

Dalessandro Soares Vianna

**Heurísticas híbridas para o
problema da filogenia**

TESE DE DOUTORADO

DEPARTAMENTO DE INFORMÁTICA
Programa de Pós-graduação em
Informática

Rio de Janeiro
Fevereiro de 2004



Dalessandro Soares Vianna

**Heurísticas híbridas para o problema da
filogenia**

Tese de Doutorado

Tese apresentada ao Programa de Pós-graduação em In-
formática do Departamento de Informática da PUC-Rio
como parte dos requisitos parciais para obtenção do título
de Doutor em Informática

Orientador: Prof. Celso Carneiro Ribeiro

Rio de Janeiro
Fevereiro de 2004



Dalessandro Soares Vianna

**Heurísticas híbridas para o problema da
filogenia**

Tese apresentada ao Programa de Pós-graduação em Informática do Departamento de Informática do Centro Técnico Científico da PUC-Rio como parte dos requisitos parciais para obtenção do título de Doutor em Informática. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Celso Carneiro Ribeiro

Orientador
Departamento de Informática — PUC-Rio

Prof. Luiz Fernando Bessa Seibel

Departamento de Informática – PUC-Rio

Prof. Sérgio Lifschitz

Departamento de Informática – PUC-Rio

Prof. Emmanuel Piseces Lopes Passos

Rede Rio

Prof. Luiz Satoru Ochi

Departamento de Ciência da Computação – UFF

Prof. Nair Maria Maia Abreu

Departamento de Engenharia da Produção –
COPPE/UFRJ

Prof. Nalvo Franco Almeida Junior

Departamento de Computação e Estatística – UFMS

Prof. José Eugenio Leal

Coordenador Setorial do Centro Técnico Científico —
PUC-Rio

Rio de Janeiro, 27 de Fevereiro de 2004

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Dalessandro Soares Vianna

Graduou-se em Ciência da Computação na Universidade Federal Fluminense, onde também cursou o mestrado em Ciência da Computação.

Ficha Catalográfica

Vianna, Dalessandro Soares

Heurísticas híbridas para o problema da filogenia/
Dalessandro Soares Vianna; orientador: Celso Carneiro
Ribeiro. — Rio de Janeiro : PUC-Rio, Departamento de
Informática, 2004.

v., 101 f: il. ; 30 cm

1. Tese (doutorado) - Pontifícia Universidade
Católica do Rio de Janeiro, Departamento de In-
formática.

Inclui referências bibliográficas.

1. Informática – Teses. 2. Heurísticas híbridas. 3.
VND. 4. Problema da filogenia. 5. GRASP. 6. Algorit-
mos genéticos. I. Ribeiro, Celso Carneiro. II. Pontifícia
Universidade Católica do Rio de Janeiro. Departamento
de Informática. III. Título.

CDD: 004

Aos meus queridos pais Anézio Martins Vianna e Iolanda Soares Vianna, à minha linda esposa Marcilene de Fátima Dianin Vianna e ao meu anjinho Vinícius Dianin Vianna, por todo amor e carinho recebido.

Agradecimentos

À Deus, por ter me dado condições de enfrentar e superar os desafios com serenidade e sucesso, trazendo mais esta conquista para minha vida.

Aos meus familiares – pai, mãe, irmãos, esposa e filho – por me oferecer tanto amor, carinho, compreensão, incentivo, paciência, confiança, dedicação, alegria, conforto e tantas outras coisas que foram e serão imprescindíveis em qualquer desafio que eu venha enfrentar.

Ao professor Celso Carneiro Ribeiro pela excelente orientação prestada. É um ótimo referencial na carreira acadêmica que estou começando a traçar.

Ao Conselho Nacional de Desenvolvimento e Pesquisa (CNPQ), pela bolsa concedida que viabilizou o meu curso.

Aos meus amigos da graduação (turma *93), pelo bom humor que permitiu bons momentos de descontração durante estes últimos anos.

À Universidade Candido Mendes (UCAM-Campos), instituição onde trabalho há um ano, por ter apoiado a finalização do meu curso oferecendo tempo suficiente para que eu me dedicasse inteiramente ao meu Doutorado.

Aos professores e funcionários que compõem o departamento de informática da PUC-Rio. Em especial aos professores Luiz Fernando Gomes Soares e Clarisse Sieckenius de Souza e às funcionárias Deborah e Emanuelle, pela atenção e disponibilidade nos momentos em que precisei de auxílio.

Resumo

Vianna, Dalessandro Soares; Ribeiro, Celso Carneiro. **Heurísticas híbridas para o problema da filogenia**. Rio de Janeiro, 2004. 101p. Tese de Doutorado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Uma filogenia é uma árvore que relaciona unidades taxonômicas, baseada na similaridade de seus conjuntos de características. O problema da filogenia consiste em encontrar uma filogenia com o número mínimo de passos evolutivos. O principal objetivo deste trabalho é desenvolver heurísticas híbridas para este problema. Duas estratégias são propostas. A primeira combina a metaheurística GRASP baseada em uma nova estrutura de vizinhança (k -SPR) proposta neste trabalho com um procedimento VND de busca local. A segunda estratégia híbrida combina algoritmos genéticos com uma estratégia de cruzamento inovadora, a qual é uma extensão da técnica de intensificação denominada reconexão por caminhos que foi originalmente aplicada no contexto de outras metaheurísticas, tais como busca tabu e GRASP. Os experimentos computacionais realizados sobre instâncias geradas aleatoriamente e instâncias da literatura científica mostram que os novos algoritmos são bastante robustos e que superaram os outros algoritmos existentes na literatura em termos de qualidade de solução e tempos computacionais obtidos.

Palavras-chave

Problema da filogenia; heurísticas híbridas; GRASP; VND; algoritmos genéticos; reconexão por caminhos.

Abstract

Vianna, Dalessandro Soares; Ribeiro, Celso Carneiro. **Hybrid heuristics for the phylogeny problem**. Rio de Janeiro, 2004. 101p. PhD. Thesis — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

A phylogeny is a tree that relates taxonomic units, based on their similarities over a set of characters. The phylogeny problem consists in finding a phylogeny with the minimum number of evolutionary steps. The main goal of this work is to develop hybrid heuristics for this problem. Two strategies are proposed. The first combines the GRASP metaheuristic using a new neighborhood structure (k -SPR) proposed in this work with a VND local search procedure. The second hybrid strategy combines genetic algorithms with an innovative optimized crossover strategy which is an extension of the path-relinking intensification technique originally applied in the context of other metaheuristics such as tabu search and GRASP. Computational results on randomly generated and benchmark instances are reported, showing that the new heuristics are quite robust and outperform the others algorithms in the literature in terms of solution quality and computational time.

Keywords

Phylogeny problem; hybrid heuristics; GRASP; VND; genetic algorithm; path-relinking.

Conteúdo

1	Introdução	12
2	O problema da filogenia	14
2.1	Aplicações	16
2.2	Revisão bibliográfica/trabalhos anteriores	17
2.3	Avaliação de uma filogenia	19
2.4	Otimizações na implementação	23
2.5	Conclusão	42
3	Um algoritmo GRASP com VND	43
3.1	Heurística GRASP	43
3.2	<i>Variable Neighborhood Descent</i> (VND)	46
3.3	Método construtivo guloso aleatorizado	47
3.4	Estruturas de vizinhança	49
3.5	Busca local usando VND	54
3.6	Heurística GRASP com VND	55
3.7	Resultados computacionais	56
3.8	Conclusão	61
4	Um algoritmo genético com reconexão por caminhos	63
4.1	Algoritmos genéticos	64
4.2	Reconexão por caminhos	70
4.3	Criação da população inicial	72
4.4	Renovação da população	76
4.5	Cruzamento de indivíduos via reconexão por caminhos	77
4.6	Resultados computacionais	81
4.7	Conclusão	86
5	Conclusões e extensões	91

Lista de Figuras

2.1	Exemplo de um conjunto de taxons e uma filogenia associada.	16
2.2	Exemplo de representação de uma filogenia.	20
2.3	Exemplo de uma filogenia bem avaliada.	20
2.4	Algoritmo para avaliação de uma filogenia s e para determinação do vetor CC de cada nó de s .	21
2.5	Exemplo de inserção de uma subárvore em uma filogenia.	24
2.6	Algoritmo para determinação do vetor CE de cada nó.	26
2.7	Algoritmo para cálculo do custo de inserção.	27
2.8	Exemplo de remoção de uma subárvore de uma filogenia.	27
2.9	Algoritmo para avaliação de uma filogenia s e para determinação dos vetores CC e $CCAux$ de cada nó de s .	28
2.10	Algoritmo para cálculo da redução de passos evolutivos ocasionada por uma remoção.	30
2.11	Algoritmo para avaliação de uma filogenia s e para determinação dos vetores CC e $CCAux$ de cada nó de s .	36
2.12	Algoritmo para determinação do vetor CE de cada nó.	37
2.13	Algoritmo para cálculo do custo de inserção.	38
2.14	Algoritmo para cálculo da redução de passos evolutivos gerada por uma remoção.	40
3.1	Algoritmo de construção padrão.	45
3.2	Algoritmo de busca local padrão.	46
3.3	Algoritmo GRASP.	47
3.4	Algoritmo VND.	48
3.5	Algoritmo básico para construção de uma filogenia.	49
3.6	Alternativas para a inserção de um novo taxon em uma filogenia com três taxons.	50
3.7	Ilustração de um movimento dentro da vizinhança SPR.	51
3.8	Pseudo-código do procedimento de busca BL^1 usando a vizinhança SPR.	52
3.9	Pseudo-código do procedimento de busca BL^2 usando a vizinhança 2-SPR.	53
3.10	Procedimento VND para busca local.	54
3.11	Pseudo-código da heurística GRASP+VND.	55
3.12	Distribuições empíricas de probabilidade do tempo-para-valor-alvo para a instância ROPA para o algoritmo GRASP em [3, 4] e a heurística GRASP+VND.	60
3.13	Distribuições empíricas de probabilidade do tempo-para-valor-alvo para a instância GOLO para o algoritmo GRASP em [3, 4] e a heurística GRASP+VND.	61
4.1	Algoritmo genético padrão.	66
4.2	Exemplo de mutação.	69
4.3	Reconexão por caminhos no contexto da busca tabu.	71

4.4	Algoritmo de reconexão por caminhos padrão.	72
4.5	Algoritmo de construção da população inicial.	73
4.6	Evolução da população.	73
4.7	Estrutura do algoritmo genético implementado.	75
4.8	Algoritmo de renovação da população.	76
4.9	Algoritmo de cruzamento de indivíduos utilizando reconexão por caminhos.	78
4.10	Exemplo de reconexão por caminhos em filogenias.	79
4.11	Distribuições empíricas de probabilidade do tempo-para-valor-alvo para a instância SCHU para os algoritmos GRASP+VND, AG+PR1 e AG+PR2.	82
4.12	Resultados comparativos para dez execuções dos algoritmos GRASP+VND e AG+PR, para as oito instâncias testes da literatura.	84
4.13	Resultados comparativos para dez execuções dos algoritmos GRASP+VND e AG+PR, para as vinte instâncias testes geradas aleatoriamente.	85
4.14	Pseudo-código da heurística GRASP+VND+PR2.	88
4.15	Distribuições empíricas de probabilidade do tempo-para-valor-alvo para a instância SCHU para os algoritmos AG+PR2, GRASP+VND+PR2 e AG+PR1E2.	89
4.16	Resultados comparativos para dez execuções dos algoritmos AG+PR2 e AG+PR1E2, para as oito instâncias testes da literatura.	89
4.17	Resultados comparativos para dez execuções dos algoritmos AG+PR2 e AG+PR1E2, para as vinte instâncias testes geradas aleatoriamente.	90

Lista de Tabelas

2.1	Melhores resultados para as oito instâncias da literatura.	18
2.2	Tempo médio de execução dos algoritmos GRASP_1 e GRASP_2 em segundos.	41
3.1	Características dos problemas gerados aleatoriamente.	57
3.2	Resultados comparativos sobre problemas gerados aleatoriamente.	58
3.3	Resultados obtidos pela heurística GRASP+VND.	58
3.4	Resultados comparativos para dez execuções do algoritmo GRASP em [3, 4] e a heurística GRASP+VND para as oito instâncias testes da literatura.	59
3.5	Resultados comparativos para dez execuções do algoritmo GRASP em [3, 4] e a heurística GRASP+VND para as instâncias testes geradas aleatoriamente.	59