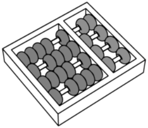
O Problema da Ordenação de Permutações por Operações de Tamanho Limitado



Guilherme Henrique Santos Miranda

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

Universidade Estadual de Campinas

03 de Outubro de 2017

Roteiro

Motivação

- Processo de Evolução;
- Comparação entre Genomas;
- Rearranjos de Genomas;
- Restrição com relevância biológica.

Fundamentação Teórica

Genoma: Representação Computacional

- $\pi = (\pi_1\, \pi_2\, \pi_3\, \ldots\, \pi_n)$, onde $\pi_i \in \{-n,\ -(n-1),\ \ldots,\ -2,\ -1,\ +1,\ +2,\ \ldots,\ +(n-1),\ +n\}$ e $|\pi_i| \neq |\pi_j| \iff i \neq j$;
- Exemplos:
  - Permutação com Sinais: $(+6\ -1\ -3\ -5\ +2\ +4)$;
  - Permutação sem Sinais: $(6\ 1\ 3\ 5\ 2\ 4)$.

Fundamentação Teórica

Modelo de Rearranjo

- Um modelo de rearranjo M determina as operações de rearranjo permitidas;
- Modelos utilizados: reversões e/ou transposições;
- Amplamente estudados na literatura.

Fundamentação Teórica

- Distância de Rearranjo ($d_M(\pi, \sigma)$): tamanho da menor sequência de operações ρ_i em M tais que $\pi \cdot \rho_1 \cdot \rho_2 \cdot \ \ldots \ \cdot \rho_t = \sigma$;
- Permutação identidade: $\iota = (1\ 2\ \ldots\ n)$;
- Distância de Ordenação: $d_M(\pi, \iota) = d_M(\pi) = t$;
- Diâmetro de Ordenação: denotado por $D_M(n)$, é igual ao valor máximo de $d_M(\pi)$ entre todas as permutações π de tamanho n.

Operações de Rearranjo

- Reversão: uma operação $\rho(i,j)$ com $1 \leq i < j \leq n$ tal que:

$$\pi \cdot \rho(i,j) = (\pi_1\ \pi_2 \ldots \pi_{i-1}\ \underline{\pi_j\ \pi_{j-1} \ldots \pi_{i+1}\ \pi_i}\ \pi_{j+1} \ldots \pi_{n-1}\ \pi_n);$$

- Se π é uma permutação *com sinais*, os sinais também são invertidos e temos $1 \leq i \leq j \leq n$;
- Distância de Ordenação por Reversões: $d_r(\pi)$ e $d_{\bar{r}}(\pi)$;
- Diâmetro de Ordenação por Reversões: $D_r(n)$ e $D_{\bar{r}}(n)$;
- O tamanho de uma reversão é dado por $j - i + 1$.

Exemplo

Dada a permutação com sinais $\pi = (+1\ +3\ +4\ -2\ -7\ -6\ -5)$, temos:

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

Neste caso, $d_{\bar{r}}(\pi) = 3$.

Operações de Rearranjo

- Transposição: uma operação $\tau(i,j,k)$ com $1 \leq i < j < k \leq n+1$ tal que:

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

- Distância de Ordenação por Transposições: $d_t(\pi)$;
- Diâmetro de Ordenação por Transposições: $D_t(n)$;
- O tamanho de uma transposição é dado por $k - i$.

Exemplo

Dada a permutação $\pi = (1\,3\,4\,2\,7\,6\,5)$, temos:

$$\begin{aligned}
\pi' &= \pi \cdot \tau(2,4,5) = (1\,\underline{2}\,\underline{3\,4}\,7\,6\,5) \\
\pi'' &= \pi' \cdot \tau(5,7,8) = (1\,2\,3\,4\,\underline{5}\,\underline{7\,6}) \\
\pi''' &= \pi'' \cdot \tau(6,7,8) = (1\,2\,3\,4\,5\,\underline{6}\,\underline{7}) = \iota
\end{aligned}$$

Neste caso, $d_t(\pi) = 3$.

Problemas de Ordenação

Ordenação de Permutações sem Sinais por Reversões

- Problema NP-Difícil [4];
- Primeiro resultado obtido por Bafna e Pevzner [1], com um algoritmo de 1.75-aproximação;
- Melhor resultado conhecido dado por Berman e coautores [3], com um algoritmo de 1.375-aproximação.

Ordenação de Permutações com Sinais por Reversões

- Solução em tempo polinomial, cuja primeira versão foi proposta por Hannenhalli e Pevzner [8];
- Melhor algoritmo dado por Tannier e coautores [14], com complexidade de tempo sub-quadrática.

Problemas de Ordenação

Ordenação de Permutações por Transposições

- Problema NP-Difícil [2];
- Primeiro algoritmo proposto por Bafna e Pevzner [2], com fator de aproximação 1.5;
- Melhor resultado conhecido dado por Elias e Hartman [6], com um algoritmo de 1.375-aproximação.

Problemas de Ordenação

Ordenação de Permutações por Reversões e Transposições

- Algoritmos de 3-aproximação e 2-aproximação para as versões sem e com sinais, respectivamente, dados por Walter e coautores [16];
- Algoritmo de $2k$-aproximação para permutações sem sinais, proposto por Rahman e coautores [13] (melhor fator de aproximação do algoritmo é $2.8386 + \epsilon$, para $\epsilon > 0$).

Exemplo

Dada a permutação sem sinais $\pi = (1\,3\,4\,2\,7\,6\,5)$, temos:

$$\begin{aligned} \pi' &= \pi \cdot \tau(2,4,5) = (1\,\underline{2}\,\underline{3\,4}\,7\,6\,5) \\ \pi'' &= \pi' \cdot \rho(5,7) = (1\,2\,3\,4\,\underline{5\,6\,7}) = \iota \end{aligned}$$

Neste caso, $d_{rt}(\pi) = 2$.

Restrições Adicionais

Operações de Tamanho Limitado

- Relevância biológica: operações que afetam grandes trechos ocorrem com pouca frequência [5, 12];
- Problema: transformar π em ι utilizando operações de tamanho no máximo λ;
- Para $\lambda = 2$ e $\lambda = 3$, tais operações são chamadas de super curtas e curtas, respectivamente;
- Algoritmos polinomiais para operações super curtas [7, 10] e de aproximação para operações curtas [7, 9, 11, 15] são conhecidos;
- Para $\lambda > 3$ ainda não há resultados conhecidos.

Restrições Adicionais

Operações de Tamanho Limitado

- λ-Reversões e λ-Transposições;
- Distâncias: $d_r^\lambda(\pi)$, $d_{\bar{r}}^\lambda(\pi)$, $d_t^\lambda(\pi)$, $d_{rt}^\lambda(\pi)$ e $d_{\bar{r}t}^\lambda(\pi)$;
- Diâmetros: $D_r^\lambda(n)$, $D_t^\lambda(n)$, $D_{rt}^\lambda(n)$, $D_{\bar{r}}^\lambda(n)$ e $D_{\bar{r}t}^\lambda(n)$;

Exemplos

- $d_r^2(\pi)$: distância de ordenação por reversões super curtas;
- $D_{\bar{r}t}^3(n)$: diâmetro de ordenação por reversões e transposições curtas para permutações com sinais de tamanho n.

Objetivos

Estudar os Problemas

1. Ordenação de permutações sem e com sinais por λ-reversões;
2. Ordenação de permutações sem sinais por λ-transposições;
3. Ordenação de permutações sem e com sinais por λ-reversões e λ-transposições.

Metodologia

- Aspecto teórico;
- Estudo de teoremas e provas existentes na literatura;
- Propor algoritmos de aproximação para os problemas estudados;
- Implementação de programas para comparar resultados.

Cronograma de Atividades

	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;
5. Participação no Programa de Estágio Docente (PED);

Cronograma de Atividades

	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																								*

6. Investigação de algoritmos para o problema de ordenação de permutações sem sinais por λ-reversões;
7. Investigação de algoritmos para o problema de ordenação de permutações com sinais por λ-reversões;

Cronograma de Atividades

	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																								*

8. Investigação de algoritmos para o problema de ordenação de permutações sem sinais por λ-transposições;
9. Investigação de algoritmos para o problema de ordenação de permutações com e sem sinais por λ-reversões e λ-transposições;

Cronograma de Atividades

	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																								*

10. Escrita da dissertação;
11. Revisão da dissertação;
12. Defesa da dissertação.

- Ordenação de permutações sem sinais:
  - por λ-reversões: fator de aproximação igual a $0.6875p(p+1)$, onde $p = \left\lceil \frac{n}{\lfloor \frac{\lambda}{2} \rfloor} \right\rceil$;
  - por λ-transposições: fator de aproximação igual a $1.375 \left\lceil \frac{\lceil \frac{n}{2} \rceil}{\lceil \frac{\lambda}{2} \rceil} \right\rceil \left\lceil \frac{\lfloor \frac{n}{2} \rfloor}{\lfloor \frac{\lambda}{2} \rfloor} \right\rceil$.
- Ordenação de permutações com sinais:
  - por λ-reversões: fator de aproximação igual a $\frac{p(p+1)}{2}$, onde $p = \left\lceil \frac{n}{\lfloor \frac{\lambda}{2} \rfloor} \right\rceil$.
- Estamos preparando um artigo com estes resultados para submissão para uma conferência ainda este ano.

Tamanhos das Reversões Utilizadas pelo GRIMM

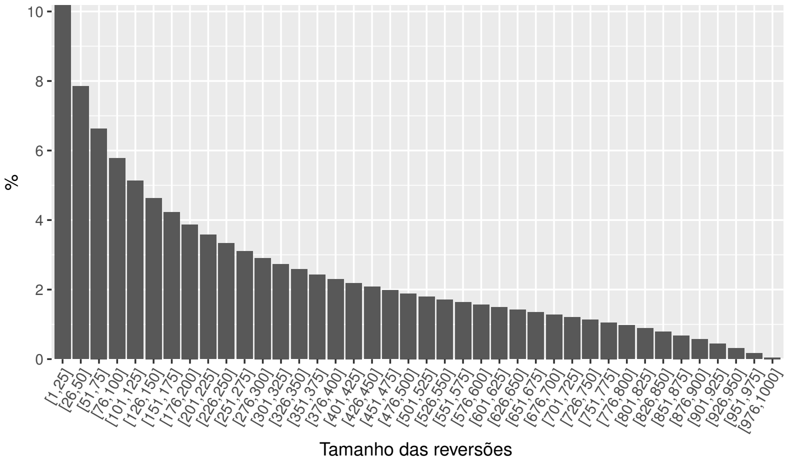


Figura: Histograma com as frequências dos tamanhos das reversões utilizadas pelo GRIMM para ordenar permutações de tamanho 1000, sem priorizar reversões menores.

Tamanhos das Reversões Utilizadas pelo GRIMM

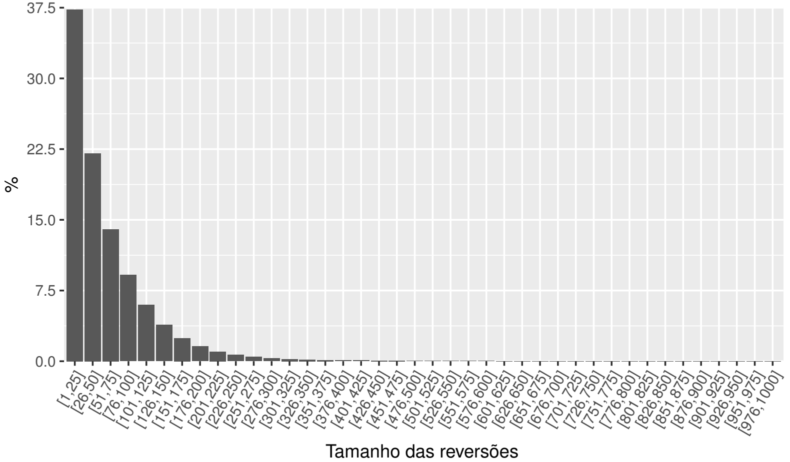


Figura: Histograma com as frequências dos tamanhos das reversões utilizadas pelo GRIMM para ordenar permutações de tamanho 1000, priorizando reversões menores.

Distância e Diâmetro para Permutações Pequenas

- Distâncias médias e diâmetro para:
  - Permutações sem sinais para valores de n e λ variando entre 2 e 10;
  - Permutações com sinais para valores de n e λ variando entre 2 a 9.
- No total foram criadas 10 tabelas, 2 para cada problema;
- Nos próximos slides apresentamos os resultados para reversões e transposições em permutações com sinais.

Distâncias para Permutações Pequenas

n	Tamanho máximo das operações (λ)							
-	2	3	4	5	6	7	8	9
2	1.25	-	-	-	-	-	-	-
3	2.33	1.83	-	-	-	-	-	-
4	3.83	2.67	2.49	-	-	-	-	-
5	5.80	3.58	3.20	3.09	-	-	-	-
6	8.26	4.65	3.93	3.73	3.67	-	-	-
7	11.22	5.87	4.75	4.42	4.30	4.26	-	-
8	14.69	7.26	5.66	5.12	4.89	4.81	4.78	-
9	18.66	8.83	6.65	5.87	5.55	5.42	5.37	5.35

Tabela: Distâncias Médias para o Problema de Ordenação por Reversões e Transposições em Permutações com Sinais ($d_{\bar{r}t}^{\lambda}(\pi)$).

Diâmetros para Permutações Pequenas

n	Tamanho máximo das operações (λ)							
-	2	3	4	5	6	7	8	9
2	2	-	-	-	-	-	-	-
3	4	3	-	-	-	-	-	-
4	7	4	4	-	-	-	-	-
5	11	5	5	4	-	-	-	-
6	16	7	6	5	5	-	-	-
7	22	9	7	6	6	6	-	-
8	29	12	8	7	7	6	6	-
9	37	15	10	8	7	7	7	7

Tabela: Diâmetros para o Problema de Ordenação por Reversões e Transposições em Permutações com Sinais ($D^{\lambda}_{\bar{r}t}(n)$).

References I

[1] V. Bafna and P. A. Pevzner. Sorting by Reversals: Genome Rearrangements in Plant Organelles and Evolutionary History of X Chromosome. *Molecular Biology and Evolution*, 12(2):239–246, 1995.

[2] V. Bafna and P. A. Pevzner. Sorting by Transpositions. *SIAM Journal on Discrete Mathematics*, 11(2):224–240, 1998.

[3] 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.

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

References II

[5] D. A. Dalevi, N. Eriksen, K. Eriksson, and S. G. E. Andersson. Measuring Genome Divergence in Bacteria: A Case Study Using Chlamydian Data. *Journal of Molecular Evolution*, 55(1):24–36, 2002.

[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] G. R. Galvão, O. Lee, and Z. Dias. Sorting Signed Permutations by Short Operations. *Algorithms for Molecular Biology*, 10(1):1–17, 2015.

[8] 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.

[9] L. S. Heath and J. P. C. Vergara. Sorting by Short Swaps. *Journal of Computational Biology*, 10(5):775–789, 2003.

References III

[10] M. R. Jerrum. The Complexity of Finding Minimum-length Generator Sequences. *Theoretical Computer Science*, 36(2-3):265–289, 1985.

[11] H. Jiang, H. Feng, and D. Zhu. An 5/4-Approximation Algorithm for Sorting Permutations by Short Block Moves. In H. Ahn and C. Shin, editors, *Proceedings of the 25th International Symposium on Algorithms and Computation (ISAAC'2014)*, Lecture Notes in Computer Science, pages 491–503. Springer International Publishing, 2014.

[12] J.-F. Lefebvre, N. El-Mabrouk, E. R. M. Tillier, and D. Sankoff. Detection and validation of single gene inversions. *Bioinformatics*, 19 (1):i190–i196, 2003.

[13] 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.

References IV

[14] E. Tannier, A. Bergeron, and M.-F. Sagot. Advances on Sorting by Reversals. *Discrete Applied Mathematics*, 155(6-7):881–888, 2007.

[15] J. P. C. Vergara. *Sorting by Bounded Permutations*. PhD thesis, Virginia Polytechnic Institute and State University, 1998.

[16] 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.

O Problema da Ordenação de Permutações por Operações de Tamanho Limitado

Guilherme Henrique Santos Miranda

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

Universidade Estadual de Campinas

03 de Outubro de 2017