

5

Conclusões e extensões

O uso de heurísticas híbridas já se mostrou eficiente para vários problemas de otimização combinatória [1, 11, 23, 30, 31, 32, 33, 47, 64]. O principal objetivo deste trabalho foi o de implementar e avaliar heurísticas híbridas para o problema da filogenia.

Este problema é um dos mais importantes problemas em biologia comparativa. Algoritmos exatos e aproximados (para problemas pequenos) para a computação de árvores filogenéticas estão espalhados pela literatura científica [3, 4, 13, 48, 57, 58]. Neste trabalho, foram analisados algoritmos que combinaram a heurística GRASP com a heurística VND e que combinaram algoritmos genéticos com a técnica de reconexão por caminhos. Todos eles se mostraram bastante robustos e quando não superaram os melhores resultados da literatura, pelo menos os igualaram.

No Capítulo 3 foi implementado o algoritmo híbrido GRASP+VND, o qual é um algoritmo GRASP que utiliza VND como método de busca local. A estrutura de vizinhança utilizada é a k -SPR, que foi proposta neste trabalho. Este algoritmo foi comparado com o algoritmo GRASP em [3, 4], que utilizava a vizinhança SPR em seu método de busca local. Os dois algoritmos foram submetidos a experimentos utilizando 8 instâncias da literatura e 20 instâncias geradas aleatoriamente. Os testes mostraram que o algoritmo proposto é superior tanto em qualidade das soluções obtidas quanto em tempo computacional. Os testes também mostraram que se ambos algoritmos forem executados durante um determinado tempo, o algoritmo GRASP+VND possui uma probabilidade maior de encontrar uma solução pelo menos tão boa quanto um determinado alvo.

No Capítulo 4 foram implementados dois algoritmos, AG+PR1 e AG+PR2. Ambos são algoritmos genéticos cuja etapa de cruzamento de indivíduos é realizada por reconexão por caminhos. O que diferencia os dois algoritmos é a maneira pela qual uma folha mal posicionada da solução inicial é selecionada para remoção e o critério utilizado para escolher em que aresta ela deve ser reinserida. Ambos algoritmos foram comparados com o algoritmo

GRASP+VND [65] descrito no Capítulo 3. Os três algoritmos foram submetidos a experimentos utilizando 8 instâncias da literatura e 20 instâncias geradas aleatoriamente. Os testes mostraram que o algoritmo AG+PR2 é superior tanto em qualidade das soluções obtidas quanto em tempo computacional, quando comparado com o algoritmo GRASP+VND. Os testes também mostraram que se os três algoritmos forem executados durante um dado tempo, os algoritmos AG+PR1 e AG+PR2 possuem uma probabilidade maior de encontrar uma solução pelo menos tão boa quanto um determinado alvo, sendo a probabilidade do algoritmo AG+PR2 maior do que a do algoritmo AG+PR1. Os resultados apresentados também mostram a eficiência do operador genético baseado na reconexão por caminhos do tipo 2. Eles mostram que em 33% das execuções, o valor alvo é atingido durante a fase de reconexão por caminhos e em 67% o alvo foi atingido após uma reconexão por caminhos seguida de uma busca local com a vizinhança SPR.

A utilização de reconexão por caminhos como operador de cruzamento é uma contribuição original desta tese.

Ainda no Capítulo 4, foram mostrados outros dois algoritmos, GRASP+VND+PR2 e AG+PR1E2. O primeiro combina o algoritmo GRASP+VND com uma etapa de intensificação da busca via reconexão por caminhos do tipo 2. O segundo é um algoritmo genético que utiliza ambos os tipos de reconexão por caminhos. Este algoritmo se mostrou muito eficiente, superando o algoritmo AG+PR2, que era o melhor algoritmo até então.

Os trabalhos futuros estão relacionados ao algoritmo genético com reconexão por caminhos, pois, neste trabalho, esta combinação se mostrou bastante eficiente e robusta para o problema da filogenia. A primeira extensão seria fazer um ajuste de parâmetros mais detalhado. Existem muitos parâmetros que não foram variados o bastante durante este trabalho. Entre eles estão o tamanho da população, a frequência para aplicação da busca local, o tamanho das classes A , B e C na estrutura do algoritmo genético e o número de renovações da população, entre outros. Um bom ajuste destes demandará bastante tempo devido à grande quantidade de parâmetros. Outra extensão seria inserir um operador de mutação que permitiria diminuir o tamanho da classe C , pois esse operador iria gerar mais diversidade na população. Além disso, avaliar qual seria a sua periodicidade (taxa de mutação), uma vez que este é um operador secundário.

Outro trabalho futuro interessante seria o de implementar um novo operador de cruzamento para se fazer novas avaliações do operador de cruzamento via reconexão por caminhos proposto neste trabalho. O novo operador seria baseado na idéia de árvores de consenso. Uma possível

implementação seria uma adaptação do *greedy-consensus* que preencheria o filho com subárvores comuns a ambos os pais. Depois, o filho seria completado escolhendo aleatoriamente pedaços de algum dos pais.

Por último, poderia ser aproveitado o paralelismo inerente aos algoritmos genéticos, paralelizando-se alguns dos algoritmos genéticos propostos. Uma idéia interessante seria a de construir um algoritmo paralelo híbrido que utilize em cada nó uma das heurísticas implementadas. As melhores heurísticas executariam em um número maior de processadores, enquanto as inferiores executariam em um número menor de processadores.