

# 计算生物学中有关基因组翻转距离的 NPC 问题<sup>\*</sup>)

NPC Problems of the Reversal Distance between Genomes in Computational Biology

栾峻峰 朱大铭 马绍汉

(山东大学计算机科学与技术学院 济南250100)

**Abstract** Problems of computing the reversal distance between genomes are discussed. Problems of computing the reversal distance between genomes are fundamental problems of Computational Biology, these problems have important meanings in studying the biological race evolution and the bio-pharmaceuticals etc. The problem of evolutionary trees based on reversal distance between genomes and it's NPC property are especially discussed.

**Keywords** Algorithm, Evolutionary tree, Genome

## 1 基因组翻转距离问题的提出

20世纪80年代末, Jeffrey Palmer 及其同事在对比甘蓝与芜菁甘蓝的基因序列时发现, 排列形成两种基因序列的分子几乎完全相同, 只是这些分子在两种基因中的排列顺序不一样。这一发现以及其后的研究表明, 基因重组是生物演化过程中的一个基本特征。基因重组研究中多次遇到要用尽可能少的次数将一条基因序列中的片断翻转连接成新的基因的问题, 这就是后来由 D. Sankoff 等提出的基因组翻转距离问题<sup>[1]</sup>, 或抽象为有向符号序列的翻转距离问题, D. Sankoff 等猜测该问题是 NP 难解问题, 然而他们没能给出证明。

**有向符号序列的翻转距离问题** 设  $S = \{1, 2, \dots, n\}$  为由  $n$  个符号组成的集合,  $S$  的反方向集合为  $S^- = \{-1, -2, \dots, -n\}$ 。考虑由  $S$  的所有有向符号组成的一个排列  $\pi = \pi_1 \pi_2 \dots \pi_n, \pi_i \in S \cup S^-$ , 如果  $|\pi| = |\pi_1| |\pi_2| \dots |\pi_n|$  恰为符号集  $S$  的一个置换, 则称  $\pi$  为长度为  $n$  的有向符号序列, 有向符号序列  $\pi$  的一次翻转  $\rho(i, j)$  使得  $\pi$  变到一个新的有向符号序列  $\pi \cdot \rho(i, j) = \pi_1 \pi_2 \dots \pi_{i-1} - \pi_i - \pi_{j-1} \dots - \pi_j \pi_{j+1} \dots \pi_n$ 。有向符号序列的翻转距离问题即给定符号集合  $S$  的两个有向符号序列  $\pi^1$  和  $\pi^2$ , 求常数  $k$ , 使得存在一组翻转  $\rho_1, \rho_2, \dots, \rho_k$ , 完成  $\pi^1 \cdot \rho_1 \cdot \rho_2 \dots \rho_k = \pi^2$ , 并且  $k$  最小。称  $k$  为序列  $\pi^1$  和  $\pi^2$  的翻转距离, 记为  $d(\pi^1, \pi^2) = k$ 。

有向符号序列的翻转距离问题是生物计算中的一个基本问题, 其极强的理论及实用价值立即吸引了许多专家的关注<sup>[2-4]</sup>, 直到1995年, 才由 S. Hannenhalli 与 P. A. Pevzner 给出多项式时间算法解决了任意两个有向符号序列的翻转距离计算问题<sup>[5]</sup>。

## 2 基因组翻转距离问题的计算

计算任意两个有向符号序列  $\pi^1$  和  $\pi^2$  的翻转距离时, 可以假设符号序列  $\pi^1$  和  $\pi^2$  的第一个符号与最后一个符号相同, 即  $\pi_1^1 = \pi_1^2 = 1, \pi_n^1 = \pi_n^2 = n$ , 否则只要在  $\pi^1$  和  $\pi^2$  的前面和后面各加一个符号即可。这样由  $\pi^1$  和  $\pi^2$  构造 RD 图  $G_R(V(S), E(\pi^1), E(\pi^2)), V(S) = \{v_{ik}, v_{jk} | 1 \leq k \leq n\}$ , 将  $V(S)$  中的点排列为  $v_{11}, v_{12}, \dots, v_{n1}, v_{n2}$ , 且若  $\pi_i^1$  方向为正, 则  $v_{i1} = v_{ik}, v_{i2} = v_{ik}$ , 若  $\pi_i^1$  方向为负, 则  $v_{i1} = v_{jk}, v_{i2} = v_{jk}$ ; 然后定义黑边集为  $E(\pi^1) = \{(v_{i2}, v_{i+1,1}) | 1 \leq k \leq n-1\}$ , 灰边集为  $E(\pi^2) = \{e(\pi_i^2, \pi_{i+1}^2) | 1 \leq$

$k \leq n-1\}$ , 其中  $e(\pi_i^2, \pi_{i+1}^2)$  取法为: 设  $|\pi_i^2| = i, |\pi_{i+1}^2| = j$ , 则  $e(\pi_i^2, \pi_{i+1}^2) =$

$$\begin{cases} (v_{i1}, v_{jk}), \pi_i^2 = i, \pi_{i+1}^2 = j, \\ (v_{i1}, v_{jk}), \pi_i^2 = i, \pi_{i+1}^2 = -j, \\ (v_{i2}, v_{jk}), \pi_i^2 = -i, \pi_{i+1}^2 = j, \\ (v_{i2}, v_{jk}), \pi_i^2 = -i, \pi_{i+1}^2 = -j. \end{cases}$$

RD 图总是由互不连通的圈组成, 将所有圈记为  $C(G_R) = \{c_1, c_2, \dots, c_k\}$ , 每个圈总是由黑边和灰边间隔相连。一条边的端点按照在  $V(S)$  中排列的顺序分为左端和右端。若一条灰边连接两条黑边的同一端, 则称其为有序边, 否则称其为无序边。至少含有一条有序边的圈称为有序圈, 否则称为无序圈。若灰边  $e_1$  的一端位于灰边  $e_2$  两端的中间,  $e_1$  的另一端位于  $e_2$  两端的外面, 则称  $e_1$  与  $e_2$  相交; 若相交的两条灰边分别属于两个圈  $c_1$  与  $c_2$ , 则称  $c_1$  与  $c_2$  相交。设  $C$  是  $C(G_R)$  的子集, 满足

- (1) 对任意  $c_i \in C$ , 总有  $c_j \in C$  与  $c_i$  相交;
- (2) 不存在  $c_k \in C(G_R) - C$  与  $C$  中的圈相交;
- (3) 不存在  $C$  的子集满足 (1)、(2),

则称  $C$  中的圈形成 RD 图的一个 C 图。若一个 C 图  $G_1$  的所有顶点均位于另一个 C 图  $G_2$  的两个顶点  $v_1$  与  $v_2$  之间, 则称  $G_2$  覆盖  $G_1$ ; 设有  $k \geq 2$  个 C 图  $G_1, G_2, \dots, G_k$  被 C 图  $G_C$  覆盖, 若对任意  $v \in V(G_{RC}), G_1, G_2, \dots, G_k$  的所有顶点均位于  $v$  的左端或右端, 则称  $G_C$  严格覆盖  $G_i, 1 \leq i \leq k$ 。若  $G_1$  的所有顶点均位于  $G_2$  任一顶点的左边或右边, 则称  $G_1$  与  $G_2$  并列。若一个 C 图中的所有圈均为无序的, 则称该 C 图为无序的。若一个无序 C 图不覆盖其他任何无序 C 图, 或者在整个 RD 图中严格覆盖其他所有无序 C 图, 则称其为 H 图。若 H 图  $G_H$  被无序 C 图  $G_C$  严格覆盖,  $G_C$  不是 H 图, 也不覆盖其他 H 图, 则称  $G_H$  为强 H 图。若 RD 图中含有奇数个强 H 图, 不含有其他 H 图, 则称 RD 图中含有一个 F 图, 显然一个 RD 图中最多只能有一个 F 图。

设 RD 图中黑边数及 C 图、H 图、F 图个数分别为  $b, c, h, f$ , 则  $d(\pi^1, \pi^2) = b - c + h + f$ <sup>[5]</sup>。在考虑  $\pi^1$  到  $\pi^2$  翻转步骤时, 从 RD 图中寻找有序圈 (如果 RD 图中没有有序圈, 可以先经一次翻转形成一个有序圈), 将有序圈对应的序列片段翻转, 可以保证翻转后的 RD 图中仍有有序圈, 从而可以继续翻转直到得到  $\pi^2$ 。

<sup>\*</sup>) 本文受国家自然科学基金支持(No. 60073042)。栾峻峰 博士生, 主要研究方向: 算法及其复杂性理论。朱大铭 教授, 博士生导师。马绍汉 教授, 博士生导师。

### 3 基于基因组翻转距离的进化树问题

生物领域中的进化概念在这里被抽象为一条有向符号序列经过一次次的翻转最终成为另一条有向符号序列的过程,生物领域中的进化树被抽象为加权树,其中叶点为对应已知生物基因的有向符号序列,分支点为猜测生物基因对应的有向符号序列,边权为顶点间的翻转距离,或称为进化距离,进化树问题的基本思想就是在给定叶点对应的序列后,计算分支点对应的有向符号序列,以使得进化树上总进化距离最小。进化树问题对于研究生物进化历史与进化规律,对于研究生物起源等问题都有重要意义。星树问题是进化树问题中最简单的一类问题,星树问题是指找一条有向符号序列使之与给定数条有向符号序列的翻转距离之和最小。星树问题在生物制药工程中有重要应用价值,在生物制药工程中,我们希望寻找能够尽可能与病毒基因接近,同时又与正常细胞基因尽可能不同的生物药品,也就是让病毒表现出尽可能多的亲药性,这对更快地发挥药效有直接的促进作用。我们把星树问题简记为 ST 问题,描述为下面的形式:

**翻转距离星树问题** 给定  $m$  条长度为  $n$  的有向符号序列  $\pi^1, \pi^2, \dots, \pi^m$ , 正整数  $M$ , 问是否存在一条有向符号序列  $\pi^*$ , 使得  $\sum_{i=1}^m d(\pi^i, \pi^*) \leq M$ 。

ST 问题的 NP 完全性已经得到了证明<sup>[4]</sup>。如果在星树问题中给定的有向符号序列数不是变量  $m$ , 而是常数 2, 则对应的问题简记为 2ST 问题。2ST 问题是很容易解决的, 这只要应用由 S. Hannenhalli 与 P. A. Pevzner 给出的多项式时间算法计算已知两条序列的翻转距离, 再与  $M$  比较即可, 由一条序列到另一条序列翻转过程中得到的任一序列, 均可作为  $\pi^*$ 。但是, 当已知序列条数为常数 3 时, 对应的 3ST 问题是否是 P 问题就不得而知了。对于 3ST 或者 4ST、5ST 等问题的 NP 性的研究仍在进行中, 我们倾向于 3ST 问题以及 4ST、5ST 等问题均为 NP 完全问题。在对 3ST 问题中的目标序列  $\pi^*$  作了一些限制后, 我们证明了新问题的 NP 完全性。这些限制及新问题描述如下:

**F3ST 问题** 给定 3 条长度为  $n$  的有向符号序列  $\pi^1, \pi^2,$

$$X_E = \begin{cases} \phi & C_{3SAT}(x_i) = \phi, \text{ or } C_{3SAT}(\bar{x}_i) = \phi, \text{ or } |C_{3SAT}(x_i)| = |C_{3SAT}(\bar{x}_i)|, \\ \{y_{i1}, y_{i2}, \dots, y_{i,2i}\}, & k = ||C_{3SAT}(x_i)| - |C_{3SAT}(\bar{x}_i)||, \end{cases}$$

$$C_E = \begin{cases} \phi, & C_{3SAT}(x_i) = \phi, \text{ or } C_{3SAT}(\bar{x}_i) = |C_{SAT}(\bar{x}_i)|, \\ \{\bar{x}_i \vee y_{i1} \vee y_{i2}, \bar{x}_i \vee y_{i3} \vee y_{i4}, \dots, \bar{x}_i \vee y_{i,2i-1} \vee y_{i,2i}\}, & k = |C_{3SAT}(x_i)| - |C_{3SAT}(\bar{x}_i)| > 0, \\ \{x_i \vee y_{i1} \vee y_{i2}, x_i \vee y_{i3} \vee y_{i4}, \dots, x_i \vee y_{i,2i-1} \vee y_{i,2i}\}, & k = |C_{3SAT}(\bar{x}_i)| - |C_{3SAT}(x_i)| > 0, \end{cases}$$

$$X_N = \begin{cases} \phi, & |C_{3SAT} \cup C_S| \text{ 为偶数}, \\ \{x_{N1}, x_{N2}, x_{N3}\}, & |C_{3SAT} \cup C_S| \text{ 为奇数}, \end{cases} E_N = \begin{cases} \phi, & |C_{3SAT} \cup C_S| \text{ 为偶数}, \\ \{x_{N1} \vee x_{N2} \vee x_{N3}\}, & |C_{3SAT} \cup C_S| \text{ 为奇数}. \end{cases}$$

当 3SAT 问题可满足时, 取  $X_E$  和  $X_N$  中的布尔变量真值为 1, 则 E3SAT 问题可满足; E3SAT 问题满足时, 3SAT 问题显然也是满足的。

#### 4.2 F3ST 问题的 NP 完全性

将 3SAT 问题归约到 F3ST 问题, 从而证明 F3ST 问题是 NP-C 问题。

在构造 F3ST 问题实例时, 先假设要构造的 F3ST 问题有最优解  $\pi^*, \pi^*$  中各符号的顺序是确定的, 大部分符号的方向也是已知的, 只有部分符号的方向待定, 然后在  $\pi^*$  的基础上根据 3SAT 问题中的项构造 RD 图, 使得所构造的 RD 图中只有 C 图, 没有 H 图和 F 图, 而且 C 图个数就是圈的个数, 并说明构造方法能够保证 3SAT 问题中的项可满足与 C 图中的圈有序是等价的, 最后根据 RD 图写出要构造的 F3ST 问题实例中的有向符号序列, 证明其最优解就是前面假设的最优解  $\pi^*$ 。下面简单说明构造过程中的一个主要环节。

$\pi^3$ , 正整数  $M$ , 以及  $S = \{1, 2, \dots, n\}$  的一个真子集  $L$ , 问是否存在一条有向符号序列  $\pi^* = \pi_1 \pi_2 \dots \pi_n$ , 使得  $\sum_{i=1}^3 d(\pi^i, \pi^*) \leq M$ , 而且  $\pi_i = i, i \in L, |\pi_i| = i, i \in S - L$ 。

### 4 星树问题的 NP 完全性

朱大铭等在 3SAT 问题基础上提出了一个 E3SAT 问题, 将 3SAT 问题多项式时间归约到 E3SAT 问题, 又将 E3SAT 问题归约到星树问题, 从而证明了 ST 问题的 NP 完全性。另外 3SAT 问题也可将多项式时间归约到 F3ST 问题, 这样 F3ST 问题的 NP 完全性也可得到证明。

#### 4.1 ST 问题的 NP 完全性

E3SAT 问题向 ST 问题的归约与后面的 3SAT 问题向 F3ST 问题的归约有极大的类似性, 事实上 3SAT 问题向 F3ST 问题的归约是在充分借鉴 E3SAT 问题向 ST 问题的归约思想后做出的, 这里省去 E3SAT 问题向 ST 问题的归约, 主要介绍 3SAT 问题向 E3SAT 问题的归约方法。3SAT 问题与 E3SAT 问题描述为:

**3SAT 问题** 给定布尔变量集合  $X = \{x_1, x_2, \dots, x_n\}$  和由布尔变量形成的项集合  $C = \{c_1, c_2, \dots, c_m\}$ , 其中  $c_j = c(x_{j1}) \vee c(x_{j2}) \vee c(x_{j3}), 1 \leq j \leq m, 1 \leq j_1 < j_2 < j_3 \leq n, c(x_i) \in \{x_i, \bar{x}_i\}, 1 \leq i \leq n$ 。问能否找到对各布尔变量一个真值指派  $a(x_i), 1 \leq i \leq m$ , 使项集合中各项取值均为真。

E3SAT 在 3SAT 问题中, 称  $x_i$  和  $\bar{x}_i$  为布尔变量  $x_i$  的正变量和负变量, 并记  $C(x_i) = \{c_k | x_i \in c_k, 1 \leq k \leq m\}, C(\bar{x}_i) = \{c_k | \bar{x}_i \in c_k, 1 \leq k \leq m\}$ , 如果

- (1)  $m$  为偶数;
- (2) 对任一  $x_i$ , 由  $C(x_i) \neq \phi, C(\bar{x}_i) \neq \phi$  可知  $|C(x_i)| = |C(\bar{x}_i)|$ ,

则称相应的 3SAT 问题为 E3SAT 问题。

**3SAT 问题向 E3SAT 问题的归约** 设 3SAT 问题的实例为  $X_{3SAT}$  和  $C_{3SAT}$ , 如下构造 E3SAT 问题:

**例**  $X_{E3SAT} = X_{3SAT} \cup X_E \cup X_N, C_{E3SAT} = C_{3SAT} \cup C_E \cup C_N$ , 其中

$$X_E = Y_{i-1}^* X_E, C_E = Y_{i-1}^* C_E,$$

设布尔变量集合  $X = \{x_1, x_2, x_3, x_4\}$ , 项集合  $C = \{c\}, c = x_1 \vee \bar{x}_3 \vee x_4$ , 假设  $\pi^*$  中前 21 个符号为  $\pi_0^* \pi_1^* \dots \pi_{20}^*$ , 每个布尔变量对应 5 个符号,  $x_i$  对应  $\pi_{5(i-1)+j}^*, 0 \leq j \leq 4$ , 其中  $\pi_{5(i-1)+j}^* = 5(i-1) + j, j = 0, 2, 3, 4, |\pi_{5(i-1)+1}^*| = 5(i-1) + 1, \pi_{5(i-1)+1}^*$  的方向由  $c = x_1 \vee \bar{x}_3 \vee x_4$  满足时  $x_i$  的真值确定,  $\pi_{5(i-1)+1}^*$  取正向当且仅当  $a(x_i) = 0, \pi_{20}^* = 20$ 。构造  $\pi^*$  时, 设计 RD 图中的一个片段(图 1), 其中每个符号用相邻两个黑点表示, 由于  $\pi_{5(i-1)+1}^*$  的方向未定, 因此  $\pi_{5(i-1)+1}^*$  对应的两个黑点暂时重合在一起。该片段中有  $c = x_1 \vee \bar{x}_3 \vee x_4$  对应的一个含七条黑边的圈, 一个  $x_2$  对应的含 5 条黑边的有序圈, 两个分别由  $x_1, x_3$  对应的含 3 条黑边的有序圈, 一个由  $x_4$  对应的含两条黑边的有序圈。由  $c = x_1 \vee \bar{x}_3 \vee x_4$  的表现形式, 构造  $\pi^1$  时取  $\pi_1^1 = 1, \pi_4^1$  任取 6 或 -6,  $\pi_{11}^1 = -11, \pi_{16}^1 = 16$ 。由图中有序圈形式取  $\pi_{5(i-1)+3}^1 = -5(i-1) - 3, 1 \leq i \leq 4$ , 其余点均取正向, 即  $\pi_{5(i-1)+j}^1 = 5(i-1) + j, 1 \leq i \leq$

4,  $j=0, 2, 4$ 。从 RD 图片段中可看出,  $c = x_1 \vee \bar{x}_3 \vee x_4$  满足与其对应的含 7 条黑边的圈有序是等价的。事实上, 若  $c = x_1 \vee \bar{x}_3 \vee x_4$  满足, 不妨设  $\bar{x}_3$  真值为 1, 即  $a(x_3) = 0$ , 则  $\pi_{11}^+ = 11$ , 再由  $\pi_{11}^- = -11$ , 可知图 1 中  $\pi_{11}^+$  对应处就存在有序灰边(图 2), 于是知

图中含 7 条黑边的圈是有序的; 反之, 若图中含 7 条黑边的圈有序, 不妨设  $\pi_i^+$  处对应灰边有序(图 3), 则由  $\pi_i^+ = 1$  知  $\pi_i^- = -1$ , 故  $a(x_1) = 1$ , 所以  $c$  满足。

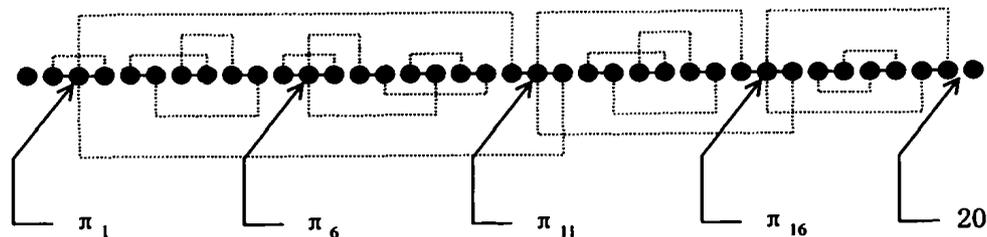


图1 RD 图片段

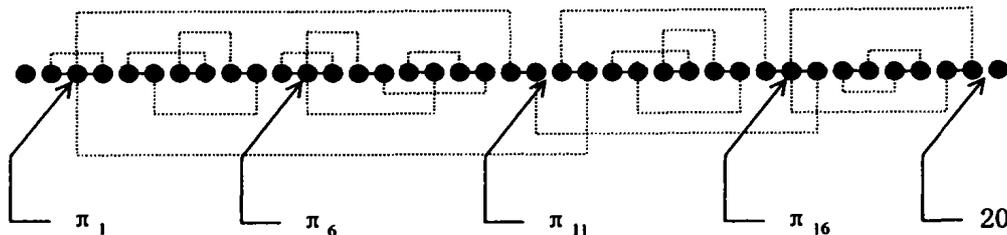


图2 RD 图片段

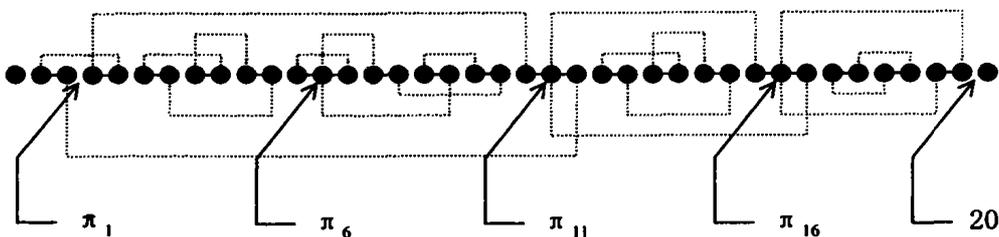


图3 RD 图片段

下面由 3SAT 问题实例构造 F3ST 问题实例。取  $M = 22mn - 2m$ , 如下构造三条长度为  $10mn + 1$  的有向符号序列  $\pi^1, \pi^2, \pi^3$  和集合  $L$ :

$$\pi^1 = \pi^{11} \pi^{12} \pi^{13} \pi^{14} \dots \pi^{1, 2m-1} \pi^{1, 2m} \pi_{10mn}^+, \pi_{10mn}^- = 10mn$$

其中序列片段  $\pi^{1k}$  的长度为  $5n$ , 符号顺序如下确定:

- 1)  $k$  为奇数时, 设  $k = 2j - 1, j = 1, \dots, m$ , 则  $\pi^{1k}$  由  $c_j = c(x_{j1}) \vee c(x_{j2}) \vee c(x_{j3})$  确定(图 4):

1. 序列  $\pi^1$  的构造

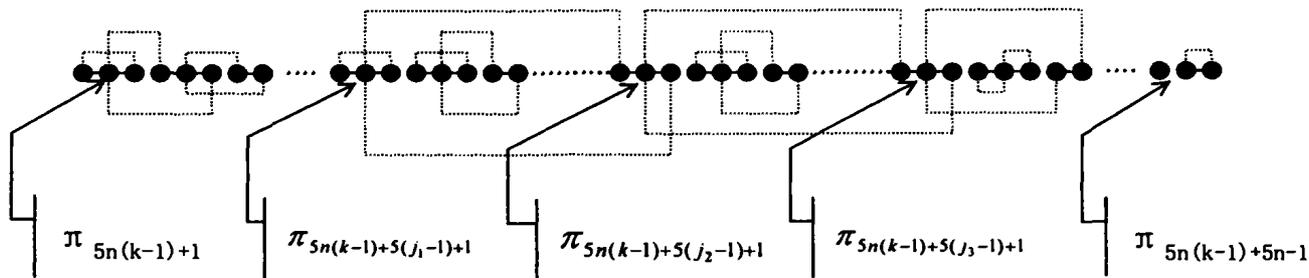


图4 欲构造的 RD 图  $G_R(V(S), E(\pi^1), E(\pi^2))$  中  $\pi^{1k}$  段( $k$  为奇数)

$$\begin{aligned} \pi^{1k} &= \pi^{1k0} \pi^{1k1} \pi^{1k2} \pi^{1k3} \pi^{1k4} \pi^{1k5} \pi^{1k6}, \\ \pi^{1k0} &= \pi^{1k} \cdot 1 \pi^{1k} \cdot 2 \dots \pi^{1k} \cdot j_1 - 1, \\ \pi^{1k1} &= \pi_{5n(k-1)+5(j_1-1)} \pi_{5n(k-1)+5(j_1-1)+2} \pi_{5n(k-1)+5(j_1-1)+4} \\ &\quad \pi_{5n(k-1)+5(j_1-1)+3}, \\ \pi^{1k2} &= \pi^{1k} \cdot j_1 + 1 \pi^{1k} \cdot j_1 + 2 \dots \pi^{1k} \cdot j_2 - 1, \\ \pi^{1k3} &= \pi_{5n(k-1)+5(j_2-1)} \pi_{5n(k-1)+5(j_2-1)+1} \pi_{5n(k-1)+5(j_2-1)+3} \\ &\quad \pi_{5n(k-1)+5(j_2-1)+2} \pi_{5n(k-1)+5(j_2-1)+4} \pi_{5n(k-1)+5(j_2-1)+3}, \\ \pi^{1k4} &= \pi^{1k} \cdot j_2 + 1 \pi^{1k} \cdot j_2 + 2 \dots \pi^{1k} \cdot j_3 - 1, \end{aligned}$$

$$\begin{aligned} \pi^{1k5} &= \pi_{5n(k-1)+5(j_2-1)+1} \pi_{5n(k-1)+5(j_2-1)+2} \pi_{5n(k-1)+5(j_2-1)+3} \\ &\quad \pi_{5n(k-1)+5(j_2-1)+4} \pi_{5n(k-1)+5(j_2-1)+1}, \\ \pi^{1k6} &= \pi^{1k} \cdot j_3 + 1 \pi^{1k} \cdot j_3 + 2 \dots \pi^{1k} \cdot n, \end{aligned}$$

其中

$$\begin{aligned} \pi^{1k \cdot i} &= \pi_{5n(k-1)+5(i-1)} \pi_{5n(k-1)+5(i-1)+2} \pi_{5n(k-1)+5(i-1)+3} \\ &\quad \pi_{5n(k-1)+5(i-1)+4} \pi_{5n(k-1)+5(i-1)+3} \end{aligned}$$

如下确定序列中各符号的方向:

$$\pi_{5n(k-1)+5j+i} = 5n(k-1) + 5j + i, j = 0, 1, 2, \dots, n-1, i =$$

0, 2, 4,

$$\pi_{5n(k-1)+5j+3} = -5n(k-1) - 5j - 3,$$

$$\pi_{5n(k-1)+5j+1} =$$

$$\begin{cases} -5n(k-1) - 5j - 1, & c(x_j) = \bar{x}_j, j \in \{j_1, j_2, j_3\} \\ 5n(k-1) + 5j + 1, & \text{others} \end{cases}$$

2) k 为偶数时, 设  $k=2j, j=1, \dots, m$ , 则  $\pi^{1k}$  由  $c_j = c(x_{j_1}) \vee c(x_{j_2}) \vee c(x_{j_3})$  确定:

令

$$\begin{cases} j'_1 = n+1-j_3, \\ j'_2 = n+1-j_2, \\ j'_3 = n+1-j_1, \end{cases}$$

将 k 为奇数时确定序列符号顺序时所用表达式中的  $j_1, j_2, j_3$  换为  $j'_1, j'_2, j'_3$ , 即得到  $\pi^{1k}$  的符号顺序, 各符号方向如下:

$$\pi_{5n(k-1)+5j+1} = 5n(k-1) + 5j + i, j=0, 1, 2, \dots, n-1, i=0, 2, 4,$$

$$\pi_{5n(k-1)+5j+3} = -5n(k-1) - 5j - 3,$$

$$\pi_{5n(k-1)+5j+1} =$$

$$\begin{cases} -5n(k-1) - 5j - 1, & c(x_{n+1-j}) = \bar{x}_{n+1-j}, j \in \{j'_1, j'_2, j'_3\} \\ 5n(k-1) + 5j + 1, & \text{others} \end{cases}$$

2. 序列  $\pi^2$  的构造

$$\pi^2 = \pi^{2^1} \pi^{2^2} \pi^{2^3} \pi^{2^4} \dots \pi^{2^{2^m-1}} \pi^{2^{2^m}} \pi_{10mn}, \pi_{10mn} = 10mn,$$

其中序列片段  $\pi^{2^k}$  的确定为(图5):

1)  $k=1, 3, \dots, 2m-1$  时,  $\pi^{2^k}$  长度为  $4n$ ,

$$\pi^{2^k} = \pi^{2^{k-1}} \pi^{2^{k-2}} \dots \pi^{2^{k-n}}$$

$$\pi^{2^k} = \pi_{5n(k-1)+5(i-1)} \pi_{5n(k-1)+5(i-1)+2} \pi_{5n(k-1)+5(i-1)+4}$$

$$\pi_{5n(k-1)+5(i-1)+3},$$

各符号方向为

$$\pi_{5n(k-1)+5(i-1)} = 5n(k-1) + 5(i-1),$$

$$\pi_{5n(k-1)+5(i-1)+2} = 5n(k-1) + 5(i-1) + 2,$$

$$\pi_{5n(k-1)+5(i-1)+4} = 5n(k-1) + 5(i-1) + 4,$$

$$\pi_{5n(k-1)+5(i-1)+3} = -5n(k-1) - 5(i-1) - 3;$$

2)  $k=2, 4, \dots, 2m$  时,  $\pi^{2^k}$  长度为  $6n$ ,

$$\pi^{2^k} = \pi^{2^{k-1}} \pi^{2^{k-2}} \dots \pi^{2^{k-n}}$$

$$\pi^{2^k} = \pi_{5n(k-1)+5(i-1)} \pi_{5n(k-2)+5(n-i)+1} \pi_{5n(k-1)+5(i-1)+2}$$

$$\pi_{5n(k-1)+5(i-1)+3} \pi_{5n(k-1)+5(i-1)+4} \pi_{5n(k-1)+5(i-1)+1},$$

各符号方向为

$$\pi_{5n(k-1)+5(i-1)} = 5n(k-1) + 5(i-1),$$

$$\pi_{5n(k-2)+5(n-i)+1} = 5n(k-2) + 5(n-i) + 1,$$

$$\pi_{5n(k-1)+5(i-1)+2} = 5n(k-1) + 5(i-1) + 2,$$

$$\pi_{5n(k-1)+5(i-1)+3} = -5n(k-1) - 5(i-1) - 3,$$

$$\pi_{5n(k-1)+5(i-1)+4} = 5n(k-1) + 5(i-1) + 4,$$

$$\pi_{5n(k-1)+5(i-1)+1} = -5n(k-1) - 5(i-1) - 1$$

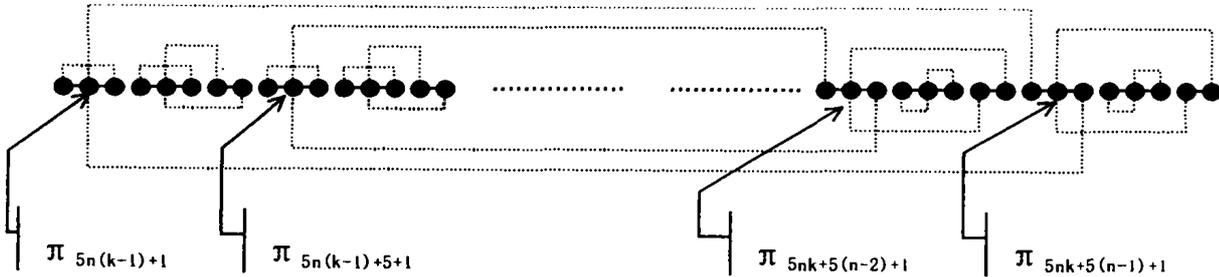


图5 欲构造的 RD 图  $G_R(V(S), E(\pi^2), E(\pi^*))$  中  $\pi^{2^k} \pi^{2^{k+1}}$  段 ( $k$  为奇数)

3. 序列  $\pi^3$  的构造

$$\pi^3 = \pi^{3^1} \pi^{3^2} \pi^{3^3} \pi^{3^4} \dots \pi^{3^{2^m-1}} \pi^{3^{2^m}} \pi_{10mn}, \pi_{10mn} = 10mn,$$

其中序列片段  $\pi^{3^k}$  的确定为(图6, 图7):

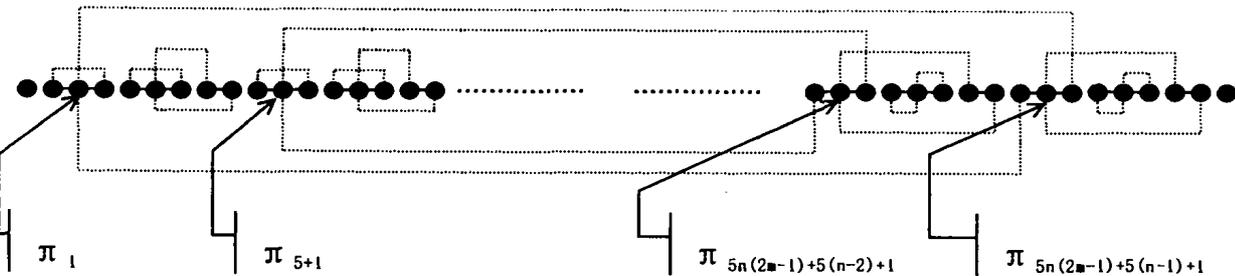


图6 欲构造的 RD 图  $G_R(V(S), E(\pi^2), E(\pi^*))$  中  $\pi^{3^1}$  与  $\pi^{3^{2^m}}$  段

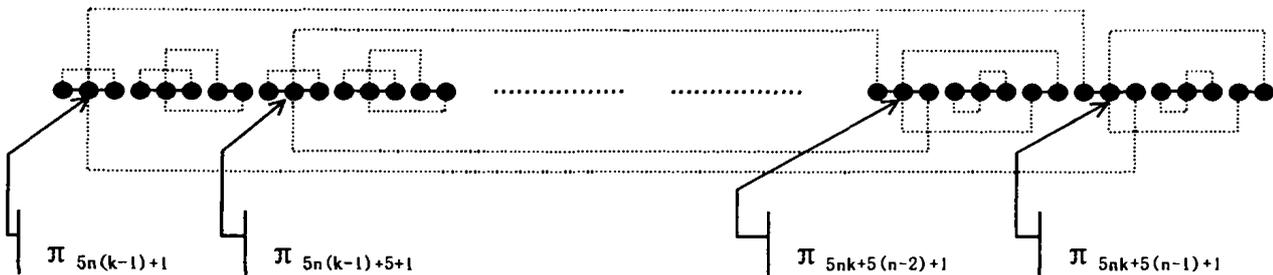


图7 欲构造的 RD 图  $G_R(V(S), E(\pi^3), E(\pi^*))$  中  $\pi^{3^k} \pi^{3^{k+1}}$  段 ( $k$  为偶数)

1)  $k=1, 2, 4, 6, \dots, 2(m-1)$  时,  $\pi^{3k}$  长度为  $4n$ ,

$$\pi^{3k} = \pi^{3k1} \pi^{3k2} \dots \pi^{3kn},$$

$$\pi^{3k1} = \pi_{5n(k-1)+5(i-1)} \pi_{5n(k-1)+5(i-1)+2} \pi_{5n(k-1)+5(i-1)+4} \pi_{5n(k-1)+5(i-1)+3},$$

各符号方向为

$$\pi_{5n(k-1)+5(i-1)} = 5n(k-1) + 5(i-1),$$

$$\pi_{5n(k-1)+5(i-1)+2} = 5n(k-1) + 5(i-1) + 2,$$

$$\pi_{5n(k-1)+5(i-1)+4} = 5n(k-1) + 5(i-1) + 4,$$

$$\pi_{5n(k-1)+5(i-1)+3} = -5n(k-1) - 5(i-1) - 3;$$

2)  $k=3, 5, 7, \dots, 2m-1$  时,  $\pi^{3k}$  长度为  $6n$ ,

$$\pi^{3k} = \pi^{3k1} \pi^{3k2} \dots \pi^{3kn},$$

$$\pi^{3k1} = \pi_{5n(k-1)+5(n-i)} \pi_{5n(k-2)+5(n-i)+1} \pi_{5n(k-1)+5(i-1)+2} \pi_{5n(k-1)+5(i-1)+3} \pi_{5n(k-1)+5(i-1)+4} \pi_{5n(k-1)+5(i-1)+1},$$

各符号方向为

$$\pi_{5n(k-1)+5(i-1)} = 5n(k-1) + 5(i-1),$$

$$\pi_{5n(k-2)+5(n-i)+1} = 5n(k-2) + 5(i-1) + 1,$$

$$\pi_{5n(k-1)+5(i-1)+2} = 5n(k-1) + 5(i-1) + 2,$$

$$\pi_{5n(k-1)+5(i-1)+3} = -5n(k-1) - 5(i-1) - 3,$$

$$\pi_{5n(k-1)+5(i-1)+4} = 5n(k-1) + 5(i-1) + 4$$

$$\pi_{5n(k-1)+5(i-1)+1} = -5n(k-1) - 5(i-1) - 1;$$

3)  $k=2m$  时,  $\pi^{3k}$  长度为  $6n$ ,

$$\pi^{3k} = \pi^{3k1} \pi^{3k2} \dots \pi^{3kn},$$

$$\pi^{3k1} = \pi_{5n(k-1)+5(i-1)} \pi_{5(n-i)+1} \pi_{5n(k-1)+5(i-1)+2} \pi_{5n(k-1)+5(i-1)+3} \pi_{5n(k-1)+5(i-1)+4} \pi_{5n(k-1)+5(i-1)+1},$$

各符号方向为

$$\pi_{5n(k-1)+5(i-1)} = 5n(k-1) + 5(i-1),$$

$$\pi_{5(n-i)+1} = -5(n-i) - 1,$$

$$\pi_{5n(k-1)+5(i-1)+2} = 5n(k-1) + 5(i-1) + 2,$$

$$\pi_{5n(k-1)+5(i-1)+3} = -5n(k-1) - 5(i-1) - 3,$$

$$\pi_{5n(k-1)+5(i-1)+4} = 5n(k-1) + 5(i-1) + 4,$$

$$\pi_{5n(k-1)+5(i-1)+1} = 5n(k-1) + 5(i-1) + 1.$$

4. 取集合  $L = \{5j+i | j=0, 1, 2, 3, \dots, 2mn-1, i=0, 2, 3, 4\} \cup \{10mn+1\}$ , 即要构造的目标序列  $\pi^* = \pi_0^* \pi_1^* \pi_2^* \dots \pi_{10mn-1}^* \pi_{10mn}^*$  中有  $8mn+1$  个符号方向已知, 只要确定另外  $2mn$  个符号的方向。

至此, F3ST 的实例已经构造完成。显然前述由 3SAT 问题实例构造 F3ST 问题实例是多项式时间的, 时间复杂度为  $O(mn)$ , 下面只要证明 3SAT 问题实例有解与所构造相应 F3ST 问题实例有解是等价的即可。

$\Rightarrow$ : 若 3SAT 问题有解, 则构造有向符号序列  $\pi^*$  为:

$$\pi^* = \pi_0^* \pi_1^* \pi_2^* \dots \pi_{10mn-1}^* \pi_{10mn}^*, \pi_{10mn}^* = 10mn,$$

其中的  $8mn$  个符号为

$$\pi_{5j+i}^* = 5j+i, j=0, 1, 2, 3, \dots, 2mn-1, i=0, 2, 3, 4,$$

另外  $2mn$  个符号由 3SAT 问题的解中布尔变量的取值确定, 对于  $j \in \{1, 2, \dots, n\}$ ,

1) 若  $x_j$  取值为 1, 即  $a(x_j) = 1$ , 则

$$\pi_{5n(k-1)+5(j-1)+1}^* = -5n(k-1) - 5(j-1) - 1,$$

$$k=1, 3, 5, \dots, 2m-1,$$

$$\pi_{5n(k-1)+5(n-j)+1}^* = -5n(k-1) - 5(n-j) - 1,$$

$$k=2, 4, 6, \dots, 2m;$$

2) 若  $x_j$  取值为 0, 即  $a(x_j) = 0$ , 则

$$\pi_{5n(k-1)+5(j-1)+1}^* = 5n(k-1) + 5(j-1) + 1,$$

$$k=1, 3, 5, \dots, 2m-1,$$

$$\pi_{5n(k-1)+5(n-j)+1}^* = 5n(k-1) + 5(n-j) + 1,$$

$$k=2, 4, 6, \dots, 2m.$$

再考虑  $\pi^*$  与  $\pi^1, \pi^2, \pi^3$  的距离。观察图 1, 不妨设  $c(x_{j_1}) = 1$ , 则不论  $c(x_{j_1}) = x_{j_1}$  还是  $c(x_{j_1}) = \bar{x}_{j_1}$ , 即不论该 RD 图片段中有

$$\begin{cases} \pi_{5n(k-1)+5(j_1-1)+1}^* = -5n(k-1) - 5(j_1-1) - 1 \\ \pi_{5n(k-1)+5(j_1-1)+1}^* = +5n(k-1) + 5(j_1-1) + 1 \end{cases}$$

还是

$$\begin{cases} \pi_{5n(k-1)+5(j_1-1)+1}^* = +5n(k-1) + 5(j_1-1) + 1 \\ \pi_{5n(k-1)+5(j_1-1)+1}^* = -5n(k-1) - 5(j_1-1) - 1 \end{cases}$$

均可以断定由  $c_j = c(x_{j_1}) \vee c(x_{j_2}) \vee c(x_{j_3})$  确定了一个含 7 条黑边的有序圈, 进一步分析容易知道, RD 图  $G_R(V(S), E(\pi^1), E(\pi^*))$  中有  $2m$  个含 7 条黑边的有序圈,  $2m(n-3)$  个含 5 条黑边的有序圈,  $2m \cdot 2$  个含 3 条黑边的有序圈,  $2m$  个含 2 条黑边的有序圈, 所以

$$d(\pi^*, \pi^1) = 2m(7-1) + 2m(n-3)(5-1) + 4m(3-1) + 2m(2-1) = 2m(4n-1);$$

观察图 2, 由于在该 RD 图片段中有  $\pi_{5n(k-1)+5(j-1)+1}^*$  与  $\pi_{5n(k-1)+5(n-j)+1}^*$  方向相同, 而  $\pi_{5n(k-1)+5(j-1)+1}^*$  与  $\pi_{5n(k-1)+5(n-j)+1}^*$  方向相反, 故  $\pi_{5n(k-1)+5(j-1)+1}^*$  与  $\pi_{5n(k-1)+5(n-j)+1}^*$  所在含 5 条黑边的圈是有序圈, 进而易得 RD 图  $G_R(V(S), E(\pi^2), E(\pi^*))$  有  $mn$  个含 5 条黑边的有序圈,  $mn$  个含 3 条黑边的有序圈,  $mn$  个含 2 条黑边的有序圈, 所以

$$d(\pi^*, \pi^2) = mn(5-1) + mn(3-1) + mn(2-1) = 7mn;$$

同理知道

$$d(\pi^*, \pi^3) = mn(5-1) + mn(3-1) + mn(2-1) = 7mn;$$

综上知  $\sum_{i=1}^3 d(\pi^i, \pi^*) = M$ , 故得所构造 F3ST 问题实例有解;

$\Leftarrow$ : 若所构造 F3ST 问题实例有解, 由于  $\pi^*$  与  $\pi^1, \pi^2, \pi^3$  的结构保证三个 RD 图中圈的个数是固定的, 因此要使  $\sum_{i=1}^3 d(\pi^i, \pi^*) \leq M$ , 必然 RD 图中有关的圈均为有序圈, 而且  $\sum_{i=1}^3 d(\pi^i, \pi^*) = M$ . 故由 RD 图  $G_R(V(S), E(\pi^2), E(\pi^*))$  与  $G_R(V(S), E(\pi^3), E(\pi^*))$  中含 5 条黑边的圈均为有序圈, 知道在  $\pi^*$  中集合  $L_i = \{\pi_{10m(k-1)+5(i-1)+1}^*, \pi_{10m(k-1)+5(n-i)+1}^* | k=1, 2, \dots, m\}$  中的  $2m$  个符号方向相同, 这里  $i \in \{1, 2, \dots, n\}$ , 共有  $n$  个集合。对 F3ST 实例, 如果  $L_i$  中符号方向为负, 则取  $a(x_j) = 1$ , 否则取  $a(x_j) = 0$ , 则由  $G_R(V(S), E(\pi^2), E(\pi^*))$  中含 7 条黑边的圈均为有序圈知道 F3ST 实例中的项  $c_j$  均满足,  $j \in \{1, 2, \dots, m\}$ , 即 F3ST 实例有解。

综上所述, 证明了 F3ST 问题是 NP-C 问题。

### 5 星树问题的近似算法

星树问题不存在绝对近似算法, 故只讨论常数近似算法。

#### 5.1 ST 问题的近似算法

ST 问题的优化形式为: 给定一组有向符号序列  $\pi^1, \pi^2, \dots, \pi^m$ , 求一条有向符号序列  $\pi^*$ , 使得  $\sum_{i=1}^m d(\pi^i, \pi^*)$  达到最小。

近似算法  $A_{ST}$  从给定的有向符号序列  $\pi^1, \pi^2, \dots, \pi^m$  中选取一条符号序列  $\pi^k$ , 使得  $\sum_{i=1}^m d(\pi^i, \pi^k) = \min\{\sum_{i=1}^m d(\pi^i, \pi^j), 1 \leq j \leq m\}$ , 输出  $\pi^k$  和  $\sum_{i=1}^m d(\pi^i, \pi^k)$ 。

近似算法  $A_{ST}$  的近似性能比为 2, 该算法主要是计算有向

符号序列的翻转距离,由 S. Hannenhalli 等给出的计算有向符号序列翻转距离的算法时间复杂度为  $O(n^5)$ ,因此算法  $A_{ST}$  的最坏时间复杂度为  $O(m^2n^5)$ 。

## 5.2 F3ST 问题的近似算法

F3ST 问题的优化形式为:给定三条长度为  $n$  的有向符号序列  $\pi^1, \pi^2, \pi^3$ , 集合  $S = \{1, 2, \dots, n\}$  的一个真子集  $L$ , 求一条有向符号序列  $\pi^* = \pi_1\pi_2\dots\pi_n$ , 使得  $\sum_{i=1}^3 d(\pi^i, \pi^*) = \min \sum_{i=1}^3 d(\pi^i, \pi^v)$ , 而且  $\pi_i = i, i \in L, |\pi_i| = i, i \in S-L$ 。

近似算法 A: 设计序列  $\pi^A = \pi_1\pi_2\dots\pi_n$ , 取  $\pi_i = i, i \in L$ ; 对其余的符号从小到大设置其方向, 对每一个未确定方向的符号  $i$ ,

1) 如果  $i$  与  $i+1$  在  $\pi^1$  中相邻并以  $i, i+1$  的顺序出现, 且  $i+1$  方向未定, 取  $|\pi_i| = i, |\pi_{i+1}| = i+1, \pi_i, \pi_{i+1}$  的方向分别与  $i, i+1$  在  $\pi^1$  中方向相同;

2) 如果  $i$  与  $i+1$  在  $\pi^1$  中相邻并以  $i+1, i$  的顺序出现, 且  $i+1$  方向未定, 取  $|\pi_i| = i, |\pi_{i+1}| = i+1, \pi_i, \pi_{i+1}$  的方向分别与  $i, i+1$  在  $\pi^1$  中方向相反;

3) 如果  $i$  与  $i+1$  在  $\pi^1$  中不相邻, 或  $i$  与  $i+1$  在  $\pi^1$  中相邻但  $i+1$  方向已定, 则任意指定  $i$  方向;

全部符号方向取定后输出  $\pi^A$  及  $d^A = \sum_{i=1}^3 d(\pi^i, \pi^A)$ 。

对 F3ST 问题的实例  $\Pi$ , 算法 A 输出的  $\pi^A$  及  $A(\Pi) = \sum_{i=1}^3 d(\pi^i, \pi^A)$  能够使得  $A(\Pi)/OPT(\Pi) \leq 8$ 。设  $\pi^*$  为最优解,  $\pi^{1*}$  为满足  $\pi_i = i, i \in L, |\pi_i| = i, i \in S-L$  的与  $\pi^1$  最近的序列, 则只要说明下式

$$\begin{aligned} A(\Pi) &= \sum_{i=1}^3 d(\pi^i, \pi^A) \leq d(\pi^1, \pi^A) + (d(\pi^2, \pi^1) + d(\pi^1, \pi^A)) + (d(\pi^3, \pi^1) + d(\pi^1, \pi^A)) \\ &\leq 3d(\pi^1, \pi^A) + (d(\pi^2, \pi^1) + d(\pi^3, \pi^1) + d(\pi^3)) \\ &\leq 6d(\pi^1, \pi^{1*}) + 2 \sum_{i=1}^3 d(\pi^i, \pi^*) \leq 6d(\pi^1, \pi^*) + \end{aligned}$$

$$\begin{aligned} &2 \sum_{i=1}^3 d(\pi^i, \pi^*) \\ &\leq 8 \sum_{i=1}^3 d(\pi^i, \pi^*) \\ &= 8OPT(\Pi) \end{aligned}$$

中的  $d(\pi^1, \pi^A) \leq 2d(\pi^1, \pi^{1*})$  即可。算法 A 使得  $G_R(V(S), E(\pi^1), E(\pi^A))$  中仅含一条黑边的圈与  $G_R(V(S), E(\pi^1), E(\pi^{1*}))$  中仅含一条黑边的圈相同, 假设其余的黑边为  $b$  条, 则  $d(\pi^1, \pi^A) \leq b, d(\pi^1, \pi^{1*}) \geq b-c$ , 其中  $c$  为  $G_R(V(S), E(\pi^1), E(\pi^{1*}))$  中至少含两条黑边的圈的个数, 由  $c \leq b/2$ , 即得  $d(\pi^1, \pi^A) \leq 2d(\pi^1, \pi^{1*})$ 。

## 参考文献

- 1 Sankoff D, Leduc G, Antoine N, Paquin B. Gen order comparisons for phylogenetic inference: Evolution of the mitochondrial . Pro. Natl. Acad. Sci. U. S. A. , 1992, 89: 6575~6579
- 2 Kececioğlu J, Sankoff D. Exact and approximation algorithms for the reversal distance between two permutations. Algorithmica, 1995, 13: 180~210
- 3 Kececioğlu J, Sankoff D. Efficient bounds for oriented chromosome inversion distance. In: Proc. 5th Annual symposium on combinatorial pattern matching, Lecture notes in Computer science-807, 1994. 307~325
- 4 Bafna V, Pevzner P. Sorting by reversals: genome rearrangements in plants organelles and evolutionary history of X chromosome . Mol. Biol. and Evol. . 1995, 12: 239~246
- 5 Hannenhalli S, Pevzner P A. Transforming cabbage into turnip- Polynomial algorithm for sorting signed permutations by reversals. In: Proc. of STOC'95, Las Vegas, USA, 1995. 178~179
- 6 朱大铭, 马绍汉, 等. 翻转距离星树问题的复杂度和近似算法. 软件学报录用

(上接第117页)

Agent 管理了一个可由其在本地通过执行基本活动访问的资源列表。基本活动管理了一个用以说明活动本身资源访问状态的调度实例列表。基本活动对资源的任何访问要求, 经 Agent 的允许, 将 Agent 资源列表里的资源加入到本活动的资源调度列表中, 同时由 Agent 修改资源的调度状态(加任务锁和活动锁)。基本活动技能调用完成, 则活动进入完成态, 等待上层的提交命令; 基本活动调用失败, 则基本活动进入未完成态, 由 Agent 撤消活动资源调度列表中资源的修改, 然后重置此资源为未锁定态, 最后将活动状态置为失败。顶层复合活动当其所有活动均为完成态, 则一次发出提交通知, 由基本活动完成其资源调度列表中所有资源的提交, 然后重置资源的调度状态, 归还资源。

**结论和后续工作** 本文提出了面向资源的基于多 Agent 虚拟组织框架的扩展, 加强了框架资源包容、调度和事务建模的能力, 为基于多 Agent 虚拟组织框架的数据流管理系统 SaFlow 提供了事务模型, 即异常和失败发生的情况下保证数据正确性和依赖性的基础。

显式的资源定义抽象了资源存取细节, 为系统提供资源存取了统一的方式和清晰的语义, 区分了工作流的共享资源和中间数据, 并灵活控制了和并发控制的粒度。资源调度作

为单独的模块从技能模块中分离, 在活动规划中动态生成调度计划, 由 Agent 提供独立的调度控制, 保持了原有框架任务规划和活动协调的灵活性和交互能力的。资源访问合成于技能之中, 实现了资源访问的灵活性以及同领域操作的协作和交互。资源访问缓冲区的引入, 以及显式的资源访问依赖性说明, 实现了活动执行的提交概念, 并为活动的撤消和恢复奠定了基础。

以此为基础, 我们将为 SaFlow 研究更加灵活可靠的异常处理机制。

## 参考文献

- 1 Alonso C, et al. Advanced Transaction Models in Workflow Contexts [R]: [Technical Report]. IBM Almaden Research Center, 1996
- 2 Hagen C, Alonso G. Flexible Exception Handling in Process Support Systems [R]: [Technical Report No. 290 ETH Zurich]. Department of Computer Science, Feb. 1998
- 3 高济, 王进. 基于 Agents 的软件合成框架 ABFSC [J]. 计算机学报, 1999, 22(10): 1050~1058
- 4 骆海波. 一种基于多软件 Agent 的工作流管理系统 [D]: [浙江大学硕士学位论文]. 1999. 2