目标序列部分确定的翻转距离星树问题*

栾峻峰+, 朱大铭, 马绍汉

(山东大学 计算机科学与技术学院,山东 济南 250100)

A Problem of Reversal Distance on Star-Tree with Object Partially Fixed

LUAN Jun-Feng+, ZHU Da-Ming, MA Shao-Han

(School of Computer Science and Technology, Shandong University, Ji'nan 250100, China)

+ Corresponding author: Phn: 86-531-8395772, E-mail: jfluan@sdu.edu.cn

http://www.sdu.edu.cn

Received 2001-10-23; Accepted 2002-02-26

Luan JF, Zhu DM, Ma SH. A problem of reversal distance on star-tree with object partially fixed. *Journal of Software*, 2003,14(2):183~189.

Abstract: A problem of reversal distance on star-tree is discussed. The problem of 3SAT is induced to the problem of the reversal distance on star-tree with object partially fixed. The fact described below is proved: if an instance has 3 sequences given, and the order and some signs of the symbols in aimed sequence are fixed, it is still NP-hard to solve the problem of reversal distance on star-tree in which only need to decide the signs of the other symbols to minimizing the sum of distance between object and the given sequences. A polynomial approximation algorithm for the problem is given.

Key words: algorithm; computational complexity; evolutionary tree; genome; reversal distance

摘 要: 讨论翻转距离星树问题,将3SAT问题归约到目标序列部分固定的翻转距离星树问题,证明实例中当有向符号序列个数为 3 时,若目标序列符号顺序固定,且有部分符号方向给定,则只确定其余符号方向以使得目标序列与已知3条给定序列翻转距离之和最小所对应的翻转距离星树问题也是NP-难解问题.同时,还给出了该问题的多项式时间近似算法.

关键词: 算法;计算复杂性;进化树;基因组;翻转距离

中图法分类号: TP301 **文献标识码**: A

基因组翻转距离计算问题由 D.Sankoff 等人提出[1],其极强的理论及实用价值吸引了许多专家的注意[2~4],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[5].D.Sankoff 等人对多种形式的进化距离,提出各种形式的进化树问题,这些进化树问题都有很强的应用背景,其中星树问题是进化树问题的最简单形式,在进化树类问题研究中具有重要价值.朱大铭等人证明了当实例中给定序列为m条时,对应的翻转距离星树问题(ST问题)是NP-难解的[6].本文证明当实例中给定序列为3条时,如果固定目标序列符号顺序,且有部分符号方向给定,则仅确定目标序列其余符号方向所对应的翻转距离星树

* Supported by the National Natural Science Foundation of China under Grant No.60073042 (国家自然科学基金)

第一作者简介: 栾峻峰(1974－),男,山东青岛人,博士生,讲师,主要研究领域为算法分析与设计.

问题也是 NP-难解的.本文将该问题简记为 F3ST 问题.

有向符号序列的翻转距离问题. 设 $S=\{1,2,...,n\}$ 为由 n 个符号组成的集合, S 的反方向集合为 $S^{-}=\{-1,-2,\ldots,-n\}$.考虑由 S 的所有有向符号组成的一个排列 $\pi=\pi_1\pi_2..\pi_n$, $\pi_i\in S\cup S^{-}$,如果 $\pi=|\pi_1||\pi_2|...|\pi_n|$ 恰为符号集 S 的一个置换,则称 π 为长度为 n 的有向符号序列,有向符号序列π的一次翻转 $\rho(i,j)$ 使得π变成一个新的有向符号序列 $\pi\cdot\rho(i,j)=\pi_1\pi_2\ldots\pi_{i-1}-\pi_j-\pi_{j-1}\ldots\pi_i\pi_{j+1}\ldots\pi_n$.有向符号序列的翻转距离问题即为给定符号集合 S 的两个有向符号序列 π^1 和 π^2,求常数 k,使得存在一组翻转 $\rho_1,\rho_2,...,\rho_k$,完成 $\pi^1\cdot\rho_1\cdot\rho_2\cdot...\cdot\rho_k=\pi^2$,并且 k 最小.称 k 为序列 π^1 和 π^2 的翻转距离,记为 $d(\pi^1,\pi^2)=k$.

当计算任意两个有向符号序列 π^1 和 π^2 的翻转距离时,先构造 RD 图 $G_R(V(S),E(\pi^1),E(\pi^2))$,然后在 RD 图中定义 C 图、H 图、F 图,并设 RD 图中黑边数及 C 图、H 图、F 图个数分别为 b,c,h,f,则 $d(\pi^1,\pi^2)=b-c+h+f$ [5].

ST 问题. 给定一组长度为 n 的有向符号序列 $\pi^1,\pi^2,...,\pi^m$ 和正整数 M,问是否存在一条有向符号序列 π^*,使得 $\sum_{i=1}^{m}d(\pi^i,\pi^*)\le M$.

F3ST 问题. 给定 3 条长度为 n 的有向符号序列 π^1,π^2,π^3,正整数 M 以及 $S=\{1,2,\ldots,n\}$的一个真子集 L,问是否存在一条有向符号序列 $\pi^*=\pi_1\pi_2..\pi_n$,使得 $\sum_{i=1}^{3}d(\pi^i,\pi^*)\le M$,而且 $\pi_i=i$, $i\in L$, $|\pi_i|=i$, $i\in S-L$.

定理 1. F3ST 问题是 NP-C 问题.

本文将 3SAT 问题归约到 F3ST 问题,从而证明 F3ST 问题是 NP-C 问题.

1 由 3SAT 问题实例构造 F3ST 问题实例

布尔变量集合 $X=\{x_1,x_2,...,x_n\}$,项集合 $C=\{c_1,c_2,...,c_m\}$,其中 $c_j=c(x_{j_1})\vee c(x_{j_2})\vee c(x_{j_3})$, $1\le j\le m$, $1\le j_1<j_2<j_3\le n$, $c(x_i)\in\{x_i,\overline{x}_i\}$, $1\le i\le n$.取 $M=22mn-2m$,并构造 3 条长度为 $10mn+1$ 的有向符号序列 π^1,π^2,π^3 和集合 L.

1.1 序列π^1的构造

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

其中序列片段π^{1k}的长度为 $5n$,符号顺序如下确定:

(1) 当 k 为奇数时,设 $k=2j-1,j=1,\ldots,m$,则π^{1k} 由 $c_j=c(x_{j_1})\vee c(x_{j_2})\vee c(x_{j_3})$ 确定(如图 1 所示).

$$\pi^{1k}=\pi^{1k0}\pi^{1k1}\pi^{1k2}\pi^{1k3}\pi^{1k4}\pi^{1k5}\pi^{1k6},$$

$$\pi^{1k0}=\pi^{1k*1}\pi^{1k*2}...\pi^{1k*,j_1-1},$$

$$\pi^{1k1}=\pi_{5n(k-1)+5(j_1-1)}\pi_{5n(k-1)+5(j_1-1)+2}\pi_{5n(k-1)+5(j_1-1)+4}\pi_{5n(k-1)+5(j_1-1)+3},$$

$$\pi^{1k2}=\pi^{1k*,j_1+1}\pi^{1k*,j_1+2}...\pi^{1k*,j_2-1},$$

$$\pi^{1k3}=\pi_{5n(k-1)+5(j_2-1)}\pi_{5n(k-1)+5(j_1-1)+1}\pi_{5n(k-1)+5(j_3-1)}\pi_{5n(k-1)+5(j_2-1)+2}\pi_{5n(k-1)+5(j_2-1)+4}\pi_{5n(k-1)+5(j_2-1)+3},$$

$$\pi^{1k4}=\pi^{1k*,j_2+1}\pi^{1k*,j_2+2}...\pi^{1k*,j_3-1},$$

$$\pi^{1k5}=\pi_{5n(k-1)+5(j_2-1)+1}\pi_{5n(k-1)+5(j_3-1)+2}\pi_{5n(k-1)+5(j_3-1)+3}\pi_{5n(k-1)+5(j_3-1)+4}\pi_{5n(k-1)+5(j_3-1)+1},$$

$$\pi^{1k6}=\pi^{1k*,j_3+1}\pi^{1k*,j_3+2}...\pi^{1k*n},$$

其中

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

下面确定序列中各符号的方向:

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

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

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

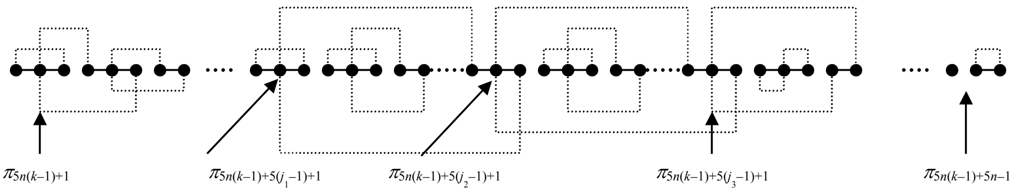

Fig.1 Segment π^{1k} constructed in $G_R(V(S),E(\pi^1),E(\pi^*))$ (k is odd)

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

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

$$j'_1=n+1-j_3,$$
$$j'_2=n+1-j_2,$$
$$j'_3=n+1-j_1,$$

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

$$\pi_{5n(k-1)+5j+i}=5n(k-1)+5j+i,\ j=0,1,2,\dots,n-1,\ i=0,2,4,$$
$$\pi_{5n(k-1)+5j+3}=-5n(k-1)-5j-3,$$
$$\pi_{5n(k-1)+5j+1}=\begin{cases}-5n(k-1)-5j-1, & c(x_{n+1-j})=\bar{x}_{n+1-j},\ j\in\{j'_1,j'_2,j'_3\}\\ 5n(k-1)+5j+1, & \text{others}\end{cases}.$$

1.2 序列π^2的构造

$$\pi^2=\pi^{21}\pi^{22}\pi^{23}\pi^{24}\dots\pi^{2,2m-1}\pi^{2,2m}\pi_{10mn},\pi_{10mn}=10mn,$$

其中序列片段π^{2k} 如下确定(如图 2 所示):

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

$$\pi^{2k}=\pi^{2k1}\pi^{2k2}\dots\pi^{2kn},$$
$$\pi^{2ki}=\pi_{5n(k-1)+5(i-1)}\pi_{5n(k-1)+5(i-1)+2}\pi_{5n(k-1)+5(i-1)+4}\pi_{5n(k-1)+5(i-1)+3},$$

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

$$\pi^{2k}=\pi^{2k1}\pi^{2k2}\dots\pi^{2kn},$$
$$\pi^{2ki}=\pi_{5n(k-1)+5(i-1)}\pi_{5n(k-2)+5(n-i)+1}\pi_{5n(k-1)+5(i-1)+2}\pi_{5n(k-1)+5(i-1)+3}\pi_{5n(k-1)+5(i-1)+4}\pi_{5n(k-1)+5(i-1)+1},$$

各符号方向为

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

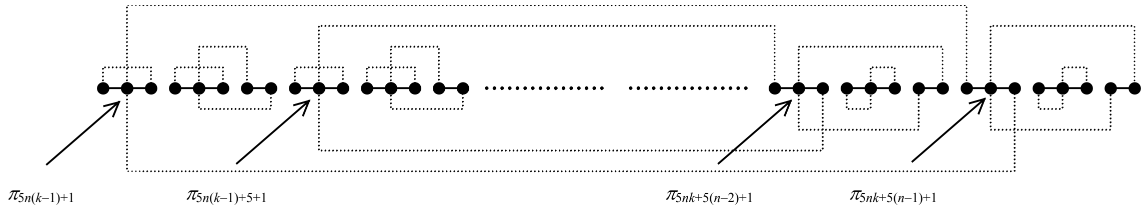

Fig.2 Segment $\pi^{2k}\pi^{2,k+1}$ constructed in $G_R(V(S),E(\pi^2),E(\pi^k))$ (k is odd)

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

1.3 序列π^3的构造

$$\pi^3=\pi^{31}\pi^{32}\pi^{33}\pi^{34}\ldots\pi^{3,2m-1}\pi^{3,2m}\pi_{10mn},\pi_{10mn}=10mn,$$

其中序列片段π^{3k}如下确定(如图 3 和图 4 所示):

(1) 当 k=1,2,4,6,…,2(m−1)时,π^{3k}的长度为 4n,

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

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

各符号方向为

$$\begin{aligned}
\pi_{5n(k-1)+5(i-1)}&=5n(k-1)+5(i-1),\\
\pi_{5n(k-1)+5(i-1)+2}&=5n(k-1)+5(i-1)+2,\\
\pi_{5n(k-1)+5(i-1)+4}&=5n(k-1)+5(i-1)+4,\\
\pi_{5n(k-1)+5(i-1)+3}&=-5n(k-1)-5(i-1)-3.
\end{aligned}$$

(2) 当 k=3,5,7,…,2m−1 时,π^{3k}的长度为 6n,

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

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

各符号方向为

$$\begin{aligned}
\pi_{5n(k-1)+5(i-1)}&=5n(k-1)+5(i-1),\\
\pi_{5n(k-2)+5(n-i)+1}&=5n(k-2)+5(n-i)+1,\\
\pi_{5n(k-1)+5(i-1)+2}&=5n(k-1)+5(i-1)+2,\\
\pi_{5n(k-1)+5(i-1)+3}&=-5n(k-1)-5(i-1)-3,\\
\pi_{5n(k-1)+5(i-1)+4}&=5n(k-1)+5(i-1)+4,\\
\pi_{5n(k-1)+5(i-1)+1}&=-5n(k-1)-5(i-1)-1.
\end{aligned}$$

(3) 当 k=2m 时,π^{3k}的长度为 6n,

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

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

各符号方向为

$$\begin{aligned}
\pi_{5n(k-1)+5(i-1)}&=5n(k-1)+5(i-1),\\
\pi_{5(n-i)+1}&=-5(n-i)-1,\\
\pi_{5n(k-1)+5(i-1)+2}&=5n(k-1)+5(i-1)+2,\\
\pi_{5n(k-1)+5(i-1)+3}&=-5n(k-1)-5(i-1)-3,\\
\pi_{5n(k-1)+5(i-1)+4}&=5n(k-1)+5(i-1)+4,\\
\pi_{5n(k-1)+5(i-1)+1}&=5n(k-1)+5(i-1)+1.
\end{aligned}$$

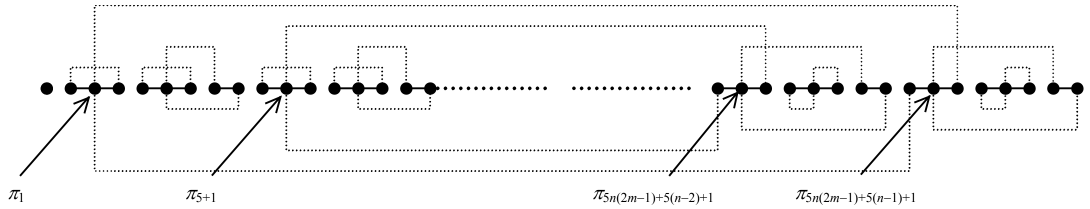

Fig.3 Segments π^{31} and $\pi^{3,2m}$ constructed in $G_R(V(S),E(\pi^2),E(\pi^*))$

图 3 欲构造的 RD 图 $G_R(V(S),E(\pi^3),E(\pi^*))$中$\pi^{31}$与$\pi^{3,2m}$段

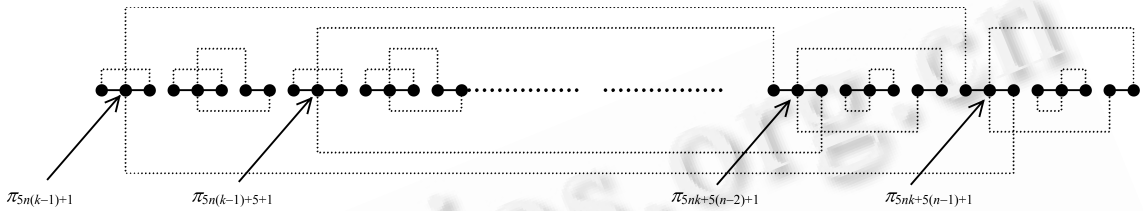

Fig.4 Segments $\pi^{3k}\pi^{3,k+1}$ constructed in $G_R(V(S),E(\pi^3),E(\pi^*))$ (k is enen)

图 4 欲构造的 RD 图 $G_R(V(S),E(\pi^3),E(\pi^*))$中$\pi^{3k}\pi^{3,k+1}$段($k$ 为偶数)

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

2 3SAT 问题实例有解与相应 F3ST 问题实例有解的等价性

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π^*为

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

其中的 $8mn$ 个符号为

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

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

(1) 若 $a(x_j)=1$,则

$$\pi_{5n(k-1)+5(j-1)+1}^*=-5n(k-1)-5(j-1)-1,k=1,3,5,...,2m-1,$$
$$\pi_{5n(k-1)+5(n-j)+1}^*=-5n(k-1)-5(n-j)-1,k=2,4,6,...,2m;$$

(2) 若 $a(x_j)=0$,则

$$\pi_{5n(k-1)+5(j-1)+1}^*=5n(k-1)+5(j-1)+1,k=1,3,5,...,2m-1,$$
$$\pi_{5n(k-1)+5(n-j)+1}^*=5n(k-1)+5(n-j)+1,k=2,4,6,...,2m.$$

再考虑 π^* 与 π^1,π^2,π^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 $c_j=c(x_{j_1})\vee c(x_{j_2})\vee c(x_{j_3})$ 确定了一个含有 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 $2m$ 个含有 7 条黑边的有序圈,$2m(n-3)$个含有 5 条黑边的有序圈,$2m\cdot 2$ 个含有 3 条黑边的有序圈,$2m$ 个含有 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 $\pi^*_{5n(k-1)+5(j-1)+1}$ 与 $\pi^*_{5nk+5(n-j)+1}$ 方向相同,而 $\pi_{5n(k-1)+5(j-1)+1}$ 与 $\pi_{5nk+5(n-j)+1}$ 方向相反,故 $\pi_{5n(k-1)+5(j-1)+1}$ 与 $\pi_{5nk+5(n-j)+1}$ 所在含有 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 个含有 5 条黑边的有序圈,mn 个含有 3 条黑边的有序圈,mn 个含有 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,

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

综上,$\sum_{i=1}^3 d(\pi^i,\pi^*)=M$,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 π^* 与 π^1,π^2,π^3 的结构保证 3 个 RD 图中圈的个数是固定的,所以要使 $\sum_{i=1}^3 d(\pi^i,\pi^*)\le M$,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 $G_R(V(S),E(\pi^3),E(\pi^*))$ 中含有 5 条黑边的圈均为有序圈,得知在 π^* 中集合 $L_i=\{\pi^*_{10n(k-1)+5(i-1)+1},\pi^*_{10n(k-1)+5(n-i)+1}|k=1,2,\ldots,m\}$中的 $2m$ 个符号方向相同,这里 $i\in\{1,2,\ldots,n\}$,共有 n 个集合.对 3SAT 实例,如果 L_i 中符号方向为负,则取 $a(x_i)=1$,否则取 $a(x_i)=0$,则由 $G_R(V(S),E(\pi^2),E(\pi^*))$中含有 7 条黑边的圈均为有序圈得知,3SAT 实例中的项 c_i 均满足,$j\in\{1,2,\ldots,m\}$,即 3SAT 实例有解.

综上所述,我们证明了 F3ST 问题是 NP-难解问题.

3 F3ST 问题的近似算法

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

算法 1. 设计序列 $\pi^A=\pi_1\pi_2\ldots\pi_n$,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,有

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

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

(3) 如果 i 与 i+1 在 π^1 中不相邻,或 i 与 i+1 在 π^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 π^A 及 $d^A=\sum_{i=1}^3 d(\pi^i,\pi^A)$.

定理 2. 对 F3ST 问题的实例Π,设算法 A 输出 π^A 及 $A(\Pi)=\sum_{i=1}^3 d(\pi^i,\pi^A)$,则 $A(\Pi)/OPT(\Pi)\le 8$.

证明:设 π^* 为最优解,π^{1*} 为满足 $\pi_i=i$, $i\in L$, $|\pi_i|=i$, $i\in S-L$ 的与 π^1 最近的序列,则我们只要说明下式中的 $d(\pi^1,\pi^A)\le 2d(\pi^1,\pi^{1*})$ 即可.

$$\begin{aligned}A(\Pi)&=\sum_{i=1}^3 d(\pi^i,\pi^A)\le d(\pi^1,\pi^A)+(d(\pi^2,\pi^1)+d(\pi^1,\pi^A))+(d(\pi^3,\pi^1)+d(\pi^1,\pi^A))\\&\le 3d(\pi^1,\pi^A)+(d(\pi^2,\pi^1)+d(\pi^3,\pi^1)+d(\pi^2,\pi^3))\\&\le 6d(\pi^1,\pi^{1*})+2\sum_{i=1}^3 d(\pi^i,\pi^*)\\&\le 6d(\pi^1,\pi^*)+2\sum_{i=1}^3 d(\pi^i,\pi^*)\\&\le 8\sum_{i=1}^3 d(\pi^i,\pi^*)\\&=8OPT(\Pi).\end{aligned}$$

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 $d(\pi^1,\pi^A)\le b$, $d(\pi^1,\pi^{1*})\ge b-c$,其中 c 为 $G_R(V(S),E(\pi^1),E(\pi^{1*}))$中至少含有两条黑边的圈的个数,由 $c\le b/2$,即得 $d(\pi^1,\pi^A)\le 2d(\pi^1,\pi^{1*})$.

References:

[1] Sankoff D, Leduc G, Antoine N, *et al*. Gene order comparisons for phylogenetic inference: evolution of the mitochondrial. Proceedings of the National Academy of Sciences, 1992,89:6575~6579.

[2] Kececioglu J, Sankoff D. Exact and approximation algorithms for the reversal distance between two permutations. Algorithmica, 1995,13(1/2):180~210.

[3] Kececioglu J, Sankoff D. Efficient bounds for oriented chromosome inversion distance. In: Proceedings of the 5th Annual Symposium on Combinatorial Pattern Matching. Lecture notes in Computer Science 807, Berlin: Springer-Verlag, 1994. 307~325.

[4] Bafna V, Pevzner PA. Sorting by reversals: genome rearrangements in plants organelles and evolutionary history of X chromosome. Molecular Biology and Evolution, 1995,12:239~246.

[5] Hannenhalli S, Pevzner PA. Transforming cabbage into turnip——polynomial algorithm for sorting signed permutations by reversals. In: Proceedings of the 27th Annual Symposium on the Theory of Computing (STOC'95). Las Vegas, Nevada, 1995. 178~189.

[6] Zhu DM, Ma SH, Lei P. Computational complexity and an approximation algorithm for star-tree phylogeny problem with reversal distance. Journal of Software, 2002,13(6):1117~1122 (in Chinese with English Abstract).

附中文参考文献:

[6] 朱大铭,马绍汉,雷鹏.翻转距离星树问题的计算复杂度和近似算法.软件学报,2002,13(6):1117~1122.