

排列短块移动排序距离的新下界

王彤 姜海涛 朱大铭

(山东大学计算机科学与技术学院 济南 250101)

(wtwq.5.16@163.com)

A New Lower Bound for the Distance of Sorting Permutations by Short Block Moves

Wang Tong, Jiang Haitao, and Zhu Daming

(School of Computer Science and Technology, Shandong University, Jinan 250101)

Abstract Over the past twenty years, computational biology has been trying tracing the rule of evolution by genome rearrangement events, therefore the rearrangement problem of genomic permutations has been researched widely and deeply. The operations of genome rearrangement include reversals, translocations, transpositions and so on. Bulteau et al proved that the problem of sorting permutations by transpositions was NP-complete. A transposition is also called a block-move. The short block-move is the most common case of block-moves. A short block-move is a block-move whose segments are moved from its original position at most two positions, so its another name is 3-bounded transposition. For the problem of the distance of sorting permutations by short block moves, we present an implicit lower bound. In this paper, we use a special permutation called “double increasing permutation” and gain a new lower bound through analyzing the double increasing permutation. On this basis, we analyse all the maximal double increasing sub-permutations of the original permutations, and then simply sum the lower bounds to obtain a new lower bound for the distance of sorting permutations by short block moves, which improves Heath and Vergara’s result greatly, so that we will lay a foundation of the design of a better approximate algorithm.

Key words short block move; hop; skip; inversion; a double increasing permutation

摘要 近 20 年来,计算生物学领域一直试图用基因组重组事件来追溯物种进化的规律,因此基因组排列的重组排序问题被广泛而深入地研究.基因组重组包含翻转、移位、转位等多种形式. Bulteau 等人证明排列的转位排序问题是 NP-完全的.一次转位操作也称为一次块移动,短块移动是最常见的一种块移动.一次短块移动是将一个元素从排列中某个位置移动到最多偏离原来 2 个位置的块移动,因此也称为 3-bounded 转位.针对排列短块移动排序距离问题,给出了一类特殊排列(称之为双递增排列)的短块移动排序次数的下界.以此为依据,分析原始排列中的所有最大双递增子排列,从而给出了任意排列短块移动排序次数的下界,改进了 Heath 和 Vergara 的负面结果,并为更好的近似算法的设计打下基础.

关键词 短块移动;跳;换位;逆序;双重递增序

中图法分类号 TP301.6

收稿日期:2014-11-17;修回日期:2015-01-20

基金项目:国家自然科学基金项目(61202014,61472222);山东省自然科学基金项目(ZR2012FQ008);中国博士后科学基金特别资助项目(2012T50614);中国博士后科学基金面上资助项目(2011M501133)

通信作者:姜海涛(htjiang@sdu.edu.cn)

基因组控制了生物体的独特性,在使用基因组数据估算 2 种生物体之间的进化距离时,我们希望重建出将一个基因组转变为另一个基因组的进化序列. 20 世纪 80 年代,人们发现一些物种的基因组在本质上由相同的基因组成,但其基因的排列顺序不同^[1-2]. 为了对比 2 个近亲生物体,可将每个生物体的基因组假设为来自相同基因集合的一个排列,通过基因组重排的若干基本操作(翻转、转位、移位等)^[3],将一个基因组(排列)转变为另一个基因组(排列). 生物学家希望得到这些操作的最少数目,并将其作为 2 个基因组间的进化距离.

一次转位操作是在一个排列上进行的重排操作. 确切地说,一次转位操作将一个片段从排列中分割出来,并将其粘贴到原排列的不同位置. 实际上,通过转位操作可以将排列中的 2 个邻接子排列进行位置交换,因此转位操作也称为块移动. 令生物学家感兴趣的是,如何通过最少次数的转位操作,将一个排列转变成另外一个排列. Bafna 和 Pevzner^[4]最先对这个问题进行了研究,并设计了一个近似性能比为 1.5 的近似算法,其时间复杂度为 $O(n^2)$. Elias 和 Hartman^[5]将其近似性能比改进至 1.375. Eriksson 等人^[6]证明了当 $n \geq 9$ 时,转位直径的界为 $\lfloor (2n-2/3) \rfloor$,并给出了一个排列的翻转排序的确切表示. 不久前, Bultheau 等人^[7]证明排列的转位排序问题是 NP-完全的.

块移动排序过程中,可以合理假定一个片段移动距离较远的可能性不大,即绝大部分块移动都是在基因组上有限长度的范围内进行的. 因此, Heath 和 Vergara^[8]提出了短块移动排序问题. 短块移动(也称为 3-bounded 转位^[9])是排列上的一个转位操作,要求进行位置交换的 2 个片段的总长度最大为 3. Heath 和 Vergara^[10]对于短块移动排序问题,设计了一个 $4/3$ -近似算法;同时,通过图匹配技巧,对于交叉双带排列这种特殊的一类排列,他们也给出了一个时间复杂度为 $O(n^3 \log n)$ 的近似算法. Mahajan 等人^[9]简化了 Heath 和 Vergara 的近似算法,并提出了一个针对 correcting-hop-free 排列最优排序的线性时间算法. Jiang 等人^[11]通过对具有“伞”结构的特殊排列排序,设计了 $O(n^2)$ 近似算法,并针对这一类逆序的个数远多于元素个数的特殊排列,给出了一个 $(1+\epsilon)$ -近似算法. 之后, Jiang 和 Zhu^[12]对于一般排列的短块移动排序问题,又设计出一个 $14/11$ -近似算法.

本文给出了一类特殊排列(双递增排列)的短块

移动排序次数的下界,以此为依据,分析原始排列中的所有最大双递增子排列,从而给出了任意排列短块移动排序次数的下界,改进了 Heath 和 Vergara 的负面结果,并为更好的近似算法设计打下基础.

1 定 义

通常,在基因重排的相关文章中,基因组以排列的方式表示,排列中的每个符号代表一个基因. 若一个排列含有 n 个符号,则采用整数集合 $I_n = \{1, 2, \dots, n\}$ 表示该排列的所有符号. 例如 $\pi = [3, 1, 2, 5, 4, 8, 6, 7, 9]$ 为一个包含 9 个符号的排列. 定义单位排列 $l_n = [1, 2, \dots, n-1, n]$, 即 l_n 中符号均已由小到大排列. 一个块(也称为子排列)是一个连续的符号片段或者仅为一个符号. 在排列 $\pi = [\pi_1, \pi_2, \dots, \pi_n]$ 中, π_i 称为 π_{i+1} 的左邻, π_{i+1} 称为 π_i 的右邻($1 \leq i \leq n-1$).

对于排列 $\pi = [\pi_1, \pi_2, \dots, \pi_n]$ ($\pi_i \in I_n, 1 \leq i \leq n$), 一次块移动操作 $\rho(i, j, k)$ ($1 \leq i < j < k \leq n+1$) 从原排列 π 中分割出块 $[\pi_i, \pi_{i+1}, \dots, \pi_{j-1}]$, 并将其迁移到 π_{k-1} 的右侧. 对排列 π 进行块移动操作 $\rho(i, j, k)$ 会产生一个新排列 $\pi' = [\pi_1, \pi_2, \dots, \pi_{i-1}, \pi_j, \dots, \pi_{k-1}, \pi_i, \dots, \pi_{j-1}, \pi_k, \dots, \pi_n]$. 过程如图 1 所示:

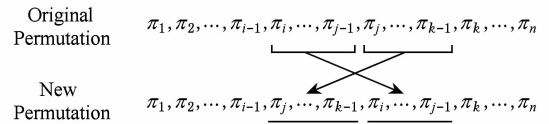


Fig. 1 A short block move.

图 1 一次块移动

由图 1 可以看到,一次块移动 $\rho(i, j, k)$ 交换了 2 个块 $[\pi_i, \dots, \pi_{j-1}]$ 和 $[\pi_j, \dots, \pi_{k-1}]$. 其中,若 $i=1$, 则 $[\pi_1, \pi_2, \dots, \pi_{i-1}] = \emptyset$; 若 $k=n+1$, 则 $[\pi_k, \dots, \pi_n] = \emptyset$. 短块移动为块移动中将进行位置交换的 2 个块的总长度限定为不超过 3 的特殊情况,即若 $k-i \leq 3$, 则称这种转位为短块移动. 若位置 i 确定, 则可能的短块移动操作有以下 3 种:

- 1) skip: $\rho([i], [i+1])$, π_i 与 π_{i+1} 换位, 也称 π_i 或 π_{i+1} 的 skip;
- 2) right-hop: $\rho([i], [i+1, i+2])$, π_i 与块 $[\pi_{i+1}, \pi_{i+2}]$ 换位, 也称 π_i 的右跳;
- 3) left-hop: $\rho([i, i+1], [i+2])$, 块 $[\pi_i, \pi_{i+1}]$ 与 π_{i+2} 换位, 也称 π_{i+2} 的左跳.

为方便起见,本文使用 $\rho(\pi_i, \pi_{i+1})$ 表示一次 skip, $\rho(\pi_i, \pi_{i+1} \pi_{i+2})$ 表示一次 right-hop, $\rho(\pi_i \pi_{i+1}, \pi_{i+2})$ 表示一次 left-hop.

π_{i+2})表示一次 left-hop. 短块移动 ρ 将排列 π 转化为一个新的排列 $\pi'=\pi \cdot \rho$.

下面对短块移动排序问题进行形式化描述:

输入: 整数排列 $\pi=[\pi_1, \pi_2, \cdots, \pi_n](\pi_i \in I_n, 1 \leq i \leq n)$.

询问: 是否存在一个短块移动序列 $\rho_1, \rho_2, \cdots, \rho_t$, 使得 $\pi \cdot \rho_1 \cdot \rho_2 \cdot \cdots \cdot \rho_t = l_n = [1, 2, \cdots, n-1, n]$, 且 t 值最小?

将排列 π 转化为单位排列 l 所需最少的短块移动次数 t , 称为 π 的短块移动距离, 记为 $K_3(\pi)$.

如果 π_i 和 π_j 满足 $\pi_i > \pi_j$ 且 $i < j$, 则称 π_i 和 π_j 为逆序排列对. 显然在单位排列 l_n 中不存在任何逆序. 将能够至少减少 π 的一个逆序的短块移动称为正确移动, 一次正确的短块移动修正被移动符号的相对顺序, 如一次正确的 skip 消除一个逆序、一次正确的 hop 消除一对逆序.

Heath 和 Vergara 证明了在寻找一个排列的最优短块移动序列时, 只需要考虑正确的移动操作即可.

定理 1^[8]. 给定排列 π , 必存在最优的短块移动序列: $\rho_1, \rho_2, \cdots, \rho_t$, 将 π 变为目标排列, 且每次移动 $\rho_i (1 \leq i \leq t)$ 都是正确的.

在下文讨论排列的短块移动距离时, 仅考虑正确的短块移动序列. 因为正确的 hop 数越多, 将排列 π 转化为单位排列 l 所需的短块移动次数越少, 则排列的短块移动排序的目标即为使正确的 hop 数最大.

在解决短块移动排序问题时, 引入排列图的概念. 给定排列 $\pi=[\pi_1, \pi_2, \cdots, \pi_n]$, 构造排列图 $G(\pi)=(V, E)$ 如下: $V=\{\pi_1, \pi_2, \cdots, \pi_n\}$, $E=\{(\pi_i, \pi_j) | i < j \text{ 且 } \pi_i > \pi_j\}$. 图 2 为一个包含 9 个符号的排列对应的排列图.



Fig. 2 The permutation graph for $\pi=[3, 1, 2, 5, 4, 8, 6, 7, 9]$.

图 2 $\pi=[3, 1, 2, 5, 4, 8, 6, 7, 9]$ 的排列图

若一个点为孤点, 则称该点是无弧的, 即该点不是任意弧的端点. $G(\pi)$ 中的每一条弧均代表了排列 π 的一个逆序. 定义排列图中的弧 (π_i, π_j) 的长度为 $k=j-i$, 并将长度为 1 的弧称为短弧. 在一个排列图中, 称 2 条弧是相关的, 当且仅当他们有相同的头或尾. 若一个弧不与其他任何弧共享头或尾, 则称这

种弧为孤立弧.

由上所述, 一次正确的短块移动操作消除排列图中对应的弧: 一次正确的 skip 消除一条弧, 一次正确的 hop 消除 2 条相关弧. 对于排列 $\pi=[3, 1, 2, 5, 4, 8, 6, 7, 9]$, 其一个最优短块移动序列为 $\rho(3, 12), \rho(8, 67), \rho(5, 4)$. 表 1 为每次短块移动消除弧的对应关系表. 若存在 2 条弧可通过一次 hop 操作消除, 则称这 2 条弧构成了一个 hop-match.

Table 1 Arcs Erased by an Optimal Short Block Move Sequence ($\pi=[3, 1, 2, 5, 4, 8, 6, 7, 9]$)

表 1 通过一个最优短块移动序列的弧消除

Block Move	Arcs
$\rho(3, 12)$	$(3, 1), (3, 2)$
$\rho(8, 67)$	$(8, 6), (8, 7)$
$\rho(5, 4)$	$(5, 4)$

方便起见, 在下文中对排列图进行讨论时, 不再对逆序与弧、排列与其对应的排列图进行区分.

文献[10]由此给出短块移动距离的上界和下界如下:

$$\left\lceil \frac{E(G(\pi))}{2} \right\rceil \leq K_3(\pi) \leq |E(G(\pi))|.$$

即任意排列的短块移动距离不大于其排列图的弧数且不小于排列图弧数的一半.

2 新下界

在 Heath 和 Vergara^[10]的方法中, 对于给定排列 $\pi=[\pi_1, \pi_2, \cdots, \pi_n]$, 考虑如何以最少的移动消除 $G(\pi)$ 中所有弧. 排序 π 时, 每个孤立弧必须通过一次 skip 消除, 显然一次消除孤立弧的移动一定是正确的. 因此设 $G(\pi)$ 共含有 m 条弧, 其中的孤立弧数为 n , 则可给出短块移动排序次数的下界为 $K_3(n) \geq$

$$\left\lceil \frac{m-n}{2} + n \right\rceil.$$

本节中, 给出了一类称之为双递增排列的特殊排列. 通过分析原始排列中的所有双递增子排列, 从而给出了任意排列短块移动排序次数的一个更紧的新下界.

在文献[12]中, 姜海涛等人初步改进了 Heath 等人^[10]给出的下界. 它们定义了一种特殊的包含 8 个元素的子排列, 称之为双蘑菇结构, 并证明任何双蘑菇结构的短块移动排序过程中至少需要 1 次 skip 操作. 图 3 为一个双蘑菇结构排列图:

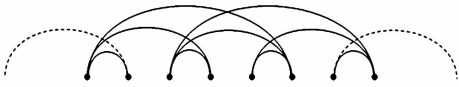


Fig. 3 The permutation graph for a double mushroom.

图3 双蘑菇结构的排列图

本文中的下界实际上是双蘑菇阵列的一种推广.

定义 1. 一个序列 $\pi = [\pi_1, \pi_2, \dots, \pi_n]$ 称为一个双递增排列, 当 π 可以划分成 2 个不相交的排列 $\pi_A = [\pi_{p_1}, \pi_{p_2}, \dots, \pi_{p_i}]$ 和 $\pi_B = [\pi_{q_1}, \pi_{q_2}, \dots, \pi_{q_j}]$, 且满足如下条件:

- 1) $p_k < p_{k+1}, \pi_{p_k} < \pi_{p_{k+1}}$, 其中 $1 \leq k \leq i-1$;
- 2) $q_r < q_{r+1}, \pi_{q_r} < \pi_{q_{r+1}}$, 其中 $1 \leq r \leq j-1$;
- 3) $p_k \neq q_r$, 其中 $1 \leq k \leq i-1$ 且 $1 \leq r \leq j-1$;
- 4) $\{\pi_{p_1}, \pi_{p_2}, \dots, \pi_{p_i}\} \cup \{\pi_{q_1}, \pi_{q_2}, \dots, \pi_{q_j}\} = \{\pi_1, \pi_2, \dots, \pi_n\}$.

由双递增排列的定义可知, 其对应排列图中顶点, 要么入度为 0, 要么出度为 0, 要么入度、出度均为 0. (不难发现, 文献[12]中的双蘑菇结构实际上也满足双递增排列的定义.) 图 2 即为一个满足双递增排列定义的排列图. 在本文中, 当提到一个双递增排列时, 我们总是假设排列 π_A 中顶点的入度均为 0, π_B 中顶点的出度均为 0. 从而, 入度为 0 的顶点组成了一个递增排列, 出度为 0 的顶点组成了另一个递增排列. 为了找到基因组排列的短块移动排序次数的下界, 我们首先需要在输入的排列中寻找双递增子排列. 对一个双递增子排列进行排序, 必须要满足以下引理:

引理 1. 给定排列 $\pi = [\pi_1, \pi_2, \dots, \pi_n]$, 满足 $\pi_{i_1} < \pi_{i_2} < \dots < \pi_{i_l} < \pi_h$, 其中 $h < i_1 < i_2 < \dots < i_l$, 则当 $|j-k| \geq 2, k \neq j$ 且 $1 \leq j, k \leq l$ 时, 弧 (π_h, π_{i_j}) 与弧 (π_h, π_{i_k}) 不能形成 hop-match.

证明. 假设弧 (π_h, π_{i_j}) 与弧 (π_h, π_{i_k}) 能够形成 hop-match, 且 $r = k - j \geq 2$ ($r \leq -2$ 同理). 在进行 hop 操作之前考虑如排列 $\pi' = [\dots, \pi_h, \pi_{i_j}, \pi_{i_{j+r}}, \dots]$ 这种情况, 则 $\pi_{i_{j+1}}$ 或者位于 π_h 的左侧, 此时会产生一个新弧 $(\pi_{i_{j+1}}, \pi_j)$; 或者位于 $\pi_{i_{j+r}}$ 的右侧, 此时会产生一个新弧 $(\pi_{i_{j+r}}, \pi_{i_{j+1}})$. 上述 2 种情况均不满足定理 1, 假设不成立. 证毕.

推论 1. 给定排列 $\pi = [\pi_1, \pi_2, \dots, \pi_n]$, 满足 $\pi_h < \pi_{i_1} < \pi_{i_2} < \dots < \pi_{i_l}$, 其中 $i_1 < i_2 < \dots < i_l < h$, 则当 $|j-k| \geq 2, k \neq j$ 且 $1 \leq j, k \leq l$ 时, 弧 (π_{i_j}, π_h) 与弧 (π_{i_k}, π_h) 不能形成 hop-match.

在引理 1 和推论 1 中, 不难发现, 双递增排列中的每个符号均可作为 π_h . 现构建一个匹配图, 记为

G^M , 用来表示可能的 hop-matches. 排列图中的每条边对应 G^M 中的一个顶点, 若 G^M 中 2 个顶点所对应的边在排列图中构成了一个 hop-match, 则在这 2 个顶点之间添加 1 条边.

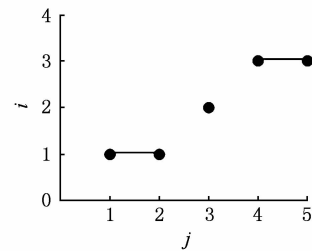
此时注意到匹配图中的最大匹配数目即为短块移动排序中 hop 操作次数的上界, 其余未匹配顶点则需要通过 skip 操作来消除.

普通排列的匹配图实际上就是排列图对应的边图, 但对于双递增排列, 因其根据引理 1 和推论 1 排除了一些边, 使得双递增排列的匹配图不同于传统意义上的边图. 在绘制一个双递增排列的匹配图时, 将 π_B 中顶点沿水平方向自左向右严格升序排列, π_A 中顶点沿垂直方向自上向下严格降序排列. 排列图中的一条边 (π_{p_i}, π_{q_j}) 对应于匹配图中一个顶点, 该顶点位于水平线 π_{p_i} 与垂直线 π_{q_j} 的交点处, 坐标记为 (j, i) .

由于构成 hop-match 的 2 个弧在排列图中具有相同的头或尾, 则匹配图中的边均为水平边或竖直边. 从而可得到点 (j, i) 和点 (j, k) 之间边的长度为 $|i-k|$. 同理, 点 (j, i) 和点 (k, i) 之间边的长度为 $|j-k|$. 根据引理 1 和推论 1, 我们可以进一步得到:

引理 2. 双递增排列对应的匹配图中, 每一条边要么为水平边, 要么为竖直边, 且每条边的长度均为 1.

对于给定双递增排列 $\pi = [3, 1, 2, 5, 4, 8, 6, 7, 9]$, 可分成 2 个不相交的递增排列 $\pi_A = [3, 5, 8, 9]$, $\pi_B = [1, 2, 4, 6, 7]$. 该排列对应排列图如图 2 所示, 其对应匹配图如图 4 所示:

Fig. 4 The match graph for $\pi = [3, 1, 2, 5, 4, 8, 6, 7, 9]$.图4 $\pi = [3, 1, 2, 5, 4, 8, 6, 7, 9]$ 的匹配图

引理 3. 双递增排列的匹配图是一个 2 分图.

证明. 给定一个双递增排列 π , 假设 π 的匹配图 G^M 中存在一个环 $C = (c_1, c_2, \dots, c_k)$, 其中 $k \geq 3$ 且 c_i 的坐标为 (x_i, y_i) ($1 \leq i \leq k$). 对于环 C 中任意一对不同顶点 c_i 和 c_j , 沿环 C 存在 2 条由 c_i 到 c_j 的路径. 其中 1 条路径包含 $|x_i - x_j| + 2r_1$ 条水平边和 $|y_i - y_j| + 2r_2$ 条竖直边, 另 1 条路径包含 $|x_i - x_j| +$

$2r_3$ 条水平边和 $|y_i - y_j| + 2r_4$ 条竖直边, 其中 r_i 为自然数. 因此, 环 C 中总的边数为 $2|x_i - x_j| + 2|y_i - y_j| + 2r_1 + 2r_2 + 2r_3 + 2r_4$, 其为偶数. 因此, 一个双递增排列 π 的匹配图中不存在奇环, 即该匹配图为一个 2 分图. 证毕.

给定双递增排列 π 的匹配图 G_n^M , 通过广度优先搜索, 可将 G_n^M 表示为 (L_n^M, R_n^M, E_n^M) . 其中, 定义 L_n^M 为广度优先搜索后得到左侧点集, R_n^M 为广度优先搜索后得到右侧点集, E_n^M 为边集.

引理 4^[13]. 2 分图 $G=(L, R, E)$ 存在一个 L -饱和匹配, 当且仅当对于所有 $X \subseteq L$, 有 $|N(X)| \geq |X|$.

引理 4 中, 定义 X 为集合 L 的任意子集, $N(X)$ 为集合 X 的邻域点集. 因为匹配图中的每个未匹配顶点对应一次 skip 操作, 则匹配图中的未匹配顶点数即为短块移动排序问题中所需 skip 操作次数的下界.

引理 5. 双递增排列 π 的匹配图 $G_n^M=(L_n^M, R_n^M, E_n^M)$, 对于任意 $X \subseteq L_n^M$, 若满足 $|N(X)| < |X|$, 则在对排列 π 进行短块移动排序过程中, 至少需要 $(|X| - |N(X)|)$ 次 skip 操作来移除这 $|X|$ 个弧.

证明. 因为 $|N(X)| < |X|$, 则匹配图中存在至少 $(|X| - |N(X)|)$ 个顶点为未匹配顶点, 每个这样的顶点对应一次 skip 操作, 则至少需要 $(|X| - |N(X)|)$ 次 skip 操作来移除这 $|X|$ 个弧. 引理 5 实际上是对引理 4 的一个简单应用. 证毕.

给定一个普通排列, 其最大双递增子排列 (maximal double increasing sub-permutation, MDP) 是一个双递增子排列, 且满足不被其他任意双递增子排列所包含. 因为对于任意最大双递增子排列, 满足第一个符号的入度为 0 且最末一个符号的出度为 0, 其余符号或入度为 0 或出度为 0 或入度、出度均为 0. 如排列 $\pi=[5, 1, 6, 2, 10, 4, 3, 11, 12, 7, 8, 9]$, 包含 2 个最大双递增子排列 $\pi_A=[5, 1, 6, 2]$, $\pi_B=[11, 12, 7, 8, 9]$. 由此, 为了找到排列的所有最大双递增子排列, 我们可从左至右对原排列进行扫描. 对于这些双递增子排列, 匹配图中的某些顶点可能会与该匹配图之外的顶点相匹配, 此时对匹配图中的这些顶点进行标记.

引理 6. 排列 π 的一个最大双递增子排列对应的匹配图中, 对于任意 $X \subseteq L_n^M$ (其中集合 X 中的顶点均为未被标记顶点). 若满足 $|N(X)| < |X|$, 则在对排列 π 进行短块移动排序过程中, 至少需要 $(|X| - |N(X)|)$ 次 skip 操作来移除这 $|X|$ 个弧.

证明. 因匹配图中的未被标记顶点, 或与图中另一顶点匹配, 或并不存在与其余顶点间的匹配, 则引理 6 成立. 证毕.

我们将引理 6 中的 skip 称为必要 skip.

定理 2. 给定一个普通排列 π , 设 $D_1, D_2, \dots, D_r$ 为 π 的最大双递增子排列, s_i 为 D_i 中的必要 skip 数目, 则在对 π 进行排序的最优解中, 至少存在 $\sum_{i=1}^r s_i$ 次 skip 操作.

证明. 由最大双递增子排列的定义可知, 排列 π 中的所有最大双递增子排列均互斥, 结合引理 6 可知, 该定理成立. 证毕.

3 结束语

相比 Heath 和 Vergara 给出的短块移动排序次数的下界, 本文通过分析原始排列中所有双递增子排列, 给出该问题一个更紧的新下界. 证明了在短块移动排序距离问题中, 其最优解中至少存在 $\sum_{i=1}^r s_i$ 次 skip 操作.

在证明该新下界存在的基础上, 如何对现有算法进行改进以提高其近似度, 从而设计出更好的近似算法, 将作为下一步研究内容的重点.

参 考 文 献

[1] Hoot S B, Palmer J D. Structural rearrangements, including parallel inversions within the chloroplast genome of anemone and related genera [J]. Journal of Molecular Evolution, 1994, 38(3): 274-281

[2] Palmer J D, Herbon L A. Tricircular mitochondrial genomes of brassica and raphanus: Reversal of repeat configurations by inversion [J]. Nucleic Acids Research, 1986, 14 (24): 9755-9764

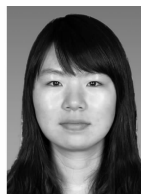
[3] Sankoff D, Leduc G, Antoine N, et al. Gene order comparisons for phylogenetic inference: Evolution of the mitochondrial genome [J]. Proc of the National Academy of Sciences of the United States of America, 1992, 89(14): 6575-6579

[4] Bafna V, Pevzner P. Sorting by transpositions [J]. SIAM Journal on Discrete Mathematics, 1998, 11(2): 224-240

[5] Elias I, Hartman T. A 1.375-approximation algorithm for sorting by transpositions [J]. IEEE/ACM Trans on Computational Biology Bioinformatics, 2006, 3(4): 369-379

[6] Eriksson H, Eriksson K, Karlander J, et al. Sorting a bridge hand [J]. Discrete Mathematics, 2001, 241(1/2/3): 289-300

- [7] Bulteau L, Fertin G, Rusu I. Sorting by transpositions is difficult [C] //Proc of the 38th Int Colloquium on Automata, Languages and Programming (ICALP). Berlin: Springer, 2011: 654-665
- [8] Heath L S, Vergara J P C. Sorting by bounded block-moves [J]. Discrete Applied Mathematics, 1998, 88(1/2/3): 181-206
- [9] Mahajan M, Rama R, Vijayakumar S. On sorting by 3-bounded transpositions [J]. Discrete Mathematics, 2006, 306(14): 1569-1585
- [10] Heath L S, Vergara J P C. Sorting by short block moves [J]. Algorithmica, 2000, 28(3): 323-354
- [11] Jiang Haitao, Zhu Daming, Zhu Binhai. A $(1 + \epsilon)$ -approximation algorithm for sorting by short block-moves [J]. Theoretical Computer Science, 2012, 439: 1-8
- [12] Jiang Haitao, Zhu Daming. A 14/11-approximation algorithm for sorting by short block-moves [J]. SCIENCE CHINA Information Sciences, 2011, 41(1): 33-47 (in Chinese)
(姜海涛, 朱大铭. 短块移动排序的 14/11 近似算法[J]. 中国科学: 信息科学, 2011, 41(1): 33-47)
- [13] Hall P. On representatives of subsets [J]. Journal London Mathematical Society, 1935, 10(1): 26-30



Wang Tong, born in 1989. Received her master degree from the School of Computer Science and Technology, Shandong University in July, 2015. Her main research interests include algorithm design analysis and bioinformatics.



Jiang Haitao, born in 1982. Co-PhD, associate professor. Member of China Computer Federation. His main research interests include algorithm design analysis, bioinformatics, and computational biology.



Zhu Daming, born in 1964. Professor and PhD supervisor. Member of China Computer Federation. His main research interests include algorithms, bioinformatics, and computational biology (dmzhu@sdu.edu.cn).

《信息安全研究》期刊简介

习近平总书记指出“没有网络安全就没有国家安全,没有信息化就没有现代化”。数字时代信息安全工具的大众化是不可阻挡的历史潮流。大众化的信息安全已经直接影响到我们每个人的利益,信息安全已成为国家、地方区域经济结构优化提升和转型发展的新机遇。在信息安全上升为国家战略、行业迎来崭新发展机遇形势下,《信息安全研究》期刊应时代而生。

《信息安全研究》是由国家发改委主管、国家信息中心主办的中文学术期刊,其宗旨是集中展示和报道国际、国内网络和信息安全研究领域研究成果及最新应用,传播信息安全基础理论和技术策略,服务国家信息安全形势发展需要。所刊登的论文均经过专家严格评审。

《信息安全研究》将于 2015 年 10 月创刊发行。刊期为月刊,每期 96 页,由《信息安全研究》杂志社出版,国内外公开发行。

《信息安全研究》将以研究致以应用,搭建信息安全领域的学术交流平台,愿意和同行业及社会各界建立联系,友好合作,共赢美好未来。欢迎大家积极投稿、赐稿,洽谈合作。

投稿邮箱:ris@cei.gov.cn

编辑部联系人:崔先生(185 0008 6481)

马先生(158 1058 2450)