Problemas de Ordenação de Permutações por Operações Ponderadas pelo Número de Fragmentações

Alexsandro Oliveira Alexandrino

Orientador: Prof. Dr. Zanoni Dias
Coorientadora: Dra. Carla Negri Lintzmayer

Universidade Estadual de Campinas

03 de Outubro de 2017

Roteiro

Introdução

- Processo Evolutivo
- Comparação entre genomas
- Rearranjos de genomas

Introdução

- Abordagem tradicional: sequência mínima de operações que transformam um genoma em outro.
- Abordagens ponderadas:
  - Pelo tamanho da operação
  - Pelo tipo de operação
  - Pelo número de fragmentações

Fundamentação Teórica

Representação de um genoma

Supondo que não haja repetição de genes, um genoma é representado por uma permutação $\pi = (\pi_1 \, \pi_2 \, \pi_3 \, \ldots \, \pi_n)$, onde $\pi_i \in \{ -n, -(n-1), \ldots, -2, -1, +1, +2, \ldots, +(n-1), +n \}$ tal que $|\pi_i| \neq |\pi_j| \iff i \neq j$.

Permutações com e sem sinais

- $(+1 \; +3 \; +2 \; +5 \; -4)$
- $(1 \; 5 \; 4 \; 3 \; 2)$

Fundamentação Teórica

Modelo de Rearranjo

- Conjunto de rearranjos permitidos para o cálculo da distância evolucionária;
- Denotado por M.

Rearranjos mais estudados na literatura

- Reversão
- Transposição

Permutação identidade

- $\iota_n = (1\ 2\ 3\ \cdots\ n)$
- $\bar{\iota}_n = (+1\ +2\ +3\ \cdots\ +n)$

Permutação reversa

- $\eta = (n\ \ldots\ 2\ 1)$
- $\overline{\eta} = (-n\ \ldots\ -2\ -1)$

Composição

A composição "$\cdot$" de duas permutações π e σ é a operação $\pi \cdot \sigma = (\pi_{\sigma_1} \ \pi_{\sigma_2} \ \ \ldots \ \pi_{\sigma_n})$.

Inversa

A inversa de uma permutação π é a permutação π^{-1}, tal que $\pi^{-1}_{\pi_i} = i$, para $1 \leq i \leq n$, e satisfaz $\pi \cdot \pi^{-1} = \pi^{-1} \cdot \pi = \iota_n$.

Ordenação de Permutações

Função de Custo

- $f : M \to \mathbb{R}$ é a função que denota o custo associado a cada operação de M

Distância

- Dados M, f, π e σ, a distância $c_M^f(\pi, \sigma)$ é o custo de uma sequência de operações $\beta_1, \beta_2, \ldots, \beta_m$ pertencentes a M, tal que $\pi \cdot \beta_1 \cdot \beta_2 \cdot \ldots$ $\ldots \cdot \beta_m = \sigma$ e $\sum_{i=1}^{m} f(\beta_i)$ é mínimo.
- O problema de transformar π em σ é equivalente ao de ordenar $\gamma = \sigma^{-1} \cdot \pi$, já que $c_M^f(\pi, \sigma) = c_M^f(\sigma^{-1} \cdot \pi, \sigma^{-1} \cdot \sigma) = c_M^f(\gamma, \iota_n) =$ $= c_M^f(\gamma)$.

Ordenação de Permutações

Diâmetro

O *diâmetro de ordenação* $D^f_M(n)$, com $n \in \mathbb{Z}^+$, é a maior distância de ordenação entre todas as permutações de tamanho n.

Definição

Uma reversão $\rho(i,j)$, com $1 \leq i < j \leq n$, inverte o segmento $\pi_i, \pi_{i+1}, \ldots, \pi_{j-1}, \pi_j$, ou seja:

$$\begin{aligned} \pi &= (\pi_1 \ \ldots \ \pi_{i-1} \ \pi_i \ \pi_{i+1} \ \ldots \ \pi_{j-1} \ \pi_j \ \pi_{j+1} \ \ldots \ \pi_n) \\ \pi \cdot \rho(i,j) &= (\pi_1 \ \ldots \ \pi_{i-1} \ \underline{\pi_j \ \pi_{j-1} \ \ldots \ \pi_{i+1} \ \pi_i} \ \pi_{j+1} \ldots \ \pi_n) \end{aligned}$$

Exemplo

$$(2\ 6\ \underline{5\ 4\ 1\ 7}\ 3) \cdot \rho(3,6) = (2\ 6\ \underline{7\ 1\ 4\ 5}\ 3)$$

Reversões com Sinais

Definição

Uma reversão $\overline{\rho}(i,j)$, com $1 \leq i \leq j \leq n$, inverte o segmento $\pi_i, \pi_{i+1}, \ldots, \pi_{j-1}, \pi_j$ e o sinal desses elementos, ou seja:

$$\pi = (\pi_1 \;\ldots\; \pi_{i-1} \; \pi_i \; \pi_{i+1} \;\ldots\; \pi_{j-1} \; \pi_j \; \pi_{j+1} \;\ldots\; \pi_n)$$

$$\pi \cdot \overline{\rho}(i,j) = (\pi_1 \;\ldots\; \pi_{i-1} \; \underline{-\pi_j \; -\pi_{j-1} \;\ldots\; -\pi_{i+1} \; -\pi_i} \; \pi_{j+1} \ldots \; \pi_n)$$

Exemplo

$$(+2 \; -6 \; \underline{-5 \; +4 \; -1} \; +3) \cdot \overline{\rho}(3,5) = (+2 \; -6 \; \underline{+1 \; -4 \; +5} \; +3)$$

Transposições

Definição

Uma transposição $\tau(i, j, k)$, com $1 \leq i < j < k \leq n+1$, troca de posição o segmento $\pi_i, \pi_{i+1}, \ldots, \pi_{j-1}$ com o segmento adjacente $\pi_j, \pi_{j+1}, \ldots,$ π_{k-1}, ou seja:

$$\begin{aligned} \pi &= (\pi_1\ \pi_2\ \pi_3\ \ldots\ \pi_{i-1}\ \underline{\pi_i\ \ldots\ \pi_{j-1}}\ \underline{\pi_j\ \ldots\ \pi_{k-1}}\ \pi_k\ \ldots\ \pi_n) \\ \pi \cdot \tau(i, j, k) &= (\pi_1\ \pi_2\ \pi_3\ \ldots\ \pi_{i-1}\ \underline{\pi_j\ \ldots\ \pi_{k-1}}\ \underline{\pi_i\ \ldots\ \pi_{j-1}}\ \pi_k\ \ldots\ \pi_n) \end{aligned}$$

Exemplo

$$(2\ \underline{6\ 5\ 4}\ \underline{1\ 7}\ 3) \cdot \tau(2, 5, 7) = (2\ \underline{1\ 7}\ \underline{6\ 5\ 4}\ 3)$$

Operações de Prefixo e Sufixo

Definição

- Operações de prefixo são aquelas que envolvem o primeiro elemento da permutação
- Operações de sufixo são aquelas que envolvem o último elemento da permutação

Problema de Ordenação de Permutações sem Sinais por Reversões

- Problema NP-Difícil [5].
- Algoritmo de 1.375-aproximação [2].

Exemplo

$\pi\ = (6\ 3\ 5\ 2\ 4\ 1)$
$\pi^1 = (6\ 3\ 5\ \underline{2\ 4}\ 1) \cdot \rho(4,5)$
$\pi^2 = (6\ 3\ \underline{5\ 4}\ 2\ 1) \cdot \rho(3,4)$
$\pi^3 = (6\ \underline{3\ 4\ 5}\ 2\ 1) \cdot \rho(2,4)$
$\pi^4 = (\underline{6\ 5\ 4\ 3\ 2\ 1}) \cdot \rho(1,6)$
$\pi^4 = (1\ 2\ 3\ 4\ 5\ 6) = \iota_6$

Problema de Ordenação de Permutações com Sinais por Reversões

- Problema polinomial [7].

Exemplo

$$
\begin{aligned}
\pi\ \ &= (+6\ +3\ +5\ +2\ +4\ +1) \\
\pi^1 &= (\underline{+6\ +3\ +5\ +2\ +4}\ +1) \cdot \rho(1,5) \\
\pi^2 &= (-4\ -2\ -5\ -3\ \underline{-6\ +1}) \cdot \rho(5,6) \\
\pi^3 &= (-4\ -2\ \underline{-5\ -3\ -1}\ +6) \cdot \rho(3,5) \\
\pi^4 &= (-4\ \underline{-2\ +1}\ +3\ +5\ +6) \cdot \rho(2,3) \\
\pi^5 &= (\underline{-4\ -1\ +2\ +3}\ +5\ +6) \cdot \rho(1,4) \\
\pi^6 &= (\underline{-3\ -2\ +1}\ +4\ +5\ +6) \cdot \rho(1,3) \\
\pi^7 &= (\underline{-1}\ +2\ +3\ +4\ +5\ +6) \cdot \rho(1,1) \\
\pi^7 &= (+1\ +2\ +3\ +4\ +5\ +6) = \bar{\iota}_6
\end{aligned}
$$

Problema de Ordenação de Permutações sem Sinais por Transposições

- Problema NP-Difícil [4].
- Algoritmo de 1.375-aproximação [6].

Exemplo

$$
\begin{aligned}
\pi \; &= (6\ 3\ 5\ 2\ 4\ 1) \\
\pi^1 &= (6\ 3\ \underline{5\ 2}\ \underline{4}\ 1) \cdot \tau(3,5,6) \\
\pi^2 &= (\underline{6}\ \underline{3\ 4\ 5}\ 2\ 1) \cdot \tau(1,2,5) \\
\pi^3 &= (\underline{3\ 4\ 5\ 6}\ \underline{2\ 1}) \cdot \tau(1,5,7) \\
\pi^4 &= (\underline{2}\ \underline{1}\ 3\ 4\ 5\ 6) \cdot \tau(1,2,3) \\
\pi^4 &= (1\ 2\ 3\ 4\ 5\ 6) = \iota_6
\end{aligned}
$$

Problema de Ordenação de Permutações por Reversões e Transposições

Problema de Ordenação de Permutações sem Sinais por Reversões e Transposições

- Complexidade desconhecida.
- Algoritmo de $2k$-aproximação [8], onde k é o fator de aproximação do algoritmo de decomposição de ciclos do grafo de breakpoints gerado a partir de π.

Problema de Ordenação de Permutações com Sinais por Reversões e Transposições

- Complexidade desconhecida.
- Algoritmo de 2-aproximação [9].

Operações Ponderadas pelo Tipo

Normalmente uma transposição recebe um custo maior que uma reversão, pois reversões ocorrem em uma proporção muito maior que transposições em cenários reais [1, 3].

Resultados Conhecidos

- Bader e Ohlebusch [1] consideraram uma nova operação chamada de transposição reversa.
- Uma transposição reversa troca dois segmentos adjacentes invertendo um deles.
- Para permutações com sinais e qualquer proporção de custos entre 1:1 e 2:1 (transposição:reversão), Bader e Ohlebusch [1] apresentam um algoritmo com fator de aproximação de 1.5, considerando que o peso de uma transposição reversa é igual ao de uma transposição.

Operações Ponderadas pelo Número de Fragmentações

- Nessa nova versão, o custo de uma operação é igual ao número de fragmentações que a operação causa na permutação.

Operações Ponderadas pelo Número de Fragmentações

$$\pi = (\pi_1 \ \ldots \ \pi_{i-1} \ \pi_i \ \pi_{i+1} \ \ldots \ \pi_{j-1} \ \pi_j \ \pi_{j+1} \ \ldots \ \pi_n) \tag{1}$$

$$\pi \cdot \rho(1, n) = (\underline{\pi_n \ \ldots \ \pi_{j+1} \ \pi_j \ \pi_{j-1} \ \ldots \ \pi_{i+1} \ \pi_i \ \pi_{i-1} \ \ldots \ \pi_1}) \tag{2}$$

$$\pi \cdot \rho(1, j) = (\underline{\pi_j \ \pi_{j-1} \ \ldots \ \pi_{i+1} \ \pi_i \ \pi_{i-1} \ \ldots \ \pi_1} \ \pi_{j+1} \ \ldots \ \pi_n) \tag{3}$$

$$\pi \cdot \rho(i, n) = (\pi_1 \ \ldots \ \pi_{i-1} \ \underline{\pi_n \ \ldots \ \pi_{j+1} \ \pi_j \ \pi_{j-1} \ \ldots \ \pi_{i+1} \ \pi_i}) \tag{4}$$

$$\pi \cdot \rho(i, j) = (\pi_1 \ \ldots \ \pi_{i-1} \ \underline{\pi_j \ \pi_{j-1} \ \ldots \ \pi_{i+1} \ \pi_i} \ \pi_{j+1} \ \ldots \ \pi_n) \tag{5}$$

Reversões

O custo de uma reversão $\rho(i, j)$ é:

- $f(\rho(i, j)) = 0$, se $i = 1$ e $j = n$
- $f(\rho(i, j)) = 1$, se $i = 1$ e $j < n$
- $f(\rho(i, j)) = 1$, se $i > 1$ e $j = n$
- $f(\rho(i, j)) = 2$, se $i > 1$ e $j < n$

Operações Ponderadas pelo Número de Fragmentações

$$
\begin{aligned}
\pi &= (\pi_1 \ \ldots \ \pi_{i-1} \ \pi_i \ \pi_{i+1} \ \ldots \ \pi_{j-1} \ \pi_j \ \pi_{j+1} \ \ldots \ \pi_{k-1} \ \pi_k \ \ldots \ \pi_n) && (6)\\
\pi \cdot \tau(1, j, n+1) &= (\underline{\pi_j \ \pi_{j+1} \ \ldots \ \pi_{k-1} \ \pi_k \ \ldots \ \pi_n} \ \underline{\pi_1 \ \ldots \ \pi_{i-1} \ \pi_i \ \pi_{i+1} \ \ldots \ \pi_{j-1}}) && (7)\\
\pi \cdot \tau(1, j, k) &= (\underline{\pi_j \ \pi_{j+1} \ \ldots \ \pi_{k-1}} \ \underline{\pi_1 \ \ldots \ \pi_{i-1} \ \pi_i \ \pi_{i+1} \ \ldots \ \pi_{j-1}} \ \pi_k \ \ldots \ \pi_n) && (8)\\
\pi \cdot \tau(i, j, n+1) &= (\pi_1 \ \ldots \ \pi_{i-1} \ \underline{\pi_j \ \pi_{j+1} \ \ldots \ \pi_{k-1} \ \pi_k \ \ldots \ \pi_n} \ \underline{\pi_i \ \pi_{i+1} \ \ldots \ \pi_{j-1}}) && (9)\\
\pi \cdot \tau(i, j, k) &= (\pi_1 \ \ldots \ \pi_{i-1} \ \underline{\pi_j \ \pi_{j+1} \ \ldots \ \pi_{k-1}} \ \underline{\pi_i \ \pi_{i+1} \ \ldots \ \pi_{j-1}} \ \pi_k \ \ldots \ \pi_n) && (10)
\end{aligned}
$$

Transposições

O custo de uma transposição $\tau(i, j, k)$ é:

- $f(\tau(i, j, k)) = 1$, se $i = 1$ e $k = n + 1$
- $f(\tau(i, j, k)) = 2$, se $i = 1$ e $k < n + 1$
- $f(\tau(i, j, k)) = 2$, se $i > 1$ e $k = n + 1$
- $f(\tau(i, j, k)) = 3$, se $i > 1$ e $k < n + 1$

Operações Ponderadas pelo Número de Fragmentações

Exemplo

Ordenação da permutação $\pi = (1\ 3\ 8\ 5\ 4\ 7\ 6\ 2)$ com reversões:

- Sequência ótima da versão não ponderada (Custo 7): $\rho(3,5)$, $\rho(5,7)$, $\rho(2,7)$ e $\rho(2,8)$
- Sequência ótima do problema de prefixo e sufixo (Custo 6): $\rho(1,7)$, $\rho(7,8)$, $\rho(5,8)$, $\rho(3,8)$, $\rho(1,6)$ e $\rho(4,8)$
- Sequência ótima do problema ponderado por fragmentações (Custo 5): $\rho(3,5)$, $\rho(1,7)$, $\rho(1,3)$, $\rho(7,8)$ e $\rho(1,8)$

Objetivos

Estudar limitantes e criar algoritmos de aproximação para os problemas:

- Ordenação de Permutações por Reversões Ponderadas pelo Número de Fragmentações (*OSRPF* e *OCRPF*);
- Ordenação de Permutações por Transposições Ponderadas pelo Número de Fragmentações (*OSTPF*);
- Ordenação de Permutações por Reversões e Transposições Ponderadas pelo Número de Fragmentações (*OSRTPF* e *OCRTPF*).

Metodologia

- Foco teórico;
- Estudo de teoremas e provas de trabalhos existentes na literatura;
- Adaptar algoritmos das versões não ponderadas;
- Propor algoritmos de aproximação para os problemas;
- Criação de experimentos para análise prática dos fatores de aproximação.

Cronograma

	2017										2018												2019	
	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F
1	*	*	*	*		*	*	*	*															
2	*	*	*	*	*	*				*				*				*						
3					*	*																		
4								*																
5													*	*	*	*								
6						*	*	*	*															
7										*	*	*	*											
8														*	*	*	*							
9																		*	*	*	*			
10								*	*			*	*			*	*			*	*	*		
11																							*	
12																								*

1. Obtenção dos créditos obrigatórios em disciplinas do programa de mestrado;
2. Revisão da literatura;
3. Escrita da proposta de mestrado;
4. Exame de Qualificação de Mestrado (EQM);
5. Participação do Programa de Estágio Docente (PED);
6. Adaptação de algoritmos existentes para os problemas estudados;

Cronograma

	2017										2018												2019	
	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F
1	*	*	*	*		*	*	*	*															
2	*	*	*	*	*	*				*				*				*						
3					*	*																		
4								*																
5													*	*	*	*								
6						*	*	*	*															
7										*	*	*	*											
8														*	*	*	*							
9																		*	*	*	*			
10								*	*			*	*			*	*			*	*	*		
11																							*	
12																								*

7. Investigação de novos limitantes para os problemas estudados;
8. Criação de heurísticas e execução de experimentos;
9. Investigação de melhores fatores de aproximação;
10. Escrita da dissertação;
11. Revisão da dissertação;
12. Defesa da dissertação.

Resultados Preliminares

Algoritmos Aproximados

- Relação entre abordagem não ponderada e ponderada por fragmentações
  - Todo algoritmo de k-aproximação para reversões é uma $2k$-aproximação para o problema de reversões ponderadas por fragmentações
  - Todo algoritmo de k-aproximação para transposições é uma $3k$-aproximação para o problema de transposições ponderadas por fragmentações
- Desenvolvimento de algoritmos de 2-aproximação para todos os 5 problemas
- Dois algoritmos para cada problema, sendo um deles guloso
- Um artigo com resultados preliminares será submetido ainda este ano para uma conferência

Referências I

[1] M. Bader and E. Ohlebusch. Sorting by Weighted Reversals, Transpositions, and Inverted Transpositions. *Journal of Computational Biology*, 14(5):615–636, 2007.

[2] P. Berman, S. Hannenhalli, and M. Karpinski. 1.375-Approximation Algorithm for Sorting by Reversals. In R. Möhring and R. Raman, editors, *Proceedings of the 10th Annual European Symposium on Algorithms (ESA'2002)*, volume 2461 of *Lecture Notes in Computer Science*, pages 200–210. Springer-Verlag Berlin Heidelberg New York, Berlin/Heidelberg, Germany, 2002.

[3] M. Blanchette, T. Kunisawa, and D. Sankoff. Parametric Genome Rearrangement. *Gene*, 172(1):GC11–GC17, 1996.

[4] L. Bulteau, G. Fertin, and I. Rusu. Sorting by Transpositions is Difficult. *SIAM Journal on Computing*, 26(3):1148–1180, 2012.

Referências II

[5] A. Caprara. Sorting Permutations by Reversals and Eulerian Cycle Decompositions. *SIAM Journal on Discrete Mathematics*, 12(1): 91–110, 1999.

[6] I. Elias and T. Hartman. A 1.375-Approximation Algorithm for Sorting by Transpositions. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 3(4):369–379, 2006.

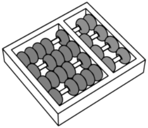
[7] S. Hannenhalli and P. A. Pevzner. Transforming Cabbage into Turnip: Polynomial Algorithm for Sorting Signed Permutations by Reversals. *Journal of the ACM*, 46(1):1–27, 1999.

[8] A. Rahman, S. Shatabda, and M. Hasan. An Approximation Algorithm for Sorting by Reversals and Transpositions. *Journal of Discrete Algorithms*, 6(3):449–457, 2008.

Referências III

[9] M. E. M. T. Walter, Z. Dias, and J. Meidanis. Reversal and Transposition Distance of Linear Chromosomes. In *Proceedings of the 5th International Symposium on String Processing and Information Retrieval (SPIRE'1998)*, pages 96–102, Los Alamitos, CA, USA, 1998. IEEE Computer Society.

Problemas de Ordenação de Permutações por Operações Ponderadas pelo Número de Fragmentações



Alexsandro Oliveira Alexandrino

Orientador: Prof. Dr. Zanoni Dias
Coorientadora: Dra. Carla Negri Lintzmayer

Universidade Estadual de Campinas

03 de Outubro de 2017