

Trabalho Prático 3

Visualização de evolução em algoritmos genéticos: Uma aplicação no alinhamento múltiplo de proteínas

**Larissa Fernandes Leijôto¹, Rubens Emílio Alves Moreira¹, Péricles Rafael Alves¹,
Vinícius Garcia¹**

¹Departamento de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG)
Caixa Postal 6627 – 31270-901 – Minas Gerais – MG – Brazil

{leijoto,periclesroalves,rubens,garcia}@dcc.ufmg.br

1. Introdução

Neste trabalho, apresentamos uma visualização para a avaliação da evolução das soluções de um Algoritmo Genético (AG) para o problema alinhamento múltiplo de proteínas. Visualizações para a classe de algoritmos que procuram abranger o espaço de busca, são extremamente complexas. E isso se deve ao fato do número exponencial do conjunto de soluções geradas que dificulta visualização. Com isso, buscamos uma maneira de visualizar as soluções que o AG gerou em sua busca, de forma a inferir propriedades do algoritmo que dificilmente seriam derivadas através de inspeção simples em dados brutos.

2. Algoritmo Genético

Os AGs são algoritmos evolucionários, baseados no princípio da evolução natural de Charles Darwin. Ele parte do pressuposto de que existe um ambiente, que é composto por uma população. Essa população é formada por indivíduos, que são submetidos a uma avaliação a cada iteração. A avaliação mede o quão adaptado ele está ao ambiente. Indivíduos que estão melhor adaptados, terão maior chance de permanecer na população e gerar novos indivíduos. A geração e a seleção de indivíduos é feito de maneira estocástica, com o intuito que a população mantenha uma diversidade. Essa diversidade é extremamente importante para que a população não fique totalmente igual, sendo assim incapaz de gerar novos indivíduos melhores do que os que já estão na população atual. O processo de visualização de um algoritmo genético, nos ajuda a analisar a sua convergência. Possibilitando que o usuário analise visualmente diferentes taxas de probabilidade para as operações possíveis, podendo alterar para mais ou menos e analisar novamente os parâmetros do algoritmo genético.

2.0.1. Espaço de Busca

O espaço de busca no AG é equivalente ao conjunto inteiro de soluções de um dado problema. Como os problemas abordados por AGs são NP, é impossível visualizar o espaço inteiro. Uma vez que não somos capazes de gerar tal espaço, devido a limitação de tempo. Assim, essa visualização foi criada com o intuito de possibilitar ao usuário, uma visualização global do sub-espaço em que o algoritmo está caminhando, podendo inferir se está ou não ocorrendo um ótimo local. Podemos também verificar na fase final a convergência do algoritmo, onde provavelmente a dispersão dos pontos será muito pequena.

2.0.2. Redução de Dimensionalidade

Criamos um método para reduzir a dimensionalidade dos indivíduos, para podermos plotá-los em um plano 2D. A equação utilizada para planificar os dados em 2D, foi criada com o objetivo de garantir que indivíduos diferentes fiquem distantes na representação visual. As fórmulas utilizadas estão descritas nas Equações 2 e 4.

$$f_i = \sum_{j=1}^A a[i][j] \times j \quad (1)$$

$$x = \frac{\sum_{i=1}^P \times f_i}{P} \quad (2)$$

$$f_j = \sum_{i=1}^P a[i][j] \times i \quad (3)$$

$$y = \frac{\sum_{j=1}^A \times f_j}{A} \quad (4)$$

2.0.3. Representação

Para representar a passagem de gerações utilizamos uma paleta de cores, de forma que a cor mais forte representa a geração atual. A medida que as gerações vão passando, a cor do indivíduo é preenchida com um tom menor. O preenchimento é feito até que o indivíduo suma da tela. Essa abordagem foi utilizada para que a densidade de objetos não prejudicasse a extração de informação.

Para quantificar o *fitness* utilizamos o tamanho do objeto. Sendo assim, a medida que as gerações vão crescendo a tendência é que os objetos vão ficando maiores. Isso mostra que o algoritmo é capaz de melhorar os indivíduos a cada geração.

3. Conclusão

A proposta do nosso trabalho era visualizar a evolução das soluções geradas em uma execução de um AG, possibilitando o usuário verificar sua dispersão inicial e sua convergência ao passar das gerações. A visualização implementada cumpriu bem o seu papel, mostrando que na fase inicial o espaçamento entre os pontos é o maior possível. Com o passar das gerações, nós podemos ver que há aumento do *fitness* médio dos indivíduos e que os indivíduos se tornam mais semelhantes indicando uma convergência. O AG para quando a população estabiliza em um ponto onde não melhora o *fitness*.

O tutorial e a visualização do trabalho prático 3 estão disponíveis em:

Tutorial

Trabalho Prático 3