

Resumo do artigo

"The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees"

Rodney Rick
email: rodneyrick@gmail.com

Resumo—Descrito pelos autores como um novo método chamado método de *Neighbor-Joining*[1] (NJ - União Vizinha) é proposto para reconstruir árvores filogenéticas a partir da distância evolutiva dos dados. Consiste em encontrar pares de unidades taxonômicas operacionais (OTUs - *Operational Taxonomic Units*) que minimizam o comprimento total entre ramos em cada estágio de um novo agrupamento de OTUs começando com uma árvore estrelada. Neste artigo foi comparado a outras técnicas que realizam a reconstrução de filogenia.

I. INTRODUÇÃO

Para a construção de árvores filogenéticas, o principal foco está na evolução mínima ou parcial máxima que é aplicado frequentemente a cada nova inclusão das OTUs. O algoritmo descreve a criação de árvores baseados neste princípio e examina todas as topologias possíveis (as possíveis novas ramificações) ou uma quantidade de número de topologias que apresentam proximidade da árvore inicial, escolhendo um novo ramo que tem a menor quantidade de mudança evolutiva para a construção da árvore final. Assim, para uma grande quantidade de OTUs, torna-se demorado esse processo de avaliação e inclusão para a construção da árvore.

Neste artigo os autores descrevem um novo método de produção de uma árvore única final e, possivelmente, mínima, chamada *Neighbor-Joining* (NJ). E realizam comparações com técnicas já conhecidas como: (a) distância de Wagner[2] (DW); (b) método de Farris[3] (MF); (c) construção de vizinhanças com Sattath e Tversky[4] (ST).

II. ALGORITMO

O algoritmo apresentado é semelhante ao método ST, como descreve os autores, sendo este para construção de topologia de árvores. Porém, em adicional, ao novo método NJ, também fornece os comprimentos de ramo da árvore final.

Antes de iniciar, necessita-se da definição do termo de vizinhos (ou vizinhança) é necessário. Um par de vizinhos é um par de OTUs conectados através de um único nó interior em uma árvore não-enraizada, logo uma bifurcação. A figura 1 (retirada do artigo) representa OTUs 1 e 2 como pares vizinhos que conectam-se com outros ramos utilizando o nó interior *A*, e assim sucessivamente para outros nós folhas e interiores. Para o cálculo de pares de vizinhos depende de N , sendo este o número de OTUs presentes. Se é par o número máximo é $N/2$, caso ímpar, então $(N - 1)/2$.

Para a soma dos comprimentos dos ramos, o cálculo segue a formulação: seja D_{ij} e L_{ab} como a distância entre OTUs i e j e o comprimento de ramo entre os nós a e b , respectivamente. A distância é dada por $L_{aX} + L_{Xb} = D_{ab}$ onde X representa o nó interior. A soma total das ramificações é descrita pela equação 1.

$$[h]S_0 = \sum_{k=1}^N L_{iX} = \frac{1}{N-1} \sum_{i < j} D_{ij} \quad (1)$$

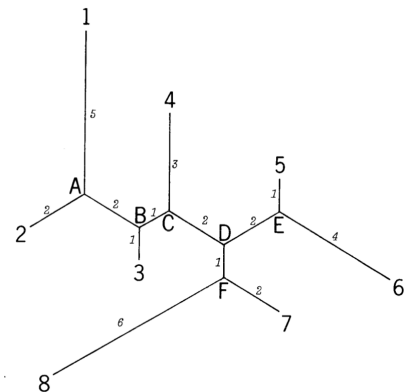


Figura 1. OTUs organizadas em uma topologia. Os nós folhas são representados por números, nós interiores por letras e números em itálico são as distâncias entre o nó interno e folha.

De modo geral, a técnica inicia-se no formato estrela (figura 2(a)), com somente um nó interno (X). Assim, utilizando uma matriz de distância entre os nós e seus vizinhos, o próximo passo considera a associação dos vizinhos mais próximos e ao mesmo tempo mais distantes de todos os outros, ramificando-se e criando um novo nó interno (Y), presente na figura (b)).

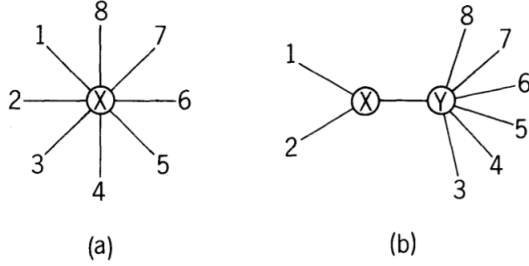


Figura 2. Processamento do primeiro passo do algoritmo de (a) para (b).

Para a construção das ramificações, conforme descrito no parágrafo interior, deve-se atentar para a utilização da fórmula 2. O primeiro termo da equação, refere-se a soma de todas as distâncias, incluindo L_{XY} e os outros dois termos exclui ramificações irrelevantes ao processo nesta etapa.

$$L_{XY} = \frac{1}{2(N-2)} \left[\sum_{k=3}^N (D_{1k} + D_{2k}) - (N-2)(L_{1X} + D_{2X}) + 2 \sum_{i=3}^N L_{iY} \right] \quad (2)$$

Respeitando-se as condições de distância e aplicando a fórmula 2, torna-se possível reconstruir a topologia da árvore filogenética, conforme figura 3.

III. CRITÉRIO DE ÁRVORE COM EVOLUÇÃO MÍNIMA

Esse critério para a construção de uma árvore demonstra que a soma da distância (d_{1N} e d_{2N}) entre dois nós, 1 e 2, até seu nó interior N , deve ser menor que a soma de qualquer outra distância de um nó i até N . Ou seja, $d_{1N} + d_{2N} < d_{1N} + d_{iN}$ ou $d_{1N} + d_{2N} < d_{2N} + d_{iN}$. Assim, para respeitar esse critério, uma matriz de distância entre as OTUs para encontrar os pares vizinhos mais adequados, reconstruindo a topologia mais aproximada.

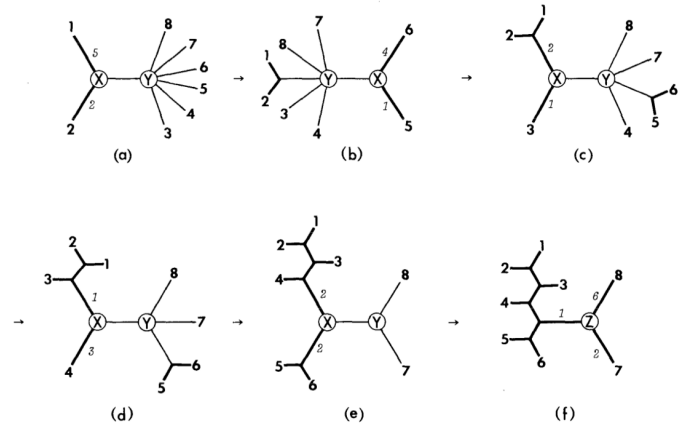


Figura 3. Representação de reconstrução de árvore topológica.

IV. EFICIÊNCIA DO MÉTODO NJ PARA RECUPERAÇÃO DA TOPOLOGIA ADEQUADA

Nesta seção, os autores demonstram as comparações de eficiência do método de NJ com cinco outros métodos. Todas as metodologias produzem uma árvore parcimoniosa única a partir de dados de distância. Os pontos considerados para avaliação foram as taxas constantes e variáveis (esperadas) de substituição de nucleotídeos.

Neste artigo, relata alguns modelos de árvore e comparações entre as metodologias. A técnica UPGMA[5] (*Unweighted Pair Group Method with Arithmetic Mean*), entre todos os modelos e tabelas comparativas é a metodologia que apresenta a performance mais fraca. Não muito distante, o método MF exibe quase o mesmo desempenho. No caso da metodologia de Wagner (DW), com os parâmetros adequados, expõe valores nas tabelas ligeiramente a frente das duas técnicas citadas anteriormente. Para quantidade de OTUs acima de 1000 pares, a metodologia LI[6] apresenta bom resultados.

Com os mesmo parâmetros e resultados mais próximos e alto desempenho, comparado com as outras três técnicas, as metodologias com melhores soluções são ST e NJ.

V. DISCUSSÃO SOBRE A METODOLOGIA NJ

O método NJ minimiza a soma de comprimentos das ramificações em cada estágio de agrupamento de novos pares de OTUs, sempre iniciando com uma árvore modelo estrela sendo esse o grande diferencial do algoritmo de evolução mínima padrão. Outro ponto a ser considerado, esta técnica não necessariamente produz uma árvore mínima, mas sim uma árvore de topologia próxima da topologia verdadeira.

Os autores descrevem os comprimentos dos ramos da árvore obtidos pelo método de NJ como estimações baseadas em mínimos quadrados determinados em cada etapa de agrupamento de novos pares de OTUs. Encontra-se mais detalhes no Apêndice A do artigo.

Alterações para estimativas negativas podem ser consideradas, valores de ramos negativos, podem ser alterados quando erros de amostragem ocorrem para zero, desconsiderando tais resultados. Esses são alguns ajustes que podem agregar na solução dependendo do caso.

REFERÊNCIAS

- [1] Naruya Saitou and Masatoshi Nei. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, 4(4):406–425, 1987.
- [2] Gustavo H. Orair, Carlos H. C. Teixeira, Wagner Meira, Jr., Ye Wang, and Srinivasan Parthasarathy. Distance-based outlier detection: Consolidation and renewed bearing. *Proc. VLDB Endow.*, 3(1-2):1469–1480, September 2010.
- [3] James S. Farris. Estimating phylogenetic trees from distance matrices. *The American Naturalist*, 106(951):645–668, 1972.
- [4] Amos Tversky, Shmuel Sattath, and Paul Slovic. Contingent weighting in judgment and choice. *Psychological review*, 95(3):371, 1988.
- [5] R.R. Sokal and P.H.A. Sneath. *Principles of Numerical Taxonomy*. Series of books in biology. W. H. Freeman, 1963.
- [6] Wen-Hsiung Li. Simple method for constructing phylogenetic trees from distance matrices. *Proceedings of the National Academy of Sciences*, 78(2):1085–1089, 1981.