

1

Introdução

Uma árvore evolutiva, árvore filogenética ou simplesmente filogenia é uma árvore que relaciona grupos de espécies, populações de espécies distintas, populações de uma mesma espécie, ou genes homólogos em populações de espécies distintas. Estas entidades podem ser classificadas sob a designação comum de *taxons* [6, 71, 77]. Em uma filogenia, as folhas representam os *taxons*, enquanto os nós interiores representam ancestrais hipotéticos.

As relações entre os taxons são baseadas na similaridade entre um conjunto de atributos, denominados *características*. Cada característica pode assumir valores dentro de um conjunto finito de estados. Cada taxon é definido pelos estados assumidos por cada característica. Características binárias são aquelas que possuem apenas dois possíveis estados, que representam a presença ou a ausência de algum atributo. As instâncias do problema da filogenia com características binárias são representadas por matrizes de zeros e uns, onde cada elemento (i, j) corresponde ao estado da característica j para o taxon i .

Um passo evolutivo é descrito como cada mudança de estado ocorrida em um ramo de uma árvore filogenética. O critério de parcimônia assume que a melhor filogenia é aquela que possui o número mínimo de passos evolutivos. Dado um conjunto de taxons definidos sobre um conjunto de características, o *problema da filogenia* consiste em encontrar a árvore filogenética com o menor número de passos evolutivos. Ele é NP-difícil segundo qualquer um dos critérios atualmente utilizados pelos sistematas [9, 12, 27, 28].

Seja E o conjunto de todos os possíveis taxons e $W = \{x^{(1)}, \dots, x^{(n)}\} \subseteq E$ o conjunto dos taxons operacionais sob análise. Uma árvore filogenética s para o conjunto de taxons operacionais W pertence ao conjunto S de todas as árvores não enraizadas com n folhas (cada folha tem um relacionamento um-para-um com um taxon operacional $x^{(i)} \in W, i=1, \dots, n$) e todos os nós internos com grau três. Seja $f: S \rightarrow \mathbb{R}$ uma função que associa cada filogenia

$s \in S$ com seu valor de parcimônia. O problema da filogenia sob parcimônia pode então ser formulado como encontrar a filogenia $s^* \in S$, tal que $f(s^*) = \min_{s \in S} f(s)$. Farris and Fitch [17, 24, 25] propuseram algoritmos polinomiais com complexidade de tempo $O(mn)$ para realizar a computação do valor de parcimônia de uma filogenia, onde n é o número de taxons operacionais e m o número de características binárias.

O objetivo deste trabalho é desenvolver algoritmos eficientes para o problema da filogenia combinando características de diferentes métodos heurísticos.

Na literatura científica existem vários trabalhos propondo heurísticas para a problema da filogenia [13, 48, 57, 58]. Andreatta e Ribeiro [3, 4] relataram e compararam os resultados computacionais obtidos por uma variedade de heurísticas sobre um conjunto de oito instâncias extraídas da literatura. Eles mostraram que as melhores soluções não eram sempre encontradas pela mesma heurística. Neste trabalho foram desenvolvidas heurísticas robustas que além de conseguir igualar os melhores resultados da literatura para essas oito instâncias, para algumas foram obtidos resultados ainda melhores.

O problema da filogenia é discutido com maiores detalhes no Capítulo 2, onde são apresentadas suas aplicações e um algoritmo para a definição do valor de parcimônia de uma filogenia. Também são mostradas algumas estratégias de otimização implementadas que permitiram realizar operações em árvores filogenéticas exigindo um menor esforço computacional em relação a outras implementações. Algumas dessas otimizações foram propostas por Goloboff [36, 37, 38] e outras são contribuições deste trabalho.

No Capítulo 3 é apresentado uma heurística GRASP utilizando uma estratégia VND como método de refinamento de suas soluções. Os experimentos realizados mostram sua robustez quando testada sobre oito instâncias da literatura e também sobre vinte instâncias geradas aleatoriamente. Este gerador também é discutido em detalhes neste capítulo.

Um algoritmo genético utilizando uma nova estratégia de cruzamento de indivíduos baseada em reconexão por caminhos é descrito no Capítulo 4. A utilização de reconexão por caminhos como operador de cruzamento é uma contribuição original desta tese. Os experimentos foram realizados sobre as oito instâncias da literatura e para as vinte instâncias geradas. Os resultados encontrados foram ainda melhores, tanto em qualidade das soluções quanto em tempo computacional, do que os da heurística GRASP, apresentada no Capítulo 3.

As conclusões são apresentadas no último capítulo.