

RELATÓRIO FINAL

**UTILIZAÇÃO DE ALGORÍTIMOS GENÉTICOS PARA DEFINIÇÃO E
AVALIAÇÃO DE ZONAS DE MANEJO**

Gabriel Silva Mesquita Souza

PIBIC

Bacharelado em Ciências da Computação - IC

UTILIZAÇÃO DE ALGORÍTIMOS GENÉTICOS PARA DEFINIÇÃO E AVALIAÇÃO DE ZONAS DE MANEJO

Relatório Final apresentado à Universidade Federal de Mato Grosso, Pró-Reitoria de Pesquisa, Programa Institucional de Bolsas de Iniciação Científica, sob orientação do Prof. Raul Teruel dos Santos/Instituto de Computação.

**Cuiabá
Agosto de 2020**

Sumário

| | |
|--|-----------|
| Introdução | 6 |
| Revisão Bibliográfica | 6 |
| Metodologia | 12 |
| Resultados e Discussões | 14 |
| Conclusões | 20 |
| Dificuldades encontradas e ações adotadas para superação das dificuldades | 21 |
| Referências Bibliográficas | 21 |
| Produtos gerados com a participação do bolsista | 23 |

Lista de Figuras

| | |
|---|-----------|
| Figura 1. Representação da codificação binária em vetor | 07 |
| Figura 2. Codificação binária utilizando matriz | 07 |
| Figura 3. Codificação em números inteiros representando os grupos | 07 |
| Figura 4. Codificação em números inteiros. representando as amostras que serão os centróides | 07 |
| Figura 5. Representação da codificação em números reais | 08 |
| Figura 6. Função Calinski Harabaz Score | 08 |
| Figura 7. Função de métrica de distância | 09 |
| Figura 8. Função soma dos erros ao quadrado | 09 |
| Figura 9. Cruzamento de ponto único | 11 |
| Figura 10. Função para cálculo da probabilidade dos operadores | 12 |
| Figura 11 Gráfico do treinamento Calinski Harabaz | 14 |
| Figura 12. Mapa do agrupamento utilizando Calinski Harabaz | 15 |
| Figura 13 Gráfico do treinamento David Bouldin | 16 |
| Figura 14. Mapa do agrupamento utilizando David Bouldin | 17 |
| Figura 15 Gráfico do treinamento FPC | 17 |
| Figura 16. Mapa do agrupamento utilizando FPC | 18 |
| Figura 17. Gráfico do treinamento Soma dos erros ao quadrado | 19 |
| Figura 18. Mapa do agrupamento utilizando Calinski Harabaz | 20 |

1. Introdução

A agricultura de precisão (AP) faz uso de gps e sensores, obtendo mais informações sobre a produtividade e distribuição de nutrientes no solo. A adoção da AP pode trazer contribuição para a produtividade e ganhos econômicos, além de tornar a produção mais sustentável Silva et al (2007), Bongiovanni e Lowenberg-Deboer (2004).

Para realizar a aplicação de insumos de forma eficiente a cultura, são definidas zonas de manejo (ZM). Estas são agrupamentos de talhões, menor unidade de produção agrícola, com características homogêneas.

As definições das ZM de forma automatizada podem ser feitas utilizando algoritmos de agrupamento Rodrigues e Cora (2015). Os algoritmos genéticos (AG) podem ser utilizados para a finalidade de agrupamento. Visando buscar a quantidade ideal de grupos e o melhor agrupamento para o problema proposto. (GIRSANG, TANZIL e UDJAJA 2016; HRUCHKA e EBECKEN 2003).

A adoção de AG em agrupamentos permite poupar recursos computacionais, visto que não irão explorar todas as possíveis soluções de tamanho de grupo/agrupamento. A busca quanto a quantidade ideal de grupos, permite a sua utilização por não-especialistas da área ao qual o conjunto de amostra pertence.

Este trabalho visa o uso dos algoritmos genéticos para o agrupamento de dados oriundos da agricultura para a definição de zonas de manejo.

2. Revisão de Literatura

A revisão de Literatura foi realizada, buscando artigos relacionados com agrupamentos utilizando algoritmos genéticos, os principais meios de busca foram o google scholar e a biblioteca digital de IEEE.

2.2. Cromossomo

O cromossomo é a estrutura de dados que irá ser trabalhada durante todo o processo do AG, nele contém as possíveis soluções do problema a ser solucionado. Na utilização de AG para agrupamentos, há três formas de montar sua estrutura, usando números binários, inteiros e reais.

2.2.1. Binário

Kuncheva e Bezdek (1997), propõe a utilização da codificação binária criando um vetor de tamanho igual a quantidade de amostras. As posições com valores 1 indicam

amostras que seriam centros de grupos e as com marcador 0 as que não foram utilizadas. A figura 1 mostra a codificação para 3 grupos em um total de 12 amostras.

Figura 1. Representação da codificação binária em vetor.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|

Fonte: Kuncheva e Bezdek (1997)

Bezdek et al. (1994) faz a utilização de uma matriz de tamanho $k.n$ (k é a quantidade de grupos do cromossomo e n é a quantidade de amostras do conjunto). Cada linha representa um grupo e as colunas as amostras. A linha que contém o valor 1 de uma coluna representa o grupo ao qual ela pertence. A Figura 2 exemplifica uma codificação para 12 amostras e 3 grupos.

Figura 2. Codificação binária utilizando matriz.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 |

Fonte: Bezdek et al.

2.2.2. Inteiro

Falkenauer (1998) utiliza os números inteiros para representar os grupos que as amostras pertencem. A Figura 3 mostra um exemplo da codificação. Hruschka e Ebecken (2003) utilizam essa codificação adicionando uma posição ao final para indicar a quantidade de grupos, que o cromossomo tem.

Figura 3. Codificação em números inteiros representando os grupos.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 3 | 1 | 2 | 3 | 2 | 2 | 2 | 3 | 3 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|

Fonte: Falkenauer (1998)

Uma segunda forma de codificação é a utilização da posição das amostras que são centros de grupos. A figura 4 ilustra um cromossomo que indica 3 grupos e as amostras 3, 8 e 11 como seus centros de grupos. Sheng e Liu (2004).

Figura 4. Codificação em números inteiros, representando as amostras que serão os centróides..

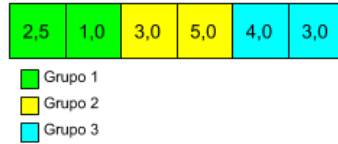
| | | |
|---|---|----|
| 3 | 8 | 11 |
|---|---|----|

Fonte: Sheng e Liu (2004)

2.2.3. Real

Esta codificação utiliza os valores numéricos dos centros de grupos para compor a sua estrutura, formando vetor de tamanho **n.k** (**n** é a dimensão das amostras e **k** é a quantidade de grupos).

Figura 5. Representação da codificação em números reais.



Fonte: Autor

2.3. População inicial

A população é formada por um grupo de cromossomos, ela permite que o AG possua uma variedade de possíveis soluções.

(LUCASIU et al. 1993; KUNCHEVA e BEZDEK 1997; BANDYOPADHYAY e MAULIK 2002) Iniciam a população usando os valores de amostras aleatórias do conjunto de amostra para compor os cromossomos.

(LU e JU 2012; BEZDEK et al. 1994), iniciam as populações gerando os indivíduos de forma aleatória.

2.4. Função de aptidão

A função de aptidão é utilizada para avaliar cada cromossomo, o seu valor resultante será usado pelo AG para fazer a seleção dos cromossomos e determinar a melhor solução.

2.4.1. Calinski Harabaz Score

A função foi proposta por Calinski e Harabaz (1974) para a avaliação de agrupamentos. Maiores valores indicam que as amostras estão próximas aos seus centros de grupos e afastadas do centro do conjunto de amostras.

Figura 6. Função Calinski Harabaz Score

$$CH(s) = \frac{\sum_{i=1}^N \|X_i - m_s\|^2 - \sum_{j=1}^K \sum_{X \in c_{sj}} \|X - m_{sj}\|^2}{\sum_{j=1}^K \sum_{X \in c_{sj}} \|X - m_{sj}\|^2} \frac{N - K_S}{K_S - 1}$$

Fonte: Calinski e Harabaz (1974)

Na equação **s** representa um solução, **N** é a quantidade de amostras, **X_i** é a amostra

da posição i no conjunto de amostras, K_s é a quantidade de grupos C_{sj} é o grupo j da interação e M_{sj} é o centro do grupo C_{sj} .

2.4.2. Métricas de distância

Lucasius et al (1993) faz a utilização de métricas de distância, visando minimizar a distância das amostras em relação ao seu centro de grupo. A figura 7 mostra a função.

Figura 7. Função de métrica de distância

$$F = \sum_{i=1}^N d(x_i, m)$$

Fonte: Lucasius et al (1993)

N é a quantidade de amostras, d é a função de distância, x_i é a amostra que está a posição i , m é o centro do grupo que a amostras x_i pertence. Bandyopadhyay e Maulik (2002) utiliza a distância euclidiana como função de aptidão.

2.4.2. Soma dos erros ao quadrado

Lu e Ju(2012) faz a utilização da função de soma dos erros ao quadrado, para a utilização.

Figura 8. Função soma dos erros ao quadrado

$$E = \sum_{j=1}^k \sum_{x \in C_j} \|x - m_j\|^2$$

Fonte: Lu e Ju (2012)

k é o número de clusters x é amostra m_j é o centro do grupo j .

A soma dos erros ao quadrado sempre vai convergir para o maior número de grupos, para corrigir isso foi utilizada uma penalização linear, multiplicando o resultado da soma pela quantidade de grupos.

2.5. Seleção

A seleção é utilizada para selecionar os cromossomos que irão sofrer o cruzamento e a mutação e os cromossomos que irão avançar para a geração seguinte.

2.5.1. Elitismo

O elitismo visa selecionar os melhores indivíduos da geração. A população é ordenada de forma crescente e são selecionados os n primeiros indivíduos, quando for problemas de minimização e os n últimos quando maximização. O valor de n é especificado na execução do AG. Kuncheva e Bezdek (1997).

Lu e Ju (2012), faz uso da seleção para impedir que os melhores indivíduos da geração passem pelos operadores de cruzamento e mutação, preservando-os para a próxima geração.

2.5.2. Roleta

A seleção por roleta, faz uso de probabilidade para selecionar os cromossomos. Para cada cromossomo é atribuída uma probabilidade de ser selecionado, dividindo sua aptidão pela soma das aptidões dos cromossomos da população. A população é ordenada de forma decrescente e um valor aleatório é gerado para ser a probabilidade inicial de seleção. Enquanto a probabilidade de seleção não for maior que o cromossomo atual da interação, some a probabilidade de seleção a probabilidade do cromossomo. O cromossomo que satisfizer a seleção será o selecionado. Esta forma de seleção foi utilizada por Mortezaee e Daneshifar (2019), Lucasius et al (1993), Bandyopadhyay e Maulik (2002).

2.5.3. Torneio

A seleção por torneio consiste em selecionar uma quantidade de indivíduos da população e selecionar o melhor indivíduo do grupo dos selecionados. Quando o número de indivíduos selecionados for 1 essa seleção passa a ser aleatória e quando é igual ao tamanho da população é uma seleção equivalente ao elitismo.

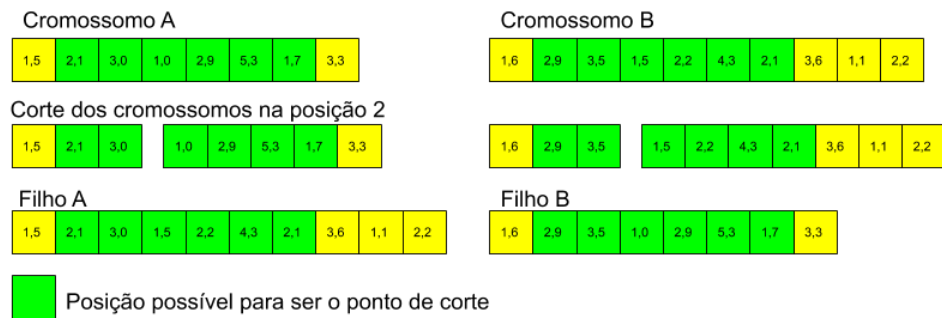
2.6. Cruzamento

O cruzamento faz a troca de informações entre dois cromossomos chamados de cromossomos pais, criando um ou mais cromossomos novos, chamado de cromossomos filhos.

2.6.1. Ponto único

O Cruzamento de ponto único define uma posição aleatória dos cromossomos pais e realiza a troca do segmento do cromossomo após a posição. Sanghamitra e Maulik (2002) propôs a utilização do limite $[1, L-1]$ para a posição do ponto de corte, sendo L posição final do cromossomo, para a posição do ponto de corte, codificação mostrada na figura 9 . Garantindo que os cromossomos filhos possuam dados dos dois pais. Sheng e Liu (2004) também adiciona mutação ao processo do cruzamento, para evitar que haja indivíduos iguais nas futuras gerações.

Figura 9. Cruzamento de ponto único.



Fonte: Autor

2.6.2. Merge

Mortezanezhad e Daneshifar (2019), faz a união dos cromossomos pais, resultando em um cromossomo filho. Os centros de grupos que estão próximos são unidos formando um único grupo.

2.6.3. Caminho

Chang, Zhang e Zheng (2009) faz a seleção de dois pais, entre os pais são aplicadas funções de maximização e minimização entre as partes dos cromossomos pais, criando os cromossomos filhos.

2.7. Mutação

A mutação faz alterações no cromossomo criando um cromossomo novo.

2.7.1 Alteração de valores

A alteração de valores, modifica um ou mais valores do cromossomo original criando um novo.

2.7.2. Adição ou remoção de grupos

Mortezanezhad e Danishifar (2019) faz uso de uma mutação que adiciona novo grupo ou remove um do cromossomo. Na adição de um grupo novo, é selecionada uma amostra aleatória do conjunto de amostras, para compor o centro do novo grupo.

2.7.3. Mutação com K-Means

Lu e Ju (2012), faz uso do K-means para a criação de novos cromossomos. Os centros de grupos da população atual são utilizados no k-means para serem clusterizados, os centros gerados pelo k-means são inseridos na nova população.

2.8 valor adaptativos dos operadores

Mortezanezhad e Daneshifar (2019), faz a utilização de valores adaptativos para a probabilidade de mutação e cruzamento, permitindo que conforme o AG vai aproximando da

sua geração final ele amplie sua área de busca, através do aumento da probabilidade dos operadores.

Figura 10. Função para cálculo da probabilidade dos operadores

$$P_m(n) = P_{mi} + \left(\frac{n}{T}\right)(P_{mf} - P_{mi})$$

Fonte: Mortezaezhad e Daneshifar (2019)

Na equação da figura 10, P_m é a probabilidade do operador a geração n , P_{mi} é o valor inicial da probabilidade do operador, T é o número total de geração e P_{mf} é probabilidade final do operador.

2.9 Avaliação de resultados

Foi utilizada a pontuação de informação mútua para comparação dos agrupamentos. O método consiste em comparar dois resultados de agrupamentos considerando a possibilidade de o rótulo dos agrupamentos ser diferente. (Vinh e Bailey, 2009)

3. Metodologia

O desenvolvimento da iniciação científica consistiu primeiramente de uma revisão da literatura, após a revisão foi feita análise dos operadores para determinar qual utilizar no AG.

3.1. Seleção dos operadores

3.1.1 Cromossomo

As codificações de vetor binário, inteiro indicando a amostra no qual é o centro de grupos e codificação do tipo real, possuem o mesmo objeto codificado, os centros dos clusters. Sendo possível adequar a utilização de um operador genético usado em uma as demais.

Primeiramente analisando o consumo de memória das três codificações tem-se: binária consumindo **n**, inteiro **k** e real **k.d**, sendo **n** a quantidade de amostras a serem agrupadas, **k** a quantidade de grupos que o cromossomo representa e **d** a dimensão dos dados das amostras.

Em segundo o espaço de busca que as codificações permitem. tanto a codificação binária quanto a de inteiros limita o espaço de busca aos valores das amostras, pois as suas codificações indicam as posições das amostras. A codificação real por fazer cópias dos valores, permite que o espaço de busca seja ampliado, permitindo que os operadores alterem os valores do cromossomo sem comprometer as amostras. Logo para encontrar a possível solução global a codificação do tipo real é melhor.

As duas codificações restantes utilizam-se da solução do agrupamento como cromossomo. Assim como as representações de centros de grupos a codificação em inteiro e matriz binária, conseguem fazer uso dos operadores, fazendo pequenas alterações.

A codificação em inteiro possui um custo de memória de n e a matriz de binários $k.n$. Entre ambas, a utilização de codificação do tipo inteiro necessita menos utilização da memória. E em operações onde troca-se o grupo no qual a amostra pertence a codificação em inteiros altera somente a posição no vetor, enquanto na binária deve-se modificar a coluna da matriz.

Tanto a codificação do tipo real quanto a de inteiros representando os agrupamentos, são passíveis para serem utilizados como codificação. A priori foi selecionado a codificação do tipo real para a execução do AG, pois a codificação com números inteiros quando o tamanho de grupos é variável, há a necessidade de adaptar os valores dos cromossomos quando os operadores genéticos forem influenciar a quantidade original de cluster do cromossomo.

3.1.2 População inicial e seleção

A população inicial foi criada através da seleção de amostras aleatórias e a seleção utilizando o método de torneio.

3.1.2 Cruzamento e mutação

O operador de cruzamento utilizado foi o de ponto único, pois permite a troca de informações dos cromossomos independente da quantidade de grupos.

A mutação utilizada é a adição ou remoção de grupos, pois esta permite haver variação na quantidade de grupos durante a execução do Algoritmo genético.

3.2. Experimento

O desenvolvimento do AG foi feito em python 3 fazendo uso da biblioteca numpy, para uso de vetores, a scikit-learn para função de agrupamento e a matplotlib para a visualização de gráficos. Código Fonte disponível no apêndice. O algoritmo foi sendo executado para os ajustes nos valores das probabilidades.

3.3. Conjunto de amostras

Os dados foram retirados de um talhão de 12,5 ha de uma fazenda de Diamantino-MT e interpolados com krigagem ordinária resultados em 100792 amostras contendo valores de Fósforo (P) e Manganês (Mn) e as coordenadas de cada amostra. Somente os valores de P e Mn foram utilizados para o agrupamento, as coordenadas são utilizadas para a representação

visual no mapa de zonas de manejo.

Os dados antes da execução dos testes foi feita a transformação dos valores para a escala de $[0, 1]$ o quadro 1 mostra o valor médio das amostras e variância após a aplicação da transformação.

Quadro 1. Descrição das média e variância das amostras.

| | P | Mn |
|------------------|------------|------------|
| Média | 0.47471216 | 0.48024922 |
| Variância | 0.01210357 | 0.02620903 |

Fonte: Autor

3.3. Parâmetros

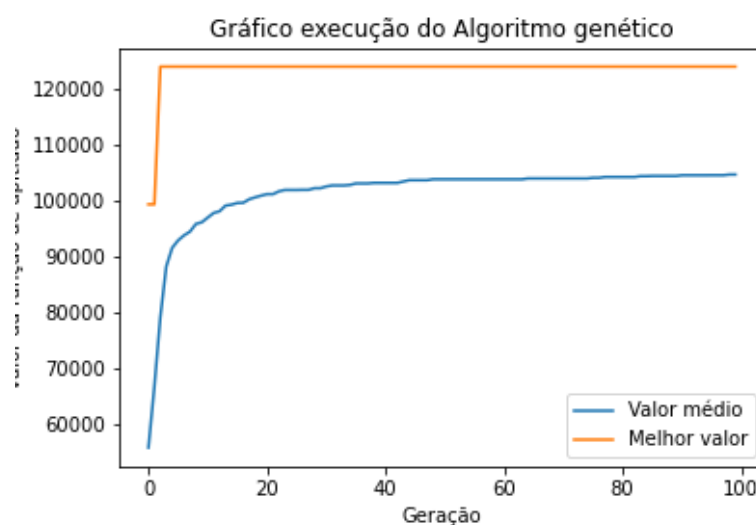
Foi utilizada uma população inicial de tamanho 40, 100 gerações, probabilidade de cruzamento 0.5, a probabilidade de mutação foi 0.1. E tamanho máximo de grupos foi de 10.

Os valores dos parâmetros foram obtidos através de tentativas anteriores, onde observava-se a convergência da função de aptidão para ajustar os parâmetros.

4. Resultados e Discussões

O primeiro teste foi executado com a função de Calinski Harabaz, a curva da fitness média inicialmente teve um alto crescimento e um pequeno crescimento após a geração 20 como mostrou a figura 11. Ao final do treinamento foi selecionado um indivíduo com 2 grupos.

Figura 11 Gráfico do treinamento Calinski Harabaz



Fonte: Autor

O grupo 1 reduz o valor de variância em relação ao conjunto de amostra tanto nos valores de P e Mn. no grupo 2 reduz a variância de Mn, porém no valor de P há um aumento. Como descrito no quadro 2

Quadro 2. Descrição dos agrupamentos utilizando Calinski Harabaz

| | P | Mn |
|---------------------------------------|------------|------------|
| Grupo 1 | 0.45940267 | 0.5991299 |
| Grupo 2 | 0.49227741 | 0.34385228 |
| Variância grupo 1 | 0.0075208 | 0.00886422 |
| Variância grupo 2 | 0.01693276 | 0.01139389 |
| Quantidade de amostras grupo 1 | 56601 | |
| Quantidade de amostras grupo 2 | 44191 | |

Fonte: Autor

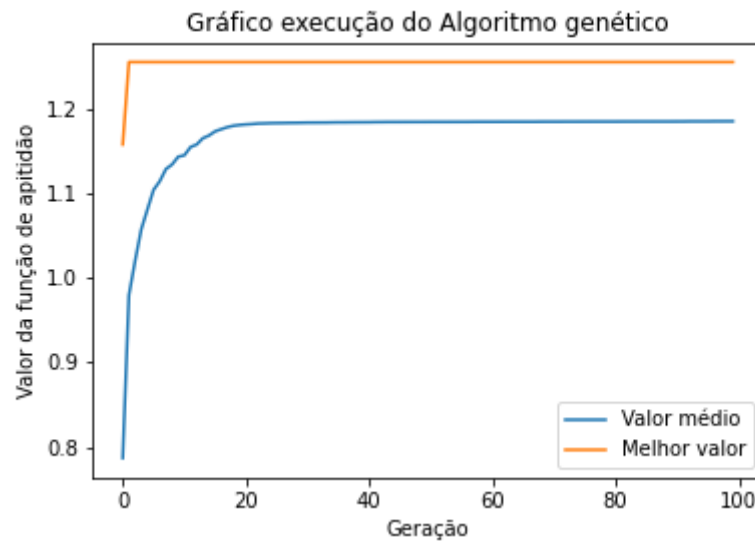
Figura 12. Mapa do agrupamento utilizando Calinski Harabaz



Fonte: Autor

O segundo teste foi com a função David Boudins, como mostrado na figura 13 a valor médio da aptidão dos indivíduos teve um alto crescimento nas gerações iniciais e estabilizou-se a partir da geração 20.

Figura 13 Gráfico do treinamento David Bouldin



Fonte: Autor

No grupo 1 houve redução no valor de variância dos valores de P e Mn e no grupo 2 houve a redução da variância no valor de Mn e aumento no valor de P. como mostrado no Quadro 3.

Quadro 3. Descrição do agrupamentos utilizando David Boudin

| | P | Mn |
|---------------------------------------|------------|------------|
| Grupo 1 | 0.4446251 | 0.56240258 |
| Grupo 2 | 0.5375479 | 0.30867495 |
| Variância grupo 1 | 0.00789707 | 0.0106519 |
| Variância grupo 2 | 0.0169881 | 0.01058563 |
| Quantidade de amostras grupo 1 | 66008 | |
| Quantidade de amostras grupo 2 | 34784 | |

Fonte: Autor

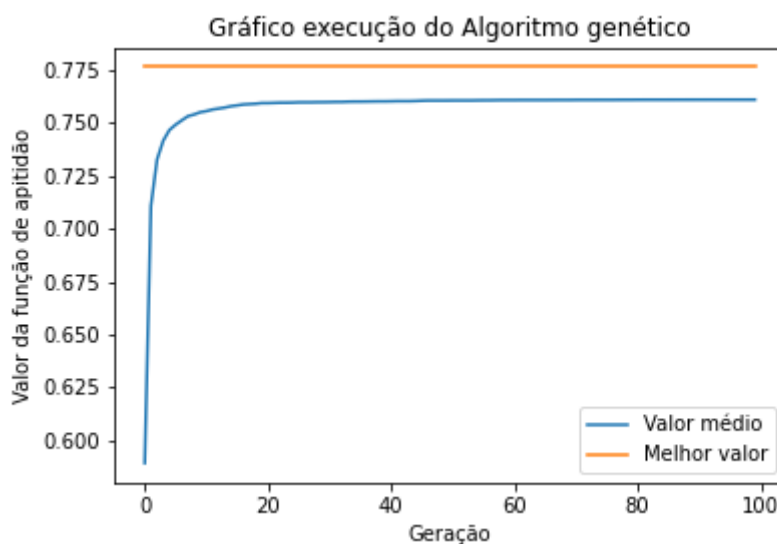
Figura 14. Mapa do agrupamento utilizando David Bouldin



Fonte: Autor

O terceiro teste foi utilizando a função FPC, a função apresentou um crescimento muito abrupto do valor médio de aptidão nas 10 primeiras gerações. Ao final do treinamento foi selecionado um indivíduo com 2 grupos. A figura 15 mostra a execução do algoritmo.

Figura 15 Gráfico do treinamento FPC



Fonte: Autor

No grupo 1 houve redução no valor de variância dos valores de P e Mn e no grupo 2 houve a redução da variância no valor de Mn e aumento no valor de P. como mostrado no Quadro 4.

Quadro 4. Descrição do agrupamentos utilizando FPC

| | P | Mn |
|---------------------------------------|------------|------------|
| Grupo 1 | 0.47324714 | 0.5444503 |
| Grupo 2 | 0.47991753 | 0.25213632 |
| Variância grupo 1 | 0.00827557 | 0.01108624 |
| Variância grupo 2 | 0.02043501 | 0.00892906 |
| Quantidade de amostras grupo 1 | 71500 | |
| Quantidade de amostras grupo 2 | 29292 | |

Fonte: Autor

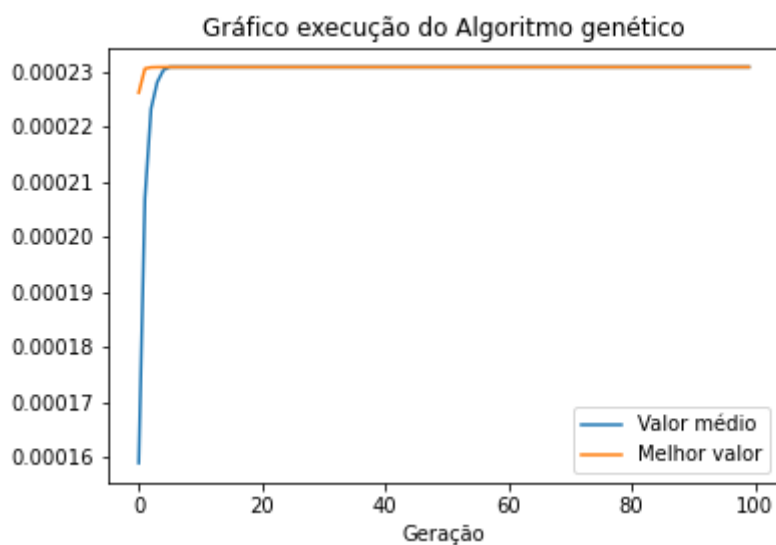
Figura 16. Mapa do agrupamento utilizando FPC



Fonte: Autor

O quarto teste foi com a soma dos erros ao quadrado, também apresentando um crescimento muito alto do valor médio de aptidão e próximo à geração 10 alcançando o valor do melhor indivíduo como mostrado na figura 17. A função também levou o algoritmo a encontrar indivíduo com 2 grupos.

Figura 17. Gráfico do treinamento Soma dos erros ao quadrado



Fonte: Autor

No grupo 1 houve redução no valor da variância de Mn e aumento no valor de P, no grupo 2 houve redução nos valores de P e Mn. Como descrito no Quadro 5.

Quadro 5. Descrição dos agrupamentos utilizando Soma dos erros ao quadrado.

| | P | Mn |
|---------------------------------------|------------|------------|
| Grupo 1 | 0.50464505 | 0.31781568 |
| Grupo 2 | 0.45625699 | 0.58039791 |
| Variância grupo 1 | 0.01840778 | 0.01051327 |
| Variância grupo 2 | 0.00732369 | 0.00958901 |
| Quantidade de amostras grupo 1 | 38442 | |
| Quantidade de amostras grupo 2 | 62350 | |

Fonte: Autor

Figura 18. Mapa do agrupamento utilizando Calinski Harabaz



Fonte: Autor

Quadro 6. Comparação dos agrupamentos.

Utilizando a pontuação de informação mútua foi comparado os agrupamentos de cada função de aptidão e do k-means entre si. A soma dos erros ao quadrado foi a que possui maior semelhança entre as demais e a FPC a que maior diverge das demais.

Quadro 6. Comparação das funções de aptidão.

| | DB | FPC | SSE | KMeans |
|---------------|-----------|------------|------------|---------------|
| CH | 0.6251 | 0.5006 | 0.7335 | 0.7337 |
| DB | --- | 0.6238 | 0.7781 | 0.7775 |
| FPC | --- | --- | 0.62076 | 0.62064 |
| SSE | --- | --- | --- | 0.9965 |
| KMeans | --- | --- | --- | --- |

Fonte: Autor

As funções de aptidões, com exceção da FPC, tiveram resultados próximos tanto visualmente nos mapas gerados quanto na função de avaliação. E em comparação com o KMeans a função de soma dos erros aos quadrados teve um bom resultado ficando próximo de gerar resultados idênticos. E a Calinski Harabaz teve um particionamento das amostras mais equilibradas.

5. Conclusões

O AG é utilizado para a resolução de problemas de busca fazendo uso de maximização/minimização de prováveis soluções. A sua utilização no agrupamento pode ser feita para encontrar um tamanho de grupo ideal e prováveis resultados de agrupamentos.

A sua execução dá-se pela criação de uma população de cromossomos, após a definição da população é aplicado operador de cruzamento, que consiste em fazer a troca de informações entre dois cromossomos, após o cruzamento os indivíduos da população passam pelo operador de mutação, que visa fazer alteração no cromossomo do indivíduo selecionado. Após aplicar os operadores, é selecionado os próximos indivíduos para compor a geração seguinte e repete-se o processo até atingir um número determinado de gerações.

A elaboração do AG neste período de pesquisa, consistiu na utilização de cromossomos com codificação do tipo real, cruzamento de ponto único e mutação foi adição e remoção de grupos. A inicialização da população foi através da seleção de amostras aleatórias. Para o operador de seleção foi utilizado a seleção por roleta para a escolha do cromossomo para cruzamento e o elitismo para selecionar indivíduos a serem preservados para a geração seguinte.

A execução do AG foi feita com amostras de Mn e P, apresentando agrupamentos que possuem a variância de cada grupo menor que a variância do conjunto de amostras. Nas amostras de Mn, houve um grupo que houve aumento na variância e os demais houve redução.

O AG permite trabalhar com o agrupamento de tamanhos variados durante a sua execução, possibilitando encontrar uma melhor quantidade adequada de ZM. A redução da variância após o agrupamento, permite a aplicação de insumos de forma homogênea na ZM, permitindo a economia de recursos na sua aplicação.

6. Dificuldades encontradas e ações adotadas para superação das dificuldades

A dificuldade encontrada foi a leitura de documentos em língua estrangeira visto que a maioria das publicações estão em inglês.

7. Referências Bibliográficas

AIBINU, A. et al. A novel Clustering based Genetic Algorithm for route optimization. **Engineering Science**

- and Technology, an International Journal**, v. 19, n. 4, p. 2022–2034, 2016.
- BANDYOPADHYAY, S.; MAULIK, U. An evolutionary technique based on K-Means algorithm for optimal clustering in RN. **Information Sciences**, v. 146, n. 1-4, p. 221–237, 2002.
- BEZDEK, J. et al. Genetic algorithm guided clustering. Proceedings of the First IEEE Conference on Evolutionary **Computation. IEEE World Congress on Computational Intelligence**, [s.d.].
- BONGIOVANNI, R.; LOWENBERG-DEBOER, J. Precision Agriculture and Sustainability. **Precision Agriculture**, v. 5, n. 4, p. 359–387, 2004.
- CALINSKI, T.; HARABASZ, J. A Dendrite Method for Cluster Analysis. **Communications in Statistics - Simulation and Computation**, v. 3, n. 1, p. 1–27, 1974.
- CHANG, D.-X.; ZHANG, X.-D.; ZHENG, C.-W. A genetic algorithm with gene rearrangement for K-means clustering. **Pattern Recognition**, v. 42, n. 7, p. 1210–1222, 2009.
- GIRSANG, A. S.; TANZIL, F.; UDJAJA, Y. Robust adaptive genetic K-Means algorithm using greedy selection for clustering. **2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS)**, 2016.
- GIRSANG, A. S.; TANZIL, F.; UDJAJA, Y. Robust adaptive genetic K-Means algorithm using greedy selection for clustering. **2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS)**, 2016.
- HRUSCHKA, E. R.; EBECKEN, N. F. A genetic algorithm for cluster analysis. **Intelligent Data Analysis**, v. 7, n. 1, p. 15–25, 2003.
- HRUSCHKA, E. et al. A Survey of Evolutionary Algorithms for Clustering. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, v. 39, n. 2, p. 133–155, 2009.
- LU, B.; JU, F. An optimized genetic K-means clustering algorithm. **2012 International Conference on Computer Science and Information Processing (CSIP)**, 2012.
- LUCASIUS, C.; DANE, A.; KATEMAN, G. On k-medoid clustering of large data sets with the aid of a genetic algorithm: background, feasibility and comparison. **Analytica Chimica Acta**, v. 282, n. 3, p. 647–669, 1993.
- LUDMILA I. KUNCHEVA, & JAMES C. BEZDEK. (1997). Selection Of Cluster Prototypes From Data By A Genetic Algorithm.
- MORTEZANEZHAD, A.; DANESHIFAR, E. Big-Data Clustering with Genetic Algorithm. **2019 5th**

Conference on Knowledge Based Engineering and Innovation (KB EI), 2019.

MORTEZANEZHAD, A.; DANESHIFAR, E. Big-Data Clustering with Genetic Algorithm. **2019 5th**

Conference on Knowledge Based Engineering and Innovation (KB EI), 2019.

SHENG, W.; LIU, X. A hybrid algorithm for k-medoid clustering of large data sets. **Proceedings of the 2004**

Congress on Evolutionary Computation (IEEE Cat. No.04TH8753), [s.d.].

VINH, N. X.; EPPS, J.; BAILEY, J. Information theoretic measures for clusterings comparison.

Proceedings of the 26th Annual International Conference on Machine Learning - ICML

'09, 2009.