

Petri 网与 DNA 计算^{*}

Petri Net and DNA Computation

陆维明

(中国科学院数学研究所 北京 100080)

摘要 本文是文[1][2]的姐妹篇,以计算机科学近年所发现的新大陆——脱氧核糖核酸(DNA)参加生化反应作为计算(DNA 计算)背景,进一步阐明 Petri 网的丰富内涵。

关键词 Petri 网,正确性,活性,安全性,DNA 计算,可靠性

世界上的事情总有不少巧合:Petri 网的研究、发展与中国计算机学会的成立、成长、壮大同步^[1];国内、外对并行计算感兴趣的人与图灵奖获得者 Robin Milner 一样对 Petri 网有相识恨晚的感受^[2];著名物理学家 T. Toffoli 在 1982 年对计算系统功能作出一种预言,但巧的是科学家 C. A. Petri 在早此 20 年就已断言计算系统只要充分运用并行性则前途无量^[3];...就在文[2]发表后不久,《科学》杂志上登载了生物化学家 L. M. Adleman 利用具有高度并行性的脱氧核糖核酸(DNA)生化反应,求解组合问题的论文^[4]。这再一次证实 C. A. Petri 的断言不但具有现实可能性,而且在分子计算系统中已经发挥作用。无独有偶,关于 DNA 计算具有高度并行性可充分利用的看法,在数学家 B. Cipra 的论文[4]中有了更详尽的说明。物理学家,生物化学家,数学家,计算机科学家,...都在关心并行计算。本文作为文[1],[2]的姐妹篇,试图进一步阐明 Petri 网对并行计算及其系统的指导作用。

一、基于 Petri 网的系统设计

C. A. Petri 于 1962 年首次在其著名博士学位论文中提出网(后称为 Petri 网)的概念之后,全世界关于研究 Petri 网应用与理论的论文、书籍不断涌现,各种研究小组的活动十分活跃,学术交流与会议频繁且遍及世界各个角落。在这样的研究热潮中,人们十分关心在设计一个物理系统,特别是并行的物理系统时 Petri 网,甚或 Petri 网的内涵可以起什么作用。诚如 Petri 早在 1962 年所说:“...采用组织

手段,你可以克服物理限制!...。并发性是系统组织中的基础、本质和必然。...”

这里,我们不妨将文[2]中用 Petri 网设计好的一个无限长堆栈复制在图 1,作一些探讨。图 1 的堆栈已假定其中有了无限多个元素,这只是为了讨论方便。随后,我们要用高级 Petri 网设计成同样堆栈,但初始状态为空栈。

很明显,除了堆栈的头部与其他段稍有不同外,其他段内涵是完全一样的,运行原理也是一样的。

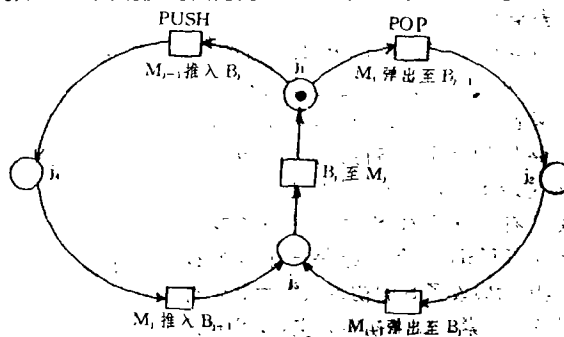


图 1 无限长堆栈 S 第 j 段,其中 M 为存储器, B 为缓冲器

对于 S 的操作:推入(PUSH)用到的是 S 的左部,弹出(POP)用到的是 S 的右部。

对于 S 的操作,PUSH 或 POP,均只要花费一个段的一个周期时间即可,亦即 S 并未因其无限长而影响其时间性能,这一点对于用 Petri 网来设计物理系统时是十分重要而吸引人的。我们深入分析可知, S 的用户在将一数据推入 S 或从 S 弹出一数据之后可能会满意地去做别的事,而 S 的某些段可能会继

^{*}国家自然科学基金资助;中国科学院管理、决策信息系统开放实验室(MADIS)支持。陆维明 研究员,博士导师,研究兴趣为 Petri 网,软件工程,算法设计与分析。

续运行那么一阵子,但这与S的性能无关。再深入分析获知,恰恰就是段间具有并发、独立的运行能力,才能使S具有好的性能。为了不引入更多的术语,对于S我们只是说它具有极高的平行性、并发性。

从S还可以看到模块化的实现非常易于分析系统的正确性、活性与安全性。

1. 正确性

只要观察POP即可,因PUSH是其完全对称部分。在用户要求从S中弹出一个元素后,系统的右边部分逐段往上弹出一个元素(经由缓冲器B),这样就保证了弹出的是栈顶元素,而任一j段的存储器中经过j段右边的Petri网部分的运行之后得到了j+1段的存储器元素,这就保证了下一次的运行的正确性;亦即,用户得到栈S的栈顶元素,并且栈S准备好了新的状态,使用户下一次操作依然有正确的保证。

2. 活性

也只要观察POP即可。对于栈S的活性,只要说明S的第j段的POP部分的三个变迁是活的即可:它们是变迁 M_j 弹出至 B_{j-1} , M_{j+1} 弹出至 B_j , B_j 至 M_j 。显然,这个简单的Petri网构成的回路在POP执行中总是每个变迁顺序执行一次而后回路回到初态,没有例外。因此, S_j 是活的(S中每个变迁均活)。

3. 安全性

只观察POP。第j段的POP部分的三个位置中只有 j_2 有两条输入弧,可是 j_2 中的标志在POP与PUSH两者择一而用之, j_2 不可能同时从POP回路与PUSH回路得到标志。至于 j_1 与 j_2 ,显见最多只会包含一个标志。

我们看到这个栈S对于用户的操作POP或者PUSH均会引起每一个段的操作,这是因为假设了S的初始状态为其中有了无限多个元素。如果初态为空,或S中只有k个元素,则用户的操作POP或PUSH不应引起每一个段都有操作,这里用高级Petri网来设计,弧上加上了谓词(对此不熟悉或不关心的读者可以跳过不读),示于图2。

同样,讨论它们的并发性、并行性、正确性、活性、安全性等均不难,不再占用篇幅。

二、DNA 计算

在我们的文[2]发表后仅三个月,科学家Adleman发表了关于DNA计算的论文[3],他说:“我们用脱氧核糖核酸分子(DNA)为计算工具,求解了有向哈密顿路问题(一个著名的NP完全问题)的一个实例。…”。

• 2 •

接着,Barry Cipra以其敏锐的科学眼光发表论文[4],高度重视DNA计算。

可以这么说,DNA计算的核心,正是Petri网的特性,即高度并发性。

为了大致了解DNA计算,这里将Adleman的实例与DNA计算算法介绍一下。图3就是他用的

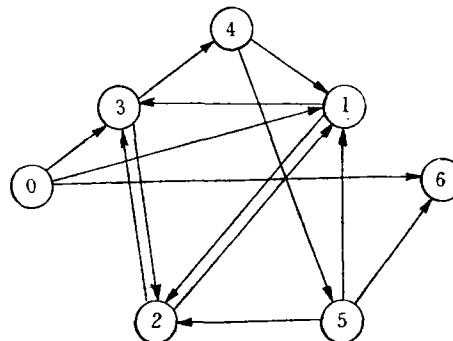


图3 DNA计算的实例:寻找有向图

中结点0→6哈密顿路

DNA计算哈密顿路的有向图;虽然此有向图只有七个结点,连接的有向弧也较简单,我们用目测就可找到结点0到结点6的路,但是使用Adleman的实验算法,利用生物化学反应找到同样的路,其意义是截然不同的。

作为一种计算,总要说明其数据结构及计算的算法;为了易于阅读,只引入必要术语。

1. DNA 数据结构

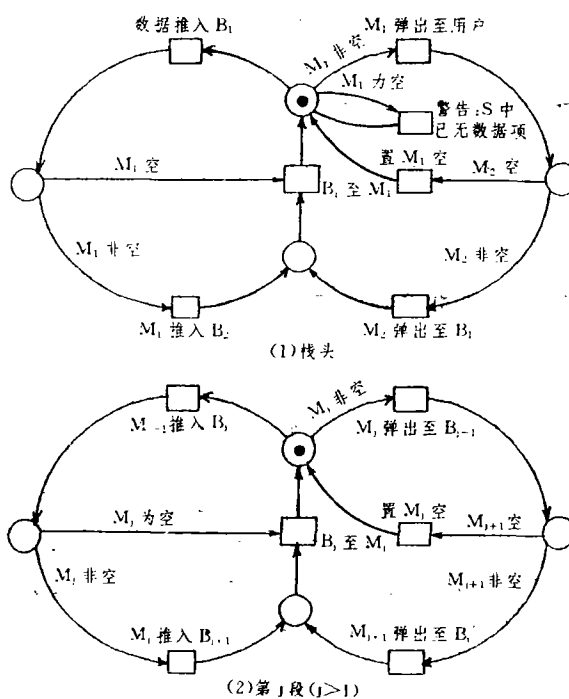


图2 初态为空的无限长堆栈S

用 DNA 单链短片段(单聚核苷酸)表示结点与结点间的有向弧;在这个片段上有四种不同的分子结构,分别记为 A, C, G, T, 其中 A 与 T 互补, C 与

G 互补。例如,下列三个片段分别表示结点 2, 有向弧 1→2 和 2→3。

很明显,这长度为 20 个分子结构的短片段已足以表

C	G	A	T	A	A	G	C	T	C	G	A	A	T	T	T	C	G	A	T	2
...										G	C	T	A	T	T	C	G	A	G	1→2
C	T	T	A	A	A	G	C	T	A	...										2→3

示图 3 所示的有向图;又,易于看出,结点 2 的前 10 个分子结构与有向弧 1→2 的后 10 个分子结构互补,而结点 2 的后 10 个分子结构与有向弧 2→3 的前 10 个分子结构互补, ..., 其他结点与有向弧的 DNA 表示类似这 3 个 DNA。

2. DNA 算法

- S1 DNA₀, ..., DNA₀, DNA₀₋₁, DNA₀₋₃, ..., DNA₅₋₁, DNA₅₋₂, DNA₅₋₆ 各取若干^{*}。
- S2 倒入同一试管;结点与有向弧的互补部分迅速结合(A 与 T 匹配, C 与 G 匹配)^{**}, 生成 DNA 双链长片段, 每个长片段表示图中一条路。
- S3 利用 PCR 反应(聚合酶链式反应)放大①始, 或⑥终的长片段。
- S4 用琼脂糖作用于算法步 S3 的结果, 并只保留长度为 140 的长片段(对应七个结点)。
- S5 将 DNA 双链 140 长片段分解为 DNA 单链 20 短片段, 并用 DNA₀, ..., DNA₆ 核实它们。
- S6 放大若干次, 以明显确定某个 DNA 双链 140 长片段是否是哈密顿路。

有两件事情使我们兴奋不已:一是建立在数学、物理基础上的计算已经拓广到生物化学的基础之上;二是 DNA 算法的超并行度(上过算法的 S2)使得象哈密顿路这种典型的 NP 完全问题有了一种实用的解决方法(不是说已解决)。

怀着同样的心情,一些科学家在探索用 DNA 计算可以解决一些什么样的问题以及研究用 DNA 计算解决除哈密顿路以外的实际问题。例如, Lipton 指出^[4], 相当多可满足问题(属 NP 完全问题)可用 DNA 计算解决, 其步数正比于逻辑表达式的长度。...执行的只是生物化学中常见的“合成”, “分离”, “放大”, “检测”。..., Boneh 与 Dunworth 报告^[4], 美国国家安全署发明了数据加密标准系统(DES), 供政府部门使用, 用以防御硅计算机已足够, 因它有约 10¹⁷种密钥, 但用 DNA 计算, 不超过 1000 步即可攻破此种保密系统。...

*)DNA 相当廉价, Adleman 做计算时约 \$ 30 可买 10¹⁷个 DNA。 **)这是超并行的计算步骤, 由生化反应实现此种超并行。

我们觉得, 一些问题, 象密码学问题, 有难以置信的并行方法来对付它们;对于有利可图的问题, DNA 计算或进而 DNA 计算机就会有经济上的推动力。当然, Adleman 制造^[4]的超并行 DNA 计算机 TT-100, 在技术上是重大进步。

三、结论

1. 从无限长堆栈 S 的设计, 我们解除了物理学家对计算系统性能的担心。系统增大, 只要充分发挥并行性, 性能不应是问题。

2. 从 DNA 计算解决哈密顿路实例, 我们看到虽然计算的基础从物理、数学转向生物化学, 但其计算系统的并行性仍是高性能的保障。

3. 从表 1 所列的内容, 我们可以认为: Petri 与 Petri 网在实践上预示 DNA 计算(机)可以成功, 也预言在将来基于原子, 基于...的计算可能有超过 10²⁰(分子计算)并行度。

表 1 并行度与系统

并行度	系 统
10 ¹	一些早期的并行多机系统, 比如, 银河, 曙光 早期系统。
10 ²	
10 ³	
10 ⁴	
...	目前的一些实验系统。
10 ²⁰	
...	当前 DNA 计算可达的并行度
∞	
	Petri 网系统

参考文献

- [1] 陆维明, Petri 网研究在中国, 计算机科学, Vol. 19, No. 3, 1992
- [2] 陆维明, 林闯, Petri 网研究: 机遇与挑战, 计算机科学, Vol. 21, No. 4, 1994
- [3] Leonard M. Adleman, Molecular Computation of Solutions to Combinatorial Problems, Science, Vol. 266, No. 11, 1994
- [4] Barry Cipra, Computer Science Discovers DNA, in WHAT'S HAPPENING IN THE MATHEMATICAL SCIENCE, Vol. 1, 1996