



UFOP

Algoritmos Genéticos para o PCV & Seleção de Recursos Baseada em Correlação Hierárquica (H-CFS)

Guilherme Oliveira Santos

2 de outubro de 2023



icea

Instituto de Ciências
Exatas e Aplicadas

DECSI

DEPARTAMENTO DE
COMPUTAÇÃO E SISTEMAS



UFOP

Sumário

1 GA para o problema do Caixeiro Viajante

► GA para o problema do Caixeiro Viajante

► H-CFS

► Referências Bibliográficas

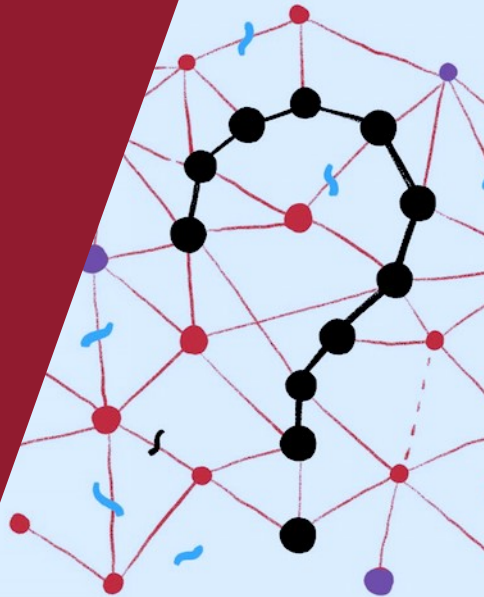


UFOP

Conteúdo

1 GA para o problema do Caixeiro Viajante

- Definição do cromossomo
- Cálculo de fitness
- Método de seleção
- Crossover e Mutação

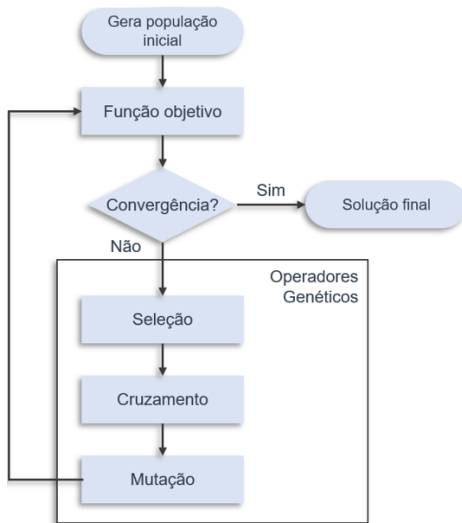




Definição do problema

1 GA para o problema do Caixeiro Viajante

Dado um conjunto de cidades com coordenadas representadas por coordenadas em um plano cartesiano, encontre a rota mais curta que visite todas as cidades utilizando um Algoritmo Genético.





Definição do cromossomo

1 GA para o problema do Caixeiro Viajante

Dado as cidades fictícias:

- Cidade A: (0, 0)
- Cidade B: (2, 4)

Foi feito um dicionário dessas cidades para facilitar a manipulação dos dados

Dictionary Cities

```
{  
1: (0,0),  
2: (2,4),  
...  
cidade n: (x,y) }
```



UFOP

Geração da população inicial

1 GA para o problema do Caixeiro Viajante

Com o dicionário, podemos facilmente criar uma lista (cromossomo) com a população inicial de possíveis caminhos de forma randômica.

Pseudocódigo da criação de um cromossomo

Inicio criarCromossomo

 cromossomo = lista vazia

 genesPossiveis = todas as cidades no dicionário

 Para i em intervalo(genesPossiveis):

 gene = escolher aleatoriamente um elemento de genesPossiveis

 Se gene não estiver em cromossomo:

 adicione gene a cromossomo

 remova o elemento no índice do último gene

6/26 retorne cromossomo



UFOP

Geração da população inicial

1 GA para o problema do Caixeiro Viajante

Com o cromossomo em mãos, repetimos o processo n vezes para a criação da matriz de população.

Pseudocódigo da criação da população

```
Inicio criarPopulacao
    populacao = [criarCromossomo em intervalo tamanho da população]
    retorne populacao
Fim criarPopulacao
```



Cálculo de Fitness

1 GA para o problema do Caixeiro Viajante

Para calcular o valor de aptidão (fitness) de um cromossomo, foi adaptado a fórmula que calcula a distância entre dois pontos de René Descartes:

$$\text{fitness}(\text{cromossomo}) = \sum_{j=0}^{n-1} \sqrt{(x_{j+1} - x_j)^2 + (y_{j+1} - y_j)^2} \quad (1)$$

- Onde, n é o tamanho do cromossomo



Seleção

1 GA para o problema do Caixeiro Viajante

- A seleção determina quais cromossomos terão a oportunidade de se reproduzir e contribuir com sua informação genética para a próxima geração. Existem várias estratégias, sendo as mais comuns: roleta e torneio.

No PCV aqui abordado, foi escolhido o método de torneio. Onde selecionamos um subconjunto aleatório da população, e indivíduos dentro deste torneio competem entre si, esse processo é repetido até que o número necessário de indivíduos seja escolhido.



Seleção

1 GA para o problema do Caixeiro Viajante

Pseudocódigo do torneio

```
Inicio torneio
  k = 0.75
  Repita n vezes
    Escolha 2 indivíduos aleatoriamente
    r = valor entre 0 e 1
    Se r < k
      o individuo com melhor fitness é escolhido
    Senão
      o individuo com pior fitness é escolhido
    Fim se
  Fim repita
Fim torneio
```

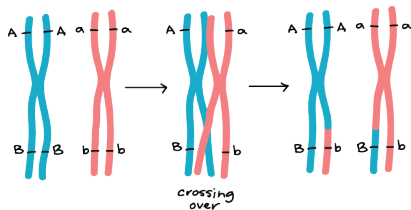


Crossover

1 GA para o problema do Caixeiro Viajante

O crossover é uma técnica inspirada no processo de evolução natural, responsável por combinar informações genéticas de indivíduos da população atual para criar novos indivíduos para a geração seguinte.

- O crossover escolhido aqui foi o PMX (Partial-Mapped Crossover) que é um método específico de crossover usado em GAs para combinar informações genéticas.





Crossover

1 GA para o problema do Caixeiro Viajante

Pseudocódigo do PMX (Partial-Mapped Crossover)

Início crossover

ponto1, ponto2 = Seleciona aleatoriamente dois pontos de corte

filho1 = copia os genes entre o corte de dois pontos dos pais

filho2 = copia os genes entre o corte de dois pontos dos pais

n = tamanho do pai

Repita n vezes:

Se ponto1 \leq n \leq ponto2:

continua

Senão:

mapping = Encontre um mapeamento do pai2 para filho1[n]

Enquanto mapping em filho1[ponto1:ponto2 + 1]:

index = recebe indice do mapping



Pseudocódigo do PMX (Partial-Mapped Crossover) -> Continuação

```
        mapping = pai2[index]
        filho1[i] = mapping

        # Repete o processo pro filho 2
    Fim senão
Fim Repita
returna filho1, filho2
Fim Crossover
```

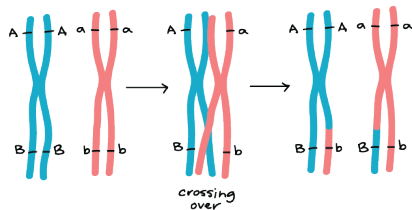


Mutação

1 GA para o problema do Caixeiro Viajante

A mutação em é um operador usado para introduzir variação genética em uma população, permitindo que o GA escape de mínimos locais e explore novas soluções.

- A função implementada é uma mutação simples em um cromossomo. Ela opera da seguinte maneira:





Mutação

1 GA para o problema do Caixeiro Viajante

Pseudocódigo de mutação

Início mutação

novoCromossomo = Cópia do cromossomo original

Se aleatório() < Probabilidade de mutação:

 índices = aleatório(2 índices aleatórios do cromossomo)

 index1, index2 = índices[0], índices[1]

 Troca os valores nos índices selecionados

Fim Se

 retorne novoCromossomo

Fim mutação



Algoritmo Genético Completo

1 GA para o problema do Caixeiro Viajante

Pseudocódigo

Inicio Algoritmo Genético

Inicialize a população

Avalie o fitness da população inicial

Encontre o melhor indivíduo na população inicial

Inicialize o vetor que armazenará o melhor fitness em cada geração

Repita i de 0 até numero_geracoes e/ou convergência:

 Seleção por torneio

 Crossover

 Mutaç o

 Substitua um indivíduo aleatório pelo melhor global (Elitismo)

 Avalie o fitness da nova população



Pseudocódigo -> Continuação

Encontre o melhor indivíduo na nova população
Armazene o melhor fitness local no vetor de histórico

Retorne o histórico do melhor fitness em cada geração
retorne melhor_vetor_global

Fim Algoritmo Genético

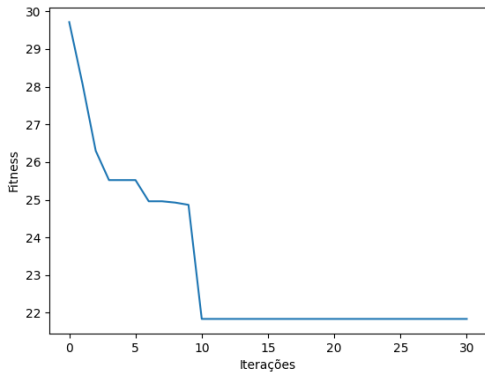


UFOP

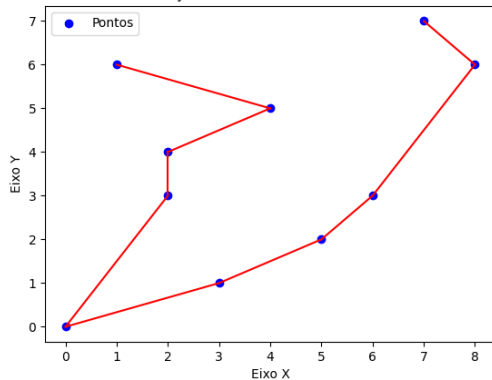
Exemplos de Output

1 GA para o problema do Caixeiro Viajante

Fitness: 21.84



Traçado de Linhas entre Pontos





UFOP

Sumário

2 H-CFS

► GA para o problema do Caixeiro Viajante

► H-CFS

► Referências Bibliográficas

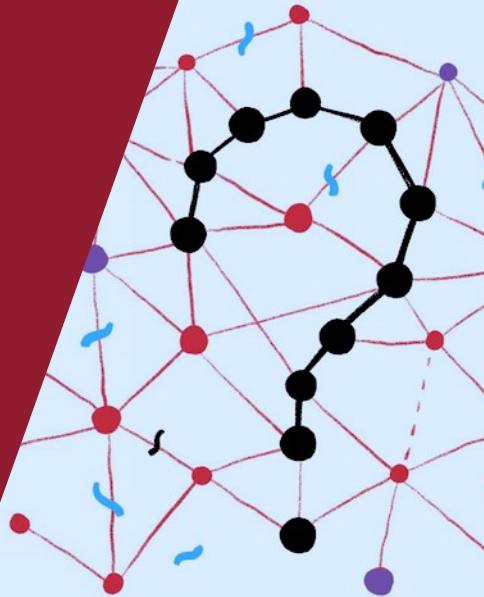


UFOP

Conteúdo

2 H-CFS

- Definição H-CFS & *Merit*





H-CFS

2 H-CFS

H-CFS (Hierarchical Correlation-based Feature Selection), faz a avaliação de um subconjunto de características, usando a equação abaixo:

$$\text{merit} = \frac{k_F \overline{r_{FL}}}{\sqrt{K_F + K_F(K_F - 1) \overline{r_{FF}}}} \quad (2)$$

- $\overline{r_{FL}}$ É a correlação média do rótulo-recurso,
- $\overline{r_{FF}}$ É a intercorrelação característica-característica média,
- F É o subconjunto de características avaliado,
- L É o conjunto de rótulos de classe,
- k_F é o número de características em F ,

E r , é o Coeficiente de Correlação de Pearson.



O termo: $\overline{r_{FF}}$

2 H-CFS

A equação abaixo descreve a correlação $\overline{r_{FF}}$, a correlação entre cada par de recursos (f_i, f_j) no conjunto de dados é calculada e, em seguida, os resultados são calculados dividindo a soma total de todos os resultados pelo número de pares de características, denotado por p .

$$\overline{r_{FF}} = \frac{\sum_{i,j=1, i>j}^{k_F} |r_{f_i f_j}|}{p} \quad (3)$$

**O termo: $\overline{r_{FL}}$**

2 H-CFS

A diferença entre o CFS e H-CFS é a forma que é estimada o termo $\overline{r_{FL}}$.

$$\overline{r_{FL}} = \frac{\sum_{f_i=1}^{k_F} |r_{fL}|}{k_F}$$

A ideia é calcular a correlação média do rótulo-atributo usando a aritmética média de todos os pares rótulo-atributo considerando a relação hierárquica entre as classes usando:

$$r_{fL} = \frac{\sum_{i=1}^h \sum_{j=1}^{k_{L_i}} w_i |r_{fL_i,j}|}{\sum_{i=1}^h w_i k_{L_i}}$$

- L_i É o set de classes no level i ,
- h É o número de níveis na hierarquia,
- k_{L_i} É o número de classes em L_i ,
- w_i é o peso atribuído ao nível i da hierarquia que é calculado como:

$$w_i = w_0^i, \text{ para } 0 \leq w_0 \leq 1 \rightarrow w_0 = 0.75$$

A ideia é que w_i diminui com a profundidade do nível i na hierarquia de classes.



Pearson's correlation coefficient

2 H-CFS

Para estimar a diferença entre duas características X, Y usamos o Coeficiente de Correlação de Pearson (r_{xy}). A equação abaixo descreve quando X e Y são duas variáveis contínuas:

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad \text{Onde,}$$

- r_{XY} representa o coeficiente de correlação entre as variáveis X e Y .
- n é o número de observações no set de treinamento.
- x_i e y_i são os valores das observações para X e Y respectivamente.
- \bar{x} e \bar{y} representam as médias das observações para X e Y respectivamente.



Pearson's correlation coefficient

2 H-CFS

- Quando uma característica é contínua e a outra discreta, calcula-se uma correlação de Pearson ponderada de acordo com as variáveis.
- Quando ambos os recursos envolvidos são discretos, Criamos recursos binários para ambos e calculamos todas as correlações ponderadas para todos combinações.
- Vale ressaltar que foram utilizados os valores absolutos do coeficiente de correlação em todas as ocorrências r_{XY} , pois tanto valores positivos quanto negativos podem representar redundância entre um par de características.



UFOP

Sumário

3 Referências Bibliográficas

► GA para o problema do Caixeiro Viajante

► H-CFS

► Referências Bibliográficas



UFOP

Referência Bibliográfica

3 Referências Bibliográficas



UFOP

Algoritmos Genéticos para o PCV & Seleção de Recursos Baseada em Correlação Hierárquica (H-CFS)

Obrigado!