ø 表示空字符。其逻辑包含关系如下:$* \supset 0, 1 \supset \emptyset$ 或 $* \supseteq 0, 1 \supseteq \emptyset$, $\Gamma=\{\emptyset, 0, 1, *\}$、$K=\{0,1,*\}$ 和 $\Sigma=\{0,1\}$ 分别是基本带、只写带和工作带的输入符号集。

基本带的作用是存储输入多值字符串,控制器 Q 通过读写头将基本带上多值字符串写入只写带中,在只写带和工作带之间存在着一个特殊的映射关系,在特定的状态下,通过有效的输入可以激活工作带上特定格子,被激活的格子是活动的。

只写带到工作带的映射函数 f 定义如下:对只写带中任意输入的多值字符 x,经过函数关系的映射变为 $f(x):=a_0a_1a_2\cdots a_{2^n-i}$,其中 $N=2^n-1$。输出 a_i(其二进制表示形式 $a_{i_1i_2i_3\cdots i_n}$)与工作带第 i(从 0 至 N)个格子一一对应。且输出为:

$$a_{i_1i_2i_3,\cdots i_n}=\begin{cases}1, & 若\ \forall A_k \supseteq i_k(k=1,\cdots,n)\\ 0, & 否则\end{cases}$$

例如写入只写带的多值字符串为 0…00∗,通过映射函数 f 可得,工作带上 0 号格子(0…000)和 1 号格子(0…001)被激活。

GTM 有两个特殊的工作状态:同时读和同时写。这两个状态是靠读写器控制的,通过读写器同时读或者写活动的格子,读写的格子数目是没有限制的。同时写可以并行地写入字符(0 或 1);在同时读时,如果活动的格子中只有“0(或 1)”(不包括空白符 Δ),则输出“0(1)”;如果既有“0”,又有“1”,输出为“∗”。

1.2　GTM 的工作原理

GTM 的工作是根据状态集 Q 来控制的。当 GTM 工作时,多值字符串会被基本带中间的有限个格子接受,其它的格子放置空白符,读写头放在输入的多值字符串的最左端,这时候只写带和工作带的内容都是空白的,只写带的带头在带子的最左端,GTM 处于 q_0 的状态,开始工作。

如果这个时候的状态为 $q\in Q-\{q_h,q_w,q_r\}$,那么可以按照普通的图灵机的工作原理来工作。而 GTM 有两个状态,一个是普通状态,另一个是特殊状态。当机器从普通状态进入特殊状态时,只写头位于最右端的格子,并且只写带上有符号行 $x\in K^n$,可以通过一步的计算完成下列各项:

1)映射函数 $f(x)\in\{0,1\}^N$ 使工作带上的一些格子处于活动的状态,这些活动的格子记做 s;

2)读写器同时扫描 s;

3)读写器将多值字符串 d 写入 s 中:如果操作是 q_w,就将 $d\in\{0,1\}$ 写入 s 中;如果操作是 q_h,那么在扫描 s 的同时,输出 $d\in\{0,1,*\}$,工作带内容保持不变;

4)读写器执行完上述的工作后,只写头回到最左边的位置并且只写带上的内容全部清空,机器回到普通状态,计算按照通常情况进行。如果当前的状态是 q_h,则表示计算结束。

2　0-1 背包问题

0-1 背包问题是一个典型的组合优化问题,在计算理论中属于 NP 完全问题,传统上采用动态规划来求解。本文是根据 GTM 模型来讨论 0-1 背包问题。

2.1　0-1 背包问题的描述

给定一个载重量为 M 的背包及 n 个物体,物体 i 的重量为 w_i、价值为 p_i,$1\leqslant i\leqslant n$,要求把这些物体装入背包,使背包内的物体价值总量最大。此处我们讨论的物体是不可分割的,通常称这种物体不可分割的背包问题为 0-1 背包问题[14,15]。

0-1 背包问题的特点是:每种物体只有一件,可以选择放或者不放。假设:x_i 表示物体 i 被装入背包的情况,$x_i=0,1$。当 $x_i=0$ 时,表示物体没有被装入背包;当 $x_i=1$ 时,表示物体被装入背包。于是该问题可以被简化为:

$$\sum_{i=1}^{n} w_ix_i \leqslant M \tag{1}$$

$$optp=\max\sum_{i=1}^{n} p_ix_i \tag{2}$$

这就是求解一个最优解的问题。

2.2　GTM 中 0-1 背包问题的分配

$$x_i=\begin{cases}1, & 第\ i\ 个背包被选中\\ 0, & 第\ i\ 个背包没被选中\end{cases}$$

其中,$i=1,2,\cdots,n$,将 0-1 背包问题 GTM 模型化后有便于用 GTM 解决该问题。只写带输入的多值字符串表示为 $x_1x_2x_3\cdots x_n(x_i\in K, i\in[1,n])$,经过映射函数 $f(x)$,共有 $N=2^n$ 个不同的值。这里用工作带上从 0 到 $N-1$ 的格子地址对应的二进制编码表示所有的值,我们可以用 $p_i(i=1,2,\cdots,n)$ 表示存储单元的内容。那么,满足约束条件且存储单元内容最大的即为背包问题的最优解,进而得到满足条件的 $x_1x_2x_3\cdots x_n(x_i\in K, i\in[1,n])$。

2.3　用 GTM 求解 0-1 背包问题的算法

GTM 求解问题的方法大体上分 3 步:1)对工作带格子进行初始化;2)将所有的 P 值写入到工作带中;3)删除不满足条件的 P 值,并得到最优解。

具体的步骤如下:

第 1 步:读取基本带内容,将 $x_0=\underbrace{**\cdots**}_{n}$ 写入只写带中,进入同时写状态,执行映射函数 f,工作带所有格子被选中并同时写入 0。

第 2 步:将 $x_i(i=1,2,\cdots,n)$ 写入只写带中。执行映射函数 f,将 P_i 与被激活的工作带单元格内容相加并存入该单元格。第 2 步操作进行 n 次。

第 3 步:利用约束条件(1)剔除非可行解(保留可行解)。即找出不满足约束条件的组合,并将这些组合对应的工作带的内容置空。重复进行第 2 步,可以排除所有的非解,从而得到问题的所有可行解。如果有解,则所有可行解中 P 值最大的对应的工作带地址的二进制编码是该问题的最优解。

3 实例说明

接下来用一个简单的例子来说明问题：

假设给定一个载重为 5.5 的背包以及 4 个物体，其中每个物体的重量是 3、2、1、2，对应的物品价值分别是 2、3、1、4，求该背包能装入的最大价值。

该问题可以抽象出以下两个关系式：

目标函数：$\max\ p=2x_1+3x_2+x_3+4x_4$；

约束条件：$3x_1+2x_2+x_3+2x_4\leqslant 5.5$。

将基本带字符串描述为

$w=WR\Delta x_0\Delta 0\Delta\Delta ADD\Delta x_1\Delta p_1\Delta\Delta\cdots\Delta\Delta ADD\Delta x_4\Delta p_4\Delta\Delta WR\Delta\cdots\cdots$

其中，Δ 是空白符，分割指令中的操作符和操作数；ΔΔ 代表分隔开两个不同的指令；*WR* 代表写操作指令，即将输入数据写入工作带格子；*ADD* 代表将 p_i 与被激活的工作带单元格内容相加并存入该单元格；*EM* 代表将存储单元内容写空。

1）求解过程：执行的操作是 $x_0\Delta WR0$，对存储单元清零，把 16 个格子的内容都写为 0。

2）在这里 $x_0=****$，$x_1=1***$，$x_2=*1**$，$x_3=**1*$，$x_4=***1$，P_1 表示选中组合的背包价值。每执行一次 $x_i\Delta WRp_i$ 我们就可以选中工作带上 8 个格子，通过执行 5 次这样的操作，并且把每次写入的值与上一次写入的值相加，我们就可以得到所有格子中的值；

3）根据约束条件我们排除非解，即把非解对应的存储单元的值置空；

4）最后一步是同时读，通过读取比较存储单元中的非空的解，计算工作带格子中的值，得到最优解，也就是数值最大的值。

表 1 中步骤 1 是将工作带中的格子内容全部写零。步骤 2 到步骤 5 是生成所有的 p 值。步骤 6—步骤 8 是将不满足条件的解删除，此时格子内容是空的。

根据表中内容可以清晰地看出工作带中对应地址为 0111 的格子中的值最大，也就是 p 最大为 8，这时候对应的 $x=0111$，回归到背包问题也就是除了第一件物品，其他的 3 件物品全部选中时背包内装入的价值最大，即为 8。

在求解的过程中，为获得 p 需要进行 $n+1$ 步。删除非真解需要 h 步（不同的 0-1 背包问题的约束条件不同，删除非真解需要的步骤也不同）。那么 0-1 背包问题的求解就需要 $n+1+h$ 步。

通过以上分析，获得 0-1 背包问题的所有可行解可在多项式时间内完成。而对于读出问题的解 0-1 背包问题需要对解进行比较后才能获得。

表 1　工作带中格子内容的变化过程

编号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
地址	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
步骤 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	000
步骤 2	0	0	0	0	0	0	0	0	2	2	2	2	2	2	2	2
步骤 3	0	0	0	0	3	3	3	3	2	2	2	2	5	5	5	5
步骤 4	0	0	1	1	3	3	4	4	2	2	3	3	5	5	6	6
步骤 5	0	4	1	5	3	7	4	8	2	6	3	7	5	9	6	10
步骤 6	0	4	1	5	3	7	4	8	2	6	3		5	9	6	10
步骤 7	0	4	1	5	3	7	4	8	2	6	3		5		6	10
步骤 8	0	4	1	5	3	7	4	8	2	6	3		1			

其中每一步执行的指令分别是：1）$WR\Delta x_0\Delta 0\Delta\Delta$；2）$ADD\Delta x_1\Delta p_1\Delta\Delta$；3）$ADD\Delta x_2\Delta p_2\Delta\Delta$；4）$ADD\Delta x_3\Delta p_3\Delta\Delta$；5）$ADD\Delta x_4\Delta p_4\Delta\Delta$；6）EMΔ1011ΔΔ；7）EMΔ1101ΔΔ；8）EMΔ111 *ΔΔ。

结束语　GTM 是由图灵机、只写带及读写头和工作带及读写网络组成的，该模型具有存数数据量大、并行性高的特点，能同时处理、存储指数个数据。该模型的应用目的就是通过大容量的空间存储来降低时间计算上的复杂度。本文通过仿真求解 0-1 背包问题实例，将时间复杂度转换为空间复杂度，从而在多项式的时间上精确地得到所求问题的全体解集，体现了 GTM 及该算法的有效性和优点。

数据读出方面是 DNA 计算的一个难题，DNA 计算通过删除非真解能找到问题的所有可行解，如何在多项式时间内将所有可行解读出需要进一步研究与探讨。

参考文献

[1] 余文，高荔，杨旭东，等. 一种基于分子计算的广义图灵模型[J]. 计算机科学（专刊），2006，33(7)：362-365

[2] Adleman L. Molecular computation of solutions to combinatorial problems[J]. Science，1994，266(11)：1021-1024

[3] Ouyang Q，Kaplan P D，Liu S，et al. DNA solution of the aximal clique problem[J]. Science，1997，278(546-9)：446-449

[4] Sakamoto K，Gouzu H，Komiya K，et al. Molecular computation by DNA hairpin formation[J]. Science，2000，288(5469)：1223-1226

[5] Zhou Kang，Gao Zun-hai，Xu Jin. An algorithm of DNA computing on 0-1 planting problem[J]. Advances in Systems Science and Applications，2002，5(4)：587-593

[6] 王子成，周康，罗亮，等. DNA 计算中编码序列的过滤函数研究[J]. 计算机工程与应用，2008，44(32)：10-11

[7] 周康，殷燕芳，李玉华，等. DNA 编码的模型分析[J]. 华中科技大学学报：自然科学版，2007，35(7)：67-70

[8] Lee J Y，Shin S Y，Park T H，et al. Solving traveling salesman problems with DNA molecules encoding numerical values[J]. BioSystem，2004，78(1/3)：39-47

[9] 周康，同小军，许进. 背包问题的闭环 DNA 算法[J]. 系统仿真学报，2008，20(17)：4605-4608

[10] Lipton R J. DNA solution of hard computational problems[J]. Science，1995，268(5210)：542-545

[11] Dirk F，Cukras A R，Lopton R J，et al. Molecular compution：RNA solution to chess problem[J]. Biochemistry，2000，97(4)：1385-1389

[12] 周康，同小军，许进. 基于粘贴 DNA 芯片模型的八皇后问题算法[J]. 系统工程学报，2008，23(3)：372-376

[13] 李艳梅，余文，宁建国. 一种广义分子计算模型及其在 NP 问题中的应用[J]. 计算机应用研究，2014

[14] 霍红卫，庄心谷. 超立方体上 0-1 背包问题的并行算法[J]. 西安电子科技大学学报，1995，22(3)：249-255

[15] 霍红卫. 序列比较问题的分治法[J]. 西安电子科技大学学报，1998，25(3)：345-347