

Análise e Construção de Árvores Filogenéticas Usando Algoritmo Genético

Roberta Vilhena Vieira[‡], Vana do Carmo Carvalho Pereira[‡]
José Homero Feitosa Cavalcanti[†]

[‡] UFPB/CCT/DSC, [†] UFPB/CCT/DEE
Av. Aprígio Veloso, 882 CEP-58.109-970 Campina Grande - PB
E-mail: homero@dee.ufpb.br

Resumo - Neste trabalho apresenta-se um algoritmo genético, baseado em pesos, usado para construção de árvores filogenéticas. Apresentam-se também resultados experimentais preliminar obtidos com a implementação do algoritmo genético na construção de árvores filogenéticas.

Abstract - This paper shows a genetic algorithm, weight based, developed for building philo-genetic trees. The preliminary experimental results obtained with the genetic algorithm implementation for building philo-genetic trees.

Introdução

O lento processo de transformação que sofrem os organismos da Terra é chamado de evolução. Árvores filogenéticas mostram como os organismos atualmente existentes se relacionam através de seus organismos ancestrais.

A fig.1 mostra um exemplo de uma árvore filogenética^{1,2} (AF) típica. Essa árvore foi construída a partir da hipótese de que existe uma filogenia perfeita para ela. A construção de árvores filogenéticas e o problema da filogenia perfeita são problemas NP-completo.

Neste trabalho considera-se que é possível ser construída a AF usando a representação em pesos, obtidos a partir da similaridade das características das espécies. O problema resultante é NP-completo e pode ser comparado com o problema do caixeiro viajante que também o é.

A solução adotada usa algoritmos genéticos³ para a geração e análise das possíveis AF a serem construídas a partir da matriz de ESPÉCIE X CARACTERÍSTICA mostrada na TABELA 1.

CARACTERÍSTICA						
ESPÉCIE	C1	C2	C3	C4	C5	C6
A	0	0	0	1	1	0
B	1	1	0	0	0	0
C	0	0	0	1	1	1
D	1	0	1	0	0	0
E	0	0	0	1	0	0

TABELA 1

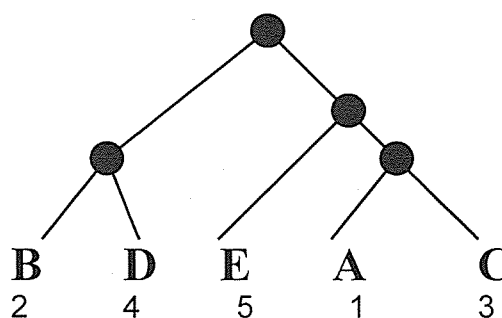


Fig.1 Árvore filogenética

A cada galho da AF é atribuído um cromossomo. Por exemplo, galhos ABCDE, cromossomo {1;2;3;4;5}. A AF terá como galhos a sequência dos cromossomos. A AF da fig.1, desenvolvida a partir da TABELA 1, tem o seguinte cromossomo: {2;4;5;1;3}.

O algoritmo genético (AG) caracteriza-se por um processo de seleção baseado no desempenho e na população de árvores gerada de uma forma similar a evolução natural. Os elementos da população são listas fixas de parâmetros a serem otimizadas baseadas numa determinada medida de desempenho ("fitness"). O AG possui 5 componentes básicos: 1)Um método para codificação dos cromossomos (p.e., sequência {1;2;3;4;5}); 2)Uma função objetivo "fitness" (maior distância e posicionamento); 3)Uma população inicial (100 cromossomos); 4)Um conjunto de operadores para formarem a evolução entre dois cromossomos de populações consecutivas (mutação e cruzamento, trocam-se as posições das espécies, p.e., {1;2;3;4;5; ->{2;1;3;4;5}); 5)Parâmetros de trabalho (p.e., 1% de probabilidade de mudança no cromossomo).

Metodologia

O algoritmo do AG utilizado está mostrado na fig.2. Inicialmente, calcula-se a matriz de similaridade entre os galhos (TABELA 2). Geram-se aleatoriamente uma população de 100 indivíduos. Escolhem-se os cromossomos para as 10 AF que tenham a maior similaridade. Por último, escolhe-se a AF que possa ser desenhada como mostrado na fig.1.

ESPÉCIE					
ESPÉCIE	A	B	C	D	E
A	0	0	2	0	1
B		0	0	1	0
C			0	0	1
D				0	0
E					0

TABELA 2

Resultados

Obtiveram-se as seguintes 10 melhores AF (todas com similaridade total 4): {5;3;1;4;2} {5;3;1;2;4} {5;1;3;2;4} {5;1;3;4;2} {5;3;1;2;4} {4;2;5;1;3} {2;4;5;1;3} {4;2;5;1;3} {3;1;5;4;2} {3;1;5;2;4}. As soluções corretas são: {4;2;5;1;3} {2;4;5;1;3} {3;1;5;2;4} e {3;1;5;4;2}.

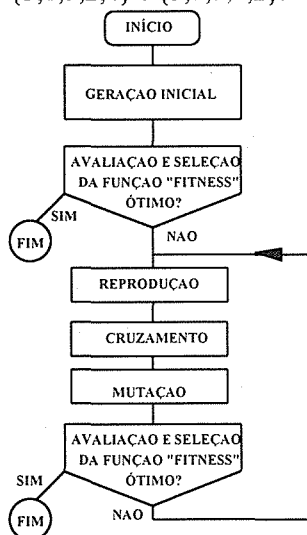


Fig.2 Algoritmo genético

Discussão e Conclusões

Foram apresentados resultados experimentais para a construção de AF usando algoritmos genéticos. Este trabalho faz parte de um sistema de desenvolvimento de programas que permitirá a construção de AF automaticamente.

Referências

- (1) Kaufmann, Morgan. **Foundations of Genetic Algorithms 2**, Campus Drive, San Mateo, California.
- (2) Meidanis, João & Setubal, João. **Uma Introdução a Biologia Computacional**. IX Escola de Computação, Recife, 1994.
- (3) Aguilar, José. **A Hybrid Approach Based on Genetic Algorithms and Simulated Annealing for Combinatorial Optimization Problems**, Anais / XXII Seminário Integrado de Software e Hardware, 31/julho a 4/agosto 1995. Vol. II, pp.1401- 1412.