O Problema da Ordenação de Permutações por Reversões e Transposições

Andre Rodrigues Oliveira

Instituto de Computação
Universidade Estadual de Campinas
Orientador: Zanoni Dias

Agenda

Motivação

- Evolução
  - Mudança das características hereditárias.
  - Resultado de mutações ocorridas no material genético.
- Mutações pontuais
  - Afetam bases individuais do DNA.
  - Ocorrem com maior frequência.
- Rearranjos de genomas
  - Mutações que afetam grandes trechos do DNA.
  - Comuns em plantas, mamíferos, vírus e bactérias.
- Comparação entre genomas.
  - Distância de edição x Distância de rearranjo.

Conceitos

- Um genoma é representado como uma *n*-tupla cujos elementos representam os genes.
- Supondo que não haja repetição de genes, esta *n*-tupla é uma permutação $\pi = (\pi_1\ \pi_2\ ...\ \pi_n)$, com $\pi_i \in \{-n, -(n-1), ...,$ $-2, -1, +1, +2, ..., +(n-1), +n\}$ e $|\pi_i| \neq |\pi_j| \Leftrightarrow i \neq j$.
- Cada elemento π_i possui um sinal, $+$ ou $-$, que indica a orientação do gene que ele representa.

Conceitos

- A *permutação identidade* é a permutação ordenada e é denotada por $\iota = (1\ 2\ \dots\ n)$.
  - Para permutações com sinal, $\iota = (+1\ \ +2\ \ \dots\ \ +n)$.
- A *permutação estendida* é obtida a partir de π inserindo-se dois novos elementos: $\pi_0 = 0$ e $\pi_{n+1} = n + 1$.

Eventos de Mutação

Para efeito de comparação entre dois genomas, consideramos apenas eventos conservativos. Os principais eventos de mutação são:

- Reversão
- Transposição
- Transposição reversa
- Fissão
- Fusão

Reversão

- Um *evento de reversão* $\rho(i,j)$, com $1 \leq i \leq j \leq n$, reverte a ordem e a orientação de $\pi[i..j]$, ou seja:

$$\pi_1 \quad \pi_2 \quad \dots \quad \pi_i \quad \pi_{i+1} \quad \dots \quad \pi_{j-1} \quad \pi_j \quad \dots \quad \pi_{n-1} \quad \pi_n$$

$$\pi_1 \quad \pi_2 \quad \dots \quad -\pi_j \quad -\pi_{j-1} \quad \dots \quad -\pi_{i+1} \quad -\pi_i \quad \dots \quad \pi_{n-1} \quad \pi_n$$

Reversão

- *Distância de reversão:* dadas as permutações π e σ, calcular a distância de reversão, $d_r(\pi, \sigma)$, consiste em obter uma sequência de k reversões $\rho_1, ..., \rho_k$, de tamanho mínimo, tal que $d_r(\pi, \sigma) = k$ e $\pi \cdot \rho_1 \cdot ... \cdot \rho_k = \sigma$.
- *Ordenação por reversões*: dada π, consiste em calcular a distância de ordenação por reversões, $d_r(\pi)$, entre π e ι, ou seja, $d_r(\pi) = d_r(\pi, \iota)$.

Transposição

- Um *evento de transposição* $\tau(i, j, k)$, com $1 \leq i < j < k \leq n + 1$, troca os blocos $\pi[i..j-1]$ e $\pi[j..k-1]$ de lugar, ou seja:

$$\pi_1 \quad \dots \quad \pi_{i-1} \quad \pi_i \quad \pi_{i+1} \quad \dots \quad \pi_{j-1} \quad \pi_j \quad \pi_{j+1} \quad \dots \quad \pi_{k-1} \quad \pi_k \quad \dots \quad \pi_n$$

$$\pi_1 \quad \dots \quad \pi_{i-1} \quad \pi_j \quad \pi_{j+1} \quad \dots \quad \pi_{k-1} \quad \pi_i \quad \pi_{i+1} \quad \dots \quad \pi_{j-1} \quad \pi_k \quad \dots \quad \pi_n$$

Transposição

- *Distância de transposição:* dadas π e σ, calcular a distância de transposição, $d_t(\pi, \sigma)$, consiste em obter uma sequência de k transposições $\tau_1, ..., \tau_k$ de tamanho mínimo, tal que $d_t(\pi, \sigma) = k$ e $\pi \cdot \tau_1 \cdot ... \cdot \tau_k = \sigma$.
- *Ordenação por transposições*: dada π, consiste em calcular a distância de ordenação por transposições, $d_t(\pi)$, entre π e ι, ou seja, $d_t(\pi) = d_t(\pi, \iota)$.

Modelo de Rearranjo

- Um *modelo de rearranjo* β determina o conjunto de operações permitidas para transformar um genoma em outro.
- A distância de ordenação, $d_\beta(\pi)$, é específica para cada permutação.
- A maior distância de ordenação de uma permutação de tamanho n é o *diâmetro* da distância de rearranjo, denotada por $D_\beta(n)$.

Breakpoints

- Um breakpoint para o problema de ordenação por reversões sem sinal ocorre entre um par de elementos π_i e π_{i+1} de π, com $0 \leq i \leq n$, se $|\pi_i \ - \ \pi_{i+1}| \ \neq \ 1$.
- O número de breakpoints de π é denotado por $b_r(\pi)$.
- $b_r(\pi) = 0 \Leftrightarrow \pi = \iota$.
- Exemplos:
  - $\pi = (0 \bullet 2 \ \ 3 \bullet 1 \bullet 4 \bullet 6 \ \ 5 \bullet 7)$ e $b_r(\pi) = 5$.
  - $\sigma = (0 \bullet 3 \ \ 4 \ \ 5 \bullet 2 \ \ 1 \bullet 6)$ e $b_r(\sigma) = 3$.

Breakpoints

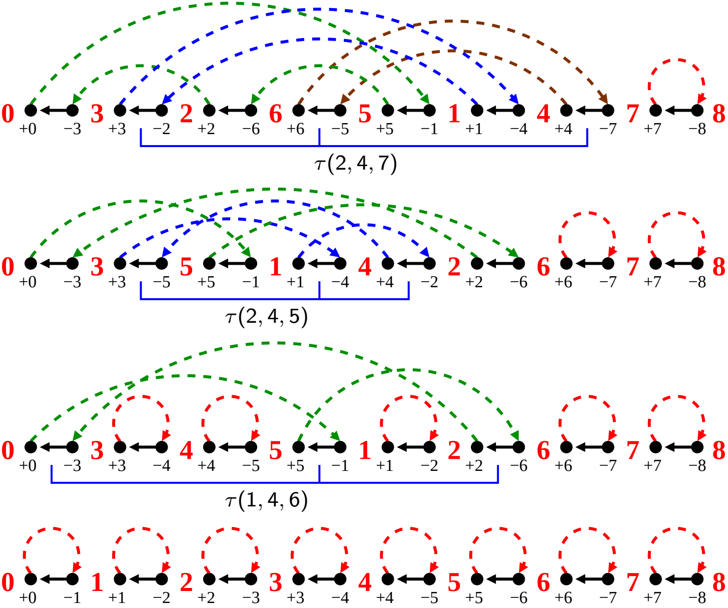
- Um breakpoint para o problema de ordenação por reversões com sinal e para o problema de ordenação por transposições ocorre entre um par de elementos π_i e π_{i+1}, com $0 \leq i \leq n$, se $\pi_{i+1} - \pi_i \neq 1$.
- O número de breakpoints de π é denotado por $b_{\overline{r}}(\pi)$ e $b_t(\pi)$, respectivamente.
- $b_{\overline{r}}(\pi) = b_t(\pi) = 0 \Leftrightarrow \pi = \iota$.
- Exemplos:
  - $\pi = (0 \bullet -1 \bullet +2 \;\; +3 \bullet -5 \;\; -4 \bullet +6 \;\; +7)$ e $b_{\overline{r}}(\pi) = 4$.
  - $\sigma = (0 \bullet 3 \;\; 4 \;\; 5 \bullet 2 \bullet 1 \bullet 6)$ e $b_t(\sigma) = 4$.

Grafo de Ciclos

- O grafo de ciclos de uma permutação π é um grafo de arestas coloridas $G(\pi)$, onde:
  - O conjunto de vértices é dado por $V = \{+0, -\pi_1, +\pi_1, ..., -\pi_n, +\pi_n, -(n+1)\}$.
  - O conjunto de arestas é dado por $E = E_P \cup E_C$, onde E_P é o conjunto de arestas pretas e E_C é o conjunto de arestas cinzas.
  - $E_P = \{(-\pi_i, +\pi_{i-1})$, para $1 \leq i \leq n+1\}$.
  - $E_C = \{(+(i-1), -i)$, para $1 \leq i \leq n+1\}$.

Grafo de Ciclos

- Denota-se por $c(\pi)$ o número de ciclos alternados de $G(\pi)$.
- Seja $c_{impar}(\pi)$ o número de ciclos alternados com número ímpar de arestas pretas de $G(\pi)$.
- Seja $c_{2+}(\pi)$ o número de ciclos alternados com duas ou mais arestas pretas de $G(\pi)$.
- A permutação ι é a única cujo grafo de ciclos possui $n+1$ ciclos, sendo todos ímpares e com apenas uma aresta preta cada ($c(\iota) = c_{impar}(\iota) = n+1$ e $c_{2+}(\iota) = 0$).


0 3 2 6 5 1 4 7 8
+0 −3 +3 −2 +2 −6 +6 −5 +5 −1 +1 −4 +4 −7 +7 −8
$\tau(2,4,7)$
0 3 5 1 4 2 6 7 8
+0 −3 +3 −5 +5 −1 +1 −4 +4 −2 +2 −6 +6 −7 +7 −8
$\tau(2,4,5)$
0 3 4 5 1 2 6 7 8
+0 −3 +3 −4 +4 −5 +5 −1 +1 −2 +2 −6 +6 −7 +7 −8
$\tau(1,4,6)$
0 1 2 3 4 5 6 7 8
+0 −1 +1 −2 +2 −3 +3 −4 +4 −5 +5 −6 +6 −7 +7 −8

O Problema de Ordenação por Reversões sem Sinal

- Em 1996, Bafna e Pevzner apresentaram os primeiros estudos sobre este problema, que resultou em um algoritmo de aproximação de fator 1.75.
- Em 1997, Caprara provou que este problema é NP-Difícil.
- Atualmente, o melhor algoritmo conhecido para o problema possui fator de aproximação 1.375, proposto por Berman e coautores em 2002.

O Problema de Ordenação por Reversões com Sinal

- O primeiro algoritmo polinomial exato para este problema foi apresentado em 1995 por Hannenhalli e Pevzner, com complexidade $O(n^4)$.
- Atualmente, o algoritmo mais eficiente disponível possui complexidade $O(n\sqrt{n \log n})$, apresentado em 2004 por Tannier e Sagot.
- Há também um algoritmo linear que retorna apenas o valor da distância de reversão de permutações, apresentado em 2001 por Bader, Moret e Yan.

O Problema de Ordenação por Transposições

- O primeiro algoritmo de aproximação foi apresentado por Bafna e Pevzner em 1998, e possui fator de aproximação 1.5.
- Em 2006, Elias e Hartman desenvolveram um algoritmo de aproximação com fator de 1.375.
- Em 2011, Bulteau, Fertin e Rusu provaram que o problema da distância de transposição é NP-Difícil.

O Problema de Ordenação por Reversões e Transposições

- Este problema permite que os eventos de reversão e os eventos de transposição ocorram durante a ordenação de uma permutação qualquer π.
- Existem duas versões para este problema, quando considera-se permutações com sinal e quando considera-se permutações sem sinal.
- A complexidade de ambas as versões é desconhecida.

O Problema de Ordenação por Reversões e Transposições

Para o caso de permutações com sinal:

- Em 1998 Walter, Dias e Meidanis apresentaram um algoritmo de aproximação com razão 2 e limitantes para o diâmetro.
- Em 1999, Gu, Peng e Sudborough acrescentaram o evento de transversão ao problema, e apresentaram um algoritmo de aproximação na razão 2 quando as três operações são permitidas.
- Em 2001, Lin e Xue apresentaram um algoritmo com fator de aproximação 1.75 quando, além das três operações citadas anteriormente, é adicionada também a operação *RevRev*.
  - Em 2005, Hartman e Sharan melhoraram esta razão para 1.5.

O Problema de Ordenação por Reversões e Transposições

Para o caso de permutações sem sinal:

- Em 1998, Walter, Dias e Meidanis apresentaram um algoritmo de aproximação com razão 3.
- Em 2008, Rahman, Shatabda e Hasan apresentaram limitantes para a distância e um algoritmo de aproximação com fator $2k$, onde k é o fator de aproximação do algoritmo utilizado para a decomposição de ciclos.
  - Dado o melhor valor de k conhecido, o fator de aproximação deste algoritmo torna-se $2.8386 + \epsilon$, para algum $\epsilon > 0$.

Limitantes

Para o Problema da Ordenação por Reversões sem sinal:

- $d_r(\pi) \geq \frac{b_r(\pi)}{2}$ (Bafna e Pevzner, 1996).
- $d_r(\pi) \leq b_r(\pi) - 1$ (Kececioglu e Sankoff, 1995).
- $D_r(n) = n - 1$ (Bafna e Pevzner, 1996).

Limitantes

Para o Problema da Ordenação por Transposições:

- $\frac{n+1-c_{impar}(G(\pi))}{2} \leq d_t(\pi) \leq \frac{3(n+1-c_{impar}(G(\pi)))}{4}$ (Bafna e Pevzner, 1998).
- $\lfloor \frac{n+1}{2} \rfloor \leq D_t(n)$ (Bafna e Pevzner, 1998).
- $D_t(n) \leq \lfloor \frac{2n-2}{3} \rfloor$ (Eriksson e coautores, 2001).

Limitantes

Para o Problema da Ordenação por Reversões e Transposições, com e sem sinal:

- $d_{rt}(\pi) \geq \frac{b(\pi) - c_{2+}(\pi)}{2}$ (Rahman Shatabda e Hasan, 2007).
- $\lfloor \frac{n}{2} \rfloor + 2 \leq D_{rt}(n)$ (Walter, Dias e Meidanis, 1998).

Conjecturas

- Uma família é um grupo de permutações que compartilham uma mesma regra de definição.
- Para estudar melhor os problemas, definimos uma família de permutações com sinal e uma família de permutações sem sinal.
- Para cada família, conjecturamos a distância exata correspondente.
- Um algoritmo de ordenação específico foi elaborado para cada família.

Conjecturas

Considere a seguinte família de permutações com sinal:

- $\pi_n^1 = (+n \ \ +(n-1) \ \ -(n-2) \ [trinca_1] \ [trinca_2] \ ... \ [trinca_k])$, com:
  - $k = \lceil \frac{n-3}{3} \rceil$;
  - $trinca_i = \ (+(3i-2) \ \ +3i \ \ -(3i-1)), \quad$ para $1 \leq i < k$;
  - $trinca_k = \begin{cases} (+(3k-2) \ \ +3k \ \ -(3k-1)), & \text{se } n \equiv 3 \bmod 3; \\ (+(3k-2) \ \ +(3k-1)), & \text{se } n \equiv 2 \bmod 3; \\ (+(3k-2)), & \text{se } n \equiv 1 \bmod 3. \end{cases}$

Exemplos:

- $\pi_{10}^1 = (+10 \ \ +9 \ \ -8 \ \ +1 \ \ +3 \ \ -2 \ \ +4 \ \ +6 \ \ -5 \ \ +7)$.
- $\pi_{11}^1 = (+11 \ \ +10 \ \ -9 \ \ +1 \ \ +3 \ \ -2 \ \ +4 \ \ +6 \ \ -5 \ \ +7 \ \ +8)$.
- $\pi_{12}^1 = (+12 \ \ +11 \ \ -10 \ \ +1 \ \ +3 \ \ -2 \ \ +4 \ \ +6 \ \ -5 \ \ +7 \ \ +9 \ \ -8)$.

Conjecturas

Lema

Para o Problema da Ordenação por Reversões e Transposições, para permutações com sinal, $d_{\bar{r}t}(\pi_n^1) \leq n - \lfloor \frac{n-2}{3} \rfloor$*, para* $n \geq 3$.

Conjectura

Para o Problema da Ordenação por Reversões e Transposições, para permutações com sinal, $d_{\bar{r}t}(\pi_n^1) = n - \lfloor \frac{n-2}{3} \rfloor$*, para* $n \geq 3$.

A conjectura acima é verdadeira para $n \leq 11$.

Conjectura

Para o Problema da Ordenação por Reversões e Transposições, para permutações com sinal, $D(n) = d_{\bar{r}t}(\pi_n^1) = n - \lfloor \frac{n-2}{3} \rfloor$*, para* $n \geq 3$.

A conjectura acima é verdadeira para $n \leq 10$.

Conjecturas

Considere a seguinte família de permutações sem sinal:

- $\pi_n^2 = \begin{cases} (n\ (n-2)\ \ldots\ 2\ (n-1)\ (n-3)\ \ldots\ 1), & \text{se } n \text{ for par;} \\ (n\ (n-2)\ \ldots\ 1\ (n-1)\ (n-3)\ \ldots\ 2), & \text{se } n \text{ for ímpar.} \end{cases}$

Exemplos:

- $\pi_{10}^2 = (10\ \ 8\ \ 6\ \ 4\ \ 2\ \ 9\ \ 7\ \ 5\ \ 3\ \ 1)$.
- $\pi_{11}^2 = (11\ \ 9\ \ 7\ \ 5\ \ 3\ \ 1\ \ 10\ \ 8\ \ 6\ \ 4\ \ 2)$.

Conjecturas

Lema

Para o Problema da Ordenação por Reversões e Transposições, considerando permutações sem sinal, $d_{rt}(\pi_n^2) \leq \lceil \frac{n}{2} \rceil$*, para* $n \geq 4$*.*

Conjectura

Para o Problema da Ordenação por Reversões e Transposições, considerando permutações sem sinal, $d_{rt}(\pi_n^2) = \lceil \frac{n}{2} \rceil$*, para* $n \geq 4$*.*

Conjectura

Para o Problema da Ordenação por Reversões e Transposições, considerando permutações sem sinal, $D(n) = d_{rt}(\pi_n^2) = \lceil \frac{n}{2} \rceil$*, para* $n \geq 4$*.*

As conjecturas acima são verdadeiras para $n \leq 13$.

Objetivos

Para o Problema da Ordenação de Permutações por Reversões e Transposições, considerando permutações com sinal e permutações sem sinal, pretendemos:

- Estudar classes de permutações que podem ajudar a entender melhor o problema.
- Melhorar os limitantes existentes.
- Obter algoritmos de aproximação com fatores melhores que os existentes.
- Provar as conjecturas propostas.

Metodologia

- O trabalho a ser realizado é, em grande parte, teórico.
- O estudo de teoremas e provas dos trabalhos já existentes na literatura são a principal base para o estabelecimento dos limitantes e das classes de permutações específicas.
- Sempre que possível e necessário, no entanto, programas serão desenvolvidos para comprovar a validade dos resultados obtidos.

Cronograma

	2013					2014												2015						
	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J
1	*	*	*	*				*	*	*	*													
2		*	*	*	*	*	*	*	*				*	*				*	*	*				
3					*	*	*																	
4								*																
5								*	*	*	*	*	*											
6													*	*	*	*	*	*						
7																		*	*	*	*	*	*	
8												*	*				*	*			*	*	*	
9																							*	
10																								*

1. Obtenção dos créditos obrigatórios em disciplinas do programa de mestrado;
2. Revisão Bibliográfica;
3. Escrita da proposta de mestrado;
4. Exame de qualificação do mestrado;
5. Investigação sobre limitantes, considerando permutações com sinal e sem sinal, para o valor do diâmetro;

Cronograma

	2013					2014												2015						
	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J
1	*	*	*	*				*	*	*	*													
2		*	*	*	*	*	*	*	*				*	*				*	*	*				
3					*	*	*																	
4								*																
5								*	*	*	*	*	*											
6													*	*	*	*	*	*						
7																		*	*	*	*	*	*	
8												*	*				*	*			*	*	*	
9																							*	
10																								*

6. Definição das classes de permutações com sinal e permutações sem sinal para as quais é possível determinar a distância exata;
7. Investigação de algoritmo com um melhor fator de aproximação, para as versões do problema que consideram permutações com e sem sinal;
8. Escrita da dissertação;
9. Revisão da dissertação;
10. Defesa da dissertação.

Obrigado!

Andre Rodrigues Oliveira

Instituto de Computação
Universidade Estadual de Campinas
Orientador: Zanoni Dias