Algoritmos para o Problema da Ordenação de Permutações por Reversões e Transposições

Andre Rodrigues Oliveira

Instituto de Computação
Universidade Estadual de Campinas
Orientador: Zanoni Dias
Coorientador: Ulisses 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.

Conceitos

- Um genoma é representado como uma n-tupla cujos elementos representam os genes (ou blocos de genes).
- Supondo que não haja repetição de genes, esta n-tupla é uma permutação $\pi = (\pi_1 \ \pi_2 \ \ldots \ \pi_n)$, com $\pi_i \in \{-n, -(n-1), \ldots, -2, -1, +1, +2, \ldots, +(n-1), +n\}$ 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.
  - Quando não há informação sobre a orientação dos genes, esse sinal é omitido, e a permutação é dita *sem sinais*.

Permutações

- Uma permutação pode representar tanto genomas lineares quanto genomas circulares.
  - Se π representa um genoma circular, temos, além das adjacências (π_i, π_{i+1}), para $1 \leq i < n$, a adjacência (π_n, π_1)
- A *permutação identidade* é a permutação ordenada e é denotada por $\iota = (1 \ 2 \ \dots \ n)$.
  - Na permutação identidade com sinais, todos os elementos possuem sinal positivo.
- 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 Rearranjo

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

- Reversão
- Transposição

Modelo de Rearranjo

- Um *modelo de rearranjo* β determina o conjunto de eventos permitidos para transformar um genoma em outro.
- A *distância* entre duas permutações π e σ com respeito a um modelo β, denotada por $d_\beta(\pi, \sigma)$, consiste em encontrar a menor sequência de operações $\rho_1, \rho_2, ..., \rho_k$ pertencentes a β tal que $\pi \cdot \rho_1 \cdot \rho_2 \cdot ... \cdot \rho_k = \sigma$, neste caso $d_\beta(\pi, \sigma) = k$.

Distância de Ordenação

- A distância de ordenação consiste em encontrar o menor conjunto de eventos permitidos que transformam π em ι, e é denotada por $d_\beta(\pi)$.
- Note que $d_\beta(\pi, \sigma) = d_\beta(\alpha)$ se tomarmos $\alpha = \sigma^{-1} \cdot \pi$, onde σ^{-1} é a inversa de σ tal que $\sigma^{-1} \cdot \sigma = \iota$, pois temos que $d_\beta(\pi, \sigma) = d_\beta(\sigma^{-1} \cdot \pi,\ \sigma^{-1} \cdot \sigma) = d_\beta(\alpha, \iota) = d_\beta(\alpha)$.
- A maior distância de ordenação entre todas as permutações de tamanho n é o *diâmetro* da distância de rearranjo, denotada por $D_\beta(n)$.

Reversão

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

$$(\pi_1 \quad \ldots \quad \pi_{i-1} \quad \color{red}{\pi_i} \quad \color{red}{\pi_{i+1}} \quad \color{red}{\ldots} \quad \color{red}{\pi_{j-1}} \quad \color{red}{\pi_j} \quad \pi_{j+1} \quad \ldots \quad \pi_n)$$

$$(\pi_1 \quad \ldots \quad \pi_{i-1} \quad \color{red}{-\pi_j} \quad \color{red}{-\pi_{j-1}} \quad \color{red}{\ldots} \quad \color{red}{-\pi_{i+1}} \quad \color{red}{-\pi_i} \quad \pi_{j+1} \quad \ldots \quad \pi_n)$$

Reversão Cíclica

- Uma *reversão cíclica* $\rho_r(i,j)$, age de forma semelhante à reversão, entretanto, não existe a restrição de que $i \leq j$ e, caso $i > j$, o bloco afetado por esta reversão será $[\pi_i..\pi_n\pi_1..\pi_j]$, ou seja:

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

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

Transposição

- Um *evento de transposição* $\rho_t(i, j, k)$, com $1 \leq i < j < k \leq n+1$, troca os blocos $[\pi_i..\pi_{j-1}]$ e $[\pi_j..\pi_{k-1}]$ de lugar, ou seja:

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

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

Transposição Cíclica

- Uma *transposição cíclica* $\rho_t(i, j, k)$, age de forma semelhante à transposição, entretanto, a restrição de que $i < j < k$ não é única, sendo permitidas também outras duas restrições:
- $k < i < j$, trocando os blocos $[\pi_i..\pi_{j-1}]$ e $[\pi_j..\pi_n\pi_1..\pi_{k-1}]$ de lugar:

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

- $j < k < i$, trocando os blocos $[\pi_i..\pi_n\pi_1..\pi_{j-1}]$ e $[\pi_j..\pi_{k-1}]$ de lugar:

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

Breakpoints

- Um breakpoint de reversão 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 reversão de π é denotado por $b_r(\pi)$.
- $b_r(\pi) = 0 \Leftrightarrow \pi = \iota$.
- Exemplos:
  - $\pi = (0 \cdot 2 \quad 3 \cdot 1 \cdot 4 \cdot 6 \quad 5 \cdot 7)$ e $b_r(\pi) = 5$.
  - $\sigma = (0 \cdot 3 \quad 4 \quad 5 \cdot 2 \quad 1 \cdot 6)$ e $b_r(\sigma) = 3$.

Breakpoints

- Um breakpoint de reversão com sinais ou de transposição 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 $b_{\bar{r}}(\pi)$ e $b_t(\pi)$, respectivamente.
- $b_{\bar{r}}(\pi) = b_t(\pi) = 0 \Leftrightarrow \pi = \iota$.
- Exemplos:
  - $\pi = (0 \cdot -1 \cdot +2 \quad +3 \cdot -5 \quad -4 \cdot +6 \quad +7)$ e $b_{\bar{r}}(\pi) = 4$.
  - $\sigma = (0 \cdot 3 \quad 4 \quad 5 \cdot 2 \cdot 1 \cdot 6)$ e $b_t(\sigma) = 4$.

Grafo de Ciclos

- O grafo de ciclos de uma permutação π é um grafo $G(\pi)$ 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

- Um *ciclo alternado* em $G(\pi)$ é tal que as cores de cada duas arestas consecutivas deste ciclo são sempre distintas.
- Denota-se por $c(\pi)$ o número de ciclos alternados de $G(\pi)$.
- Denota-se por $c_{impar}(\pi)$ o número de ciclos alternados com número ímpar de arestas pretas de $G(\pi)$.
- Um ciclo c é dito *curto*, caso possua no máximo 3 arestas pretas e *longo*, caso contrário.
  - Uma permutação π é dita *simples* se $G(\pi)$ possui apenas ciclos curtos.
- A permutação ι é a única cujo grafo de ciclos possui $n+1$ ciclos, sendo todos ímpares e com apenas uma aresta preta cada $(c(\iota) = c_{impar}(\iota) = n+1)$.

Grafo de Ciclos - Exemplo

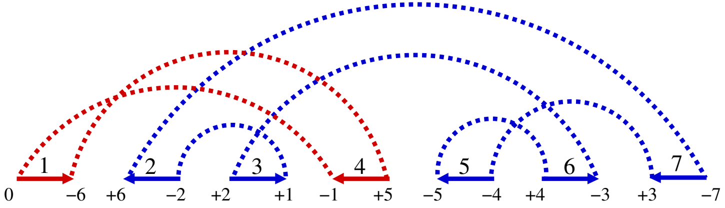

Figura: Grafo de ciclos da permutação com sinais $\pi = (+6\ +2\ -1\ -5\ +4\ +3)$, contendo os ciclos $c_1 = (7, 5, -6, -3, 2)$ e $c_2 = (4, -1)$. As arestas pretas estão representadas por retas horizontais e as arestas cinzas estão representadas por arcos pontilhados.

- $c(\pi) = 2$ e $c_{impar}(\pi) = 1$.
- Note que c_1 é um ciclo ímpar e c_2 é um ciclo par.
- Além disso, c_1 é um ciclo longo e c_2 é um ciclo curto.
- π não é uma permutação simples, uma vez que possui um ciclo longo.

Variações de Problemas de Ordenação por Rearranjos

- Pesquisas sugerem que as reversões que ocorrem durante a evolução tendem a não ser muito longas.
- Assim, uma das variações considera eventos ponderados pelo tamanho das operações.
- Uma reversão ponderada $\rho_r(i, j)$ possui peso $j - i + 1$.
- Uma transposição ponderada $\rho_t(i, j, k)$ possui peso $k - i$.

Problema da Ordenação por Rearranjos Ponderados

Dada uma permutação π, encontrar a sequência de eventos de rearranjo pertencentes a um modelo de rearranjo com menor peso que transforma π na permutação identidade ι.

Variações de Problemas de Ordenação por Rearranjos

- Uma reversão $\rho_r(i,j)$ é dita uma k-reversão se $k = j - i + 1$.
  - Uma k-reversão é dita curta se $k \leq 3$ e é dita super curta se $k \leq 2$.
- Uma transposição $\rho_t(i,j,k)$ é dita uma (x,y)-transposição se $x = j - i$ e $y = k - j$.
  - Uma (x,y)-transposição é dita curta se $x + y \leq 3$ e é dita super curta se $x + y = 2$.
- Pesquisas mostraram a prevalência de reversões curtas na evolução de alguns genomas bacterianos e eucariotos.

Problema da Ordenação por Rearranjos Curtos ou Super Curtos

Dada uma permutação π, encontrar a menor sequência de eventos de rearranjo curtos (ou super curtos) pertencentes a um modelo de rearranjo que transforma π na permutação identidade ι.

Estado da Arte dos Problemas de Ordenação por Rearranjos

Modelo de Rearranjo	**Complexidade**	**Melhor Solução**
Reversões com Sinais	Polinomial	Algoritmo exato $O(n^{\frac{3}{2}})$
Reversões sem Sinais	NP-Difícil	Algoritmo 1.375-aproximado
Transposições	NP-Difícil	Algoritmo 1.375-aproximado
Reversões e Transposições com Sinais	Desconhecida	Algoritmo 2-aproximado
Reversões e Transposições sem Sinais	Desconhecida	Algoritmo $2k$-aproximado
Reversões Curtas com Sinais	Desconhecida	Algoritmo 5-aproximado
Reversões Curtas sem Sinais	Desconhecida	Algoritmo 2-aproximado
Transposições Curtas	Desconhecida	Algoritmo 5/4-aproximado
Reversões e Transposições Curtas com Sinais	Desconhecida	Algoritmo 3-aproximado
Reversões e Transposições Curtas sem Sinais	Desconhecida	Algoritmo 2-aproximado
Reversões (ou Transposições) Super Curtas sem Sinais	Polinomial	Algoritmo exato $O(n^2)$
Reversões (ou Transposições) Super Curtas Circulares sem Sinais	Polinomial	Algoritmo exato $O(n^2)$
Reversões Super Curtas com Sinais	Polinomial	Algoritmo exato $O(n^3)$
Reversões Super Curtas Circulares com Sinais	Polinomial	Algoritmo exato $O(n^2)$
Reversões e Transposições Super Curtas com Sinais	Polinomial	Algoritmo exato $O(n^3)$
Reversões e Transposições Super Curtas Circulares com Sinais	Desconhecida	Inexistente

Objetivos

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

- Baixar os fatores de aproximação atuais.
- Criar algoritmos que retornem resultados melhores que os existentes na literatura.

Objetivos

Para o Problema da Ordenação de Permutações por Reversões e Transposições Curtas e Super Curtas:

- Encontrar um algoritmo polinomial exato para o Problema da Ordenação de Permutações Circulares com Sinais por Reversões e Transposições Super Curtas.
- Criar algoritmos com fatores de aproximação melhores que os existentes na literatura quando consideramos reversões e transposições curtas.

Objetivos

Para o Problema da Ordenação de Permutações por Reversões e Transposições Ponderadas:

- Verificar a existência de algoritmos aproximados para este problema, considerando permutações com sinais e sem sinais.
- Criar heurísticas que retornem bons resultados médios.

Objetivos

Para o Problema da Ordenação de Permutações por Reversões e Transposições considerando técnicas de Aprendizado de Máquina:

- Aplicar conceitos de classificação em aprendizado de máquina para ordenar permutações.
- Testar e comparar os diferentes métodos de classificação existentes na literatura.

Estágio no Exterior

- Início em Novembro/2016 pelo período de 1 ano na *Université de Nantes*, sob supervisão do Prof. Dr. Guillaume Fertin.
- Parte de um projeto de cooperação Capes/Cofecub entre pesquisadores brasileiros e franceses.
- Atividades previstas:
  - Trabalhar com as próximas etapas do Problema da Ordenação de Permutações de Reversões e Transposições.
  - Finalizar a prova do algoritmo polinomial para o Problema da Ordenação de Permutações Circulares com Sinais por Reversões e Transposições Super Curtas.

Cronograma

	2015					2016												2017						
	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								*	*	*	*													
12												*	*				*	*		*	*			
13			*	*				*	*				*	*				*	*				*	*

1. Obtenção dos créditos obrigatórios em disciplinas do programa.
2. Revisão Bibliográfica.
3. Programa de Estágio Docente (PED).
4. Escrita da proposta de doutorado.
5. Exame de Qualificação Específico (EQE).
6. Realização do Estágio no Exterior.

Cronograma

	2015					2016												2017						
	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								*	*	*	*													
12												*	*				*	*		*	*			
13			*	*				*	*				*	*				*	*				*	*

7. Investigação de algoritmo com um melhor fator de aproximação para as versões do Problema de Ordenação de Permutações por Reversões e Transposições que consideram permutações com sinais e sem sinais.
8. Investigação de algoritmo exato polinomial para o Problema da Ordenação de Permutações com Sinais Circulares por Reversões e Transposições Super Curtas.

Cronograma

	2015					2016												2017						
	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								*	*	*	*													
12												*	*				*	*		*	*			
13			*	*				*	*				*	*				*	*				*	*

9. Aplicação de técnicas de Aprendizado de Máquina em Ordenação de Permutações por Reversões e Transposições.
10. Investigação de heurísticas para o Problema da Ordenação de Permutações por Reversões e Transposições Ponderadas.
12. Escrita de artigos.
13. Escrita e revisão da tese.

Cronograma

	2017					2018												2019						
	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J
2		*	*				*	*				*	*				*	*		*	*			
6	*	*	*																					
9	*	*	*	*	*																			
10											*	*	*	*	*	*	*	*	*	*	*	*		
11					*	*	*	*	*	*	*	*	*	*	*	*								
12					*	*										*	*					*	*	
13			*	*				*	*				*	*			*	*		*	*	*	*	*
14																								*

2. Revisão Bibliográfica.
6. Realização do Estágio no Exterior.
9. Aplicação de técnicas de Aprendizado de Máquina em Ordenação de Permutações por Reversões e Transposições.
10. Investigação de heurísticas para o Problema da Ordenação de Permutações por Reversões e Transposições Ponderadas.

Cronograma

	2017					2018												2019						
	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J
2		*	*				*	*				*	*				*	*		*	*			
6	*	*	*																					
9	*	*	*	*	*																			
10											*	*	*	*	*	*	*	*	*	*	*	*		
11					*	*	*	*	*	*	*	*	*	*	*	*								
12					*	*										*	*					*	*	
13			*	*				*	*				*	*			*	*		*	*	*	*	*
14																								*

11. Investigação de algoritmo de aproximação com fatores melhores que os existentes para o Problema da Ordenação de Permutações por Reversões e Transposições Curtas.
12. Escrita de artigos.
13. Escrita e revisão da tese.
14. Defesa da tese.

Resultados Iniciais

Para o Problema da Ordenação de Permutações por Reversões e Transposições:

- Artigo submetido com uma heurística que utiliza os conceitos de breakpoints e grafo de ciclos cujos resultados médios são melhores que os existentes na literatura.

Resultados Iniciais

Para o Problema da Ordenação de Permutações por Reversões e Transposições:

- Exploração de um novo limitante para a criação dos bancos de permutações cujo fator de aproximação máximo até o momento é de $\frac{54}{29} \approx 1.8621$.
- O banco antigo possui 4.447.976 configurações, com aproximações com fator até 2.
- Novo banco possui 91.435.768 configurações, e falta encontrar uma sequência válida para 52.601 configurações.

Resultados Iniciais

Para o Problema da Ordenação de Permutações por Reversões e Transposições Ponderados:

- Relatório técnico publicado com um estudo inicial do problema e análise de resultados de heurísticas criadas.

Para o Problema da Ordenação de Permutações Circulares com Sinais por Reversões e Transposições Super Curtas:

- Esboço de prova que este problema admite um algoritmo polinomial exato.

Algoritmos para o Problema da Ordenação de Permutações por Reversões e Transposições

Andre Rodrigues Oliveira

Instituto de Computação
Universidade Estadual de Campinas
Orientador: Zanoni Dias
Coorientador: Ulisses Dias