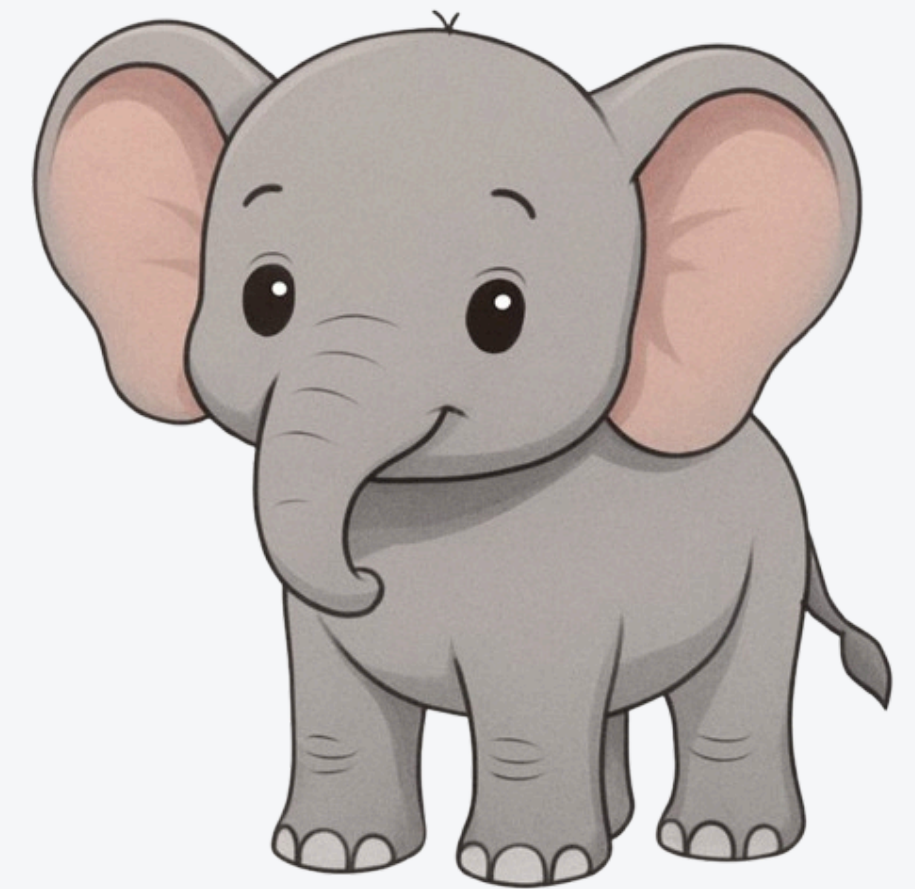


Elephant Herding Optimization

Método de otimização baseado no
comportamento de elefantes

Ana Cláudia Machado, Davi dos Reis de Jesus e Gabriel Silva Prenassi



1. Elephant Herding Optimization (EHO)

- **Algoritmo de Inteligência de Enxame:** baseado no comportamento coletivo dos elefantes
 - Elefantes vivem em **clãs**, liderados por uma **matriarca**
 - Machos se **separam** do clã ao crescerem
- **Elephant Herding Optimization [1]:**
 - Indivíduos divididos em **clãs**
 - **Matriarca** é o melhor indivíduo do clã (**melhor fitness**)
 - A cada **geração**, um número fixo de indivíduos (**piores fitness**) são separados do clã

2. Questões de projeto

- Função objetivo
- Representação real
- Dois operadores:
 - Atualização de clã (parâmetros α e β)
 - Separação

2.1. Função objetivo

- Minimização
- Utilizadas 15 funções benchmark da literatura, incluindo função de **Ackley**

$$f(x) = -20\epsilon^{-0.2} \sqrt{\frac{1}{n} \sum x_i^2} - \epsilon^{\frac{1}{n}} \sum \cos(2\pi x_i) + 20 + \epsilon$$

- $X = \{x_1, x_2, \dots, x_n\}$ é a posição do elefante no espaço n-dimensional

2.2. Operador de Atualização de Clã

- Atualiza a posição de todos os elefantes em um clã
- As novas posições dos elefantes são influenciadas pela posição da matriarca:

$$x_{new,ci,j} = x_{ci,j} + \alpha \times (x_{best,ci} - x_{ci,j}) \times r$$

- $x_{new,ci,j}$: nova posição do elefante j no clã ci
- $x_{ci,j}$: posição anterior do elefante j no clã ci
- $x_{best,ci}$: posição da matriarca do clã ci
- α : determina o fator de influência, no intervalo $[0, 1]$, da matriarca sobre o elefante j
- r : número aleatório, no intervalo $[0,1]$, com distribuição uniforme

2.2. Operador de Atualização de Clã

- A matriarca também é atualizada, mas utilizando outra equação:

$$x_{new,ci,j} = \beta \times x_{center,ci}$$

- $x_{new,ci,j}$: nova posição da matriarca
- β : determina o fator de influência, no intervalo $[0, 1]$, do centro do cluster na nova posição da matriarca
- centro do *cluster*, considerando a posição de todos os elefantes e as D dimensões:

$$x_{center,ci,d} = \frac{1}{n_{ci}} \times \sum_{j=1}^{n_{ci}} x_{ci,j,d}$$

2.3. Operador de Separação

- Simulando partida dos elefantes machos na puberdade
- Indivíduo com **pior fitness** ganha uma nova posição **aleatória** no intervalo de possíveis posições

$$x_{worst,ci} = x_{\min} + (x_{\max} - x_{\min} + 1) \times rand$$

- $x_{worst,ci}$: nova posição do antigo pior elefante de ci
- x_{\max} : limite superior das possíveis posições dos elefantes
- x_{\min} : limite inferior das possíveis posições dos elefantes
- $rand$: valor aleatório no intervalo $[0,1]$, seguindo uma distribuição uniforme

3. Implementação

- Representação
 - Real
- Estrutura para indivíduos e *fitness*
- Parametrização
 - Número de clãs
 - Número de elefantes em cada clã
 - **default:** mesmo número de elefantes em cada clã
 - Número de gerações
 - Alfa
 - Beta

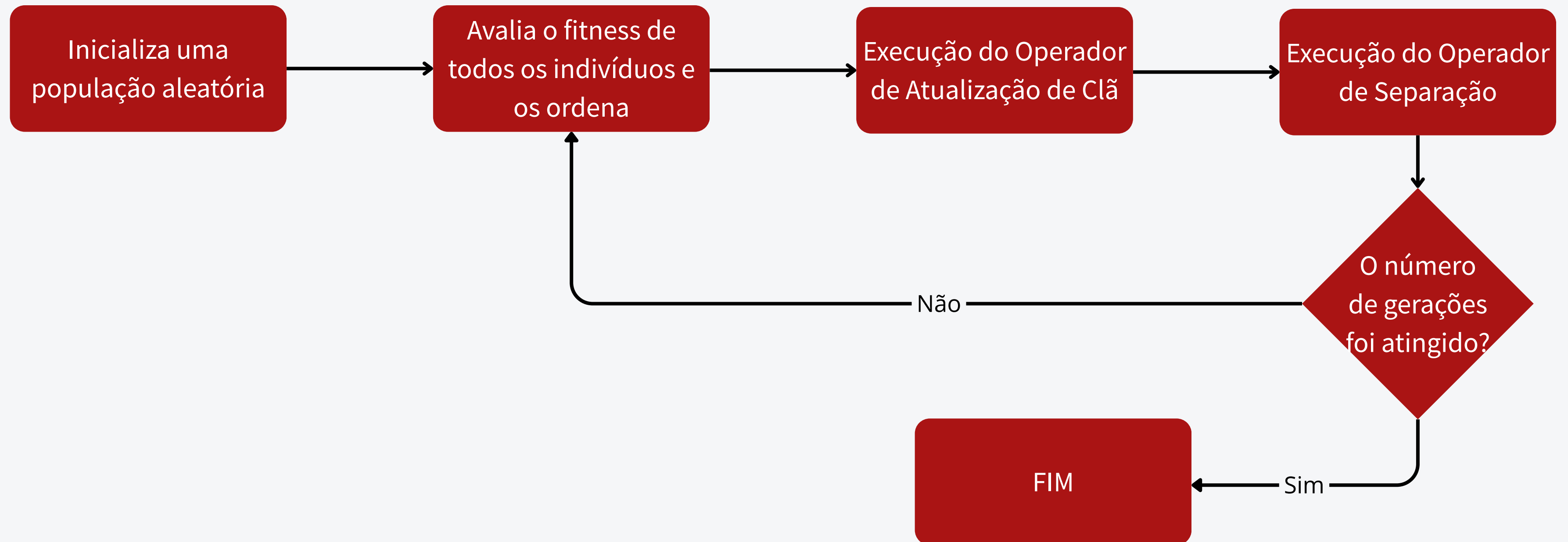
4. Testes

- Parâmetros:
 - Alfa: 0.5
 - Beta: 0.1
 - Número de Gerações: 50
 - Número de Clãs: 5
 - Tamanho da População: 50 (10 indivíduos por clã)
 - Número de parâmetros: 3

4. Testes

- Utilizando 100 execuções
- Melhor valor obtido:
 - EHO (paper): $1.3\text{E-}3$
 - EHO (reprodução): $1.35\text{E-}3$

5. Fluxograma

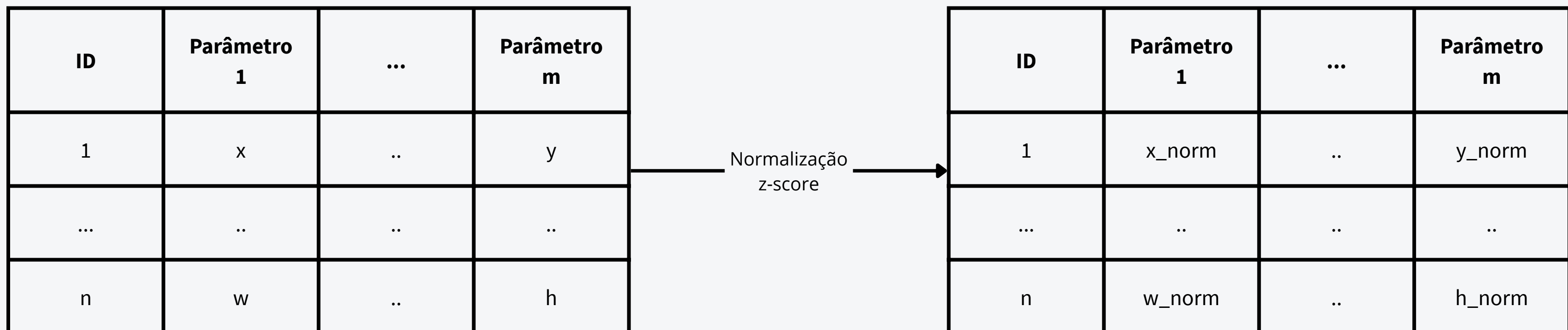


6. Aplicação

- Título: An Optimization K-Modes Clustering Algorithm with Elephant Herding Optimization Algorithm for Crime Clustering (Journal of Advances in Computer Engineering and Technology, 2020) [2]
- Proposta de um modelo híbrido baseado no algoritmo K-Modes (clusterização) e EHO para analisar o agrupamento de crimes com o objetivo de detectar similaridades entre eles.
- EHO é utilizado para determinar a quantidade de centros e definir suas posições, que serão empregadas no algoritmo K-Modes.

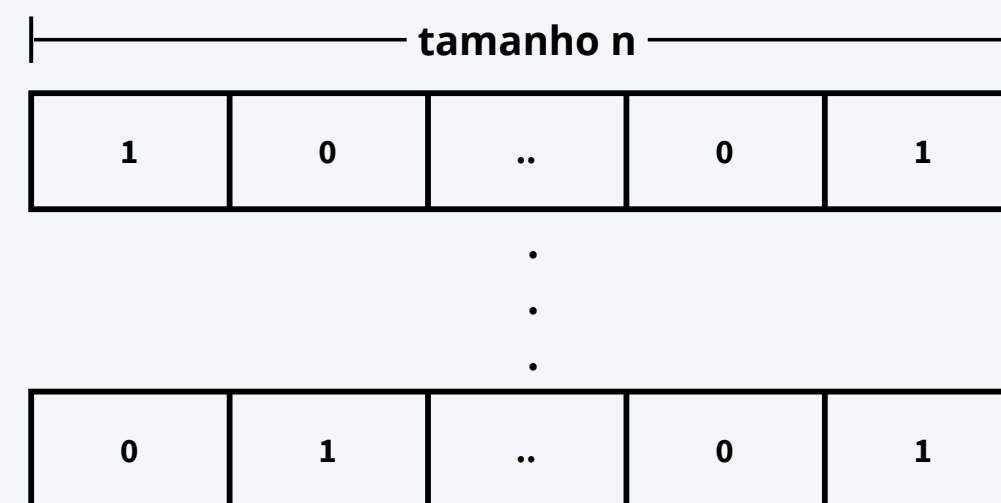
6.1. Modelo Proposto

- Passo 1 - Normalização dos Dados
 - A normalização ajusta os dados para uma mesma faixa de valores, evitando desequilíbrios que podem afetar o desempenho da função de fitness.



6.1. Modelo Proposto

- Passo 2 - Execução do EHO
 - População: cada elefante é um vetor binário que indica quais amostras do conjunto de dados serão usadas como centros de cluster.



- Fitness: distância intra-cluster entre os centros dos clusters.
- Parâmetros utilizados: número de elefantes igual a 50, apenas 1 clã, α igual a 0.5, β igual a 0.6 e número de gerações variável (100 ou 200).

6.1. Modelo Proposto

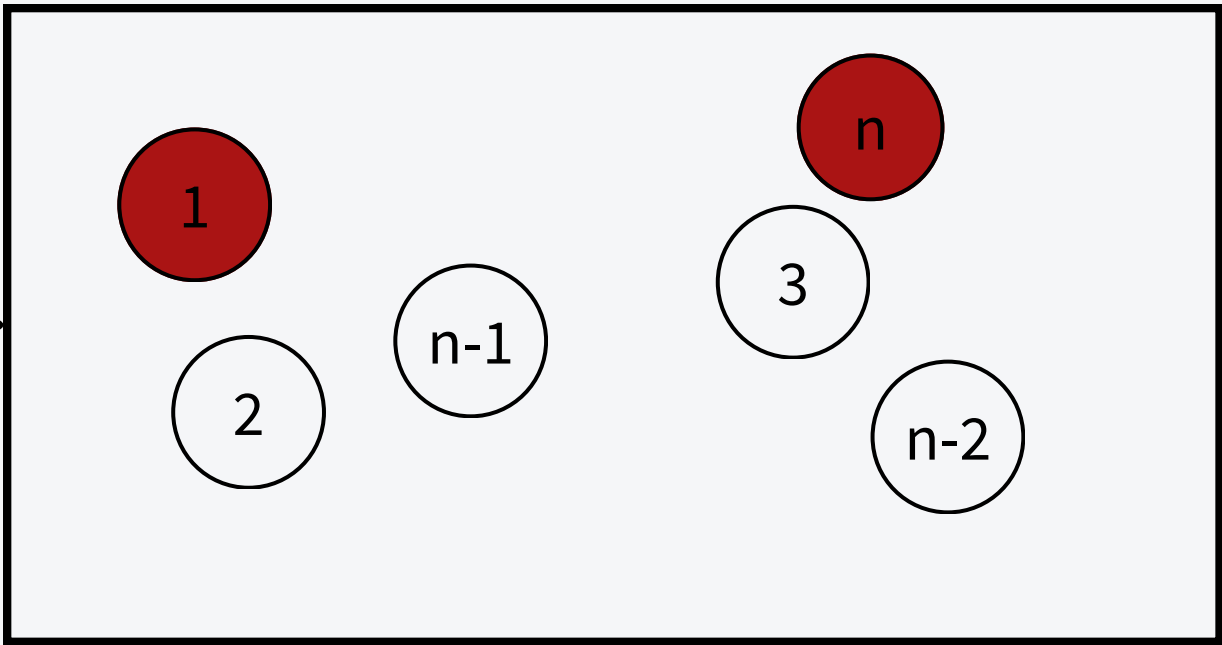
- Passo 2 - Execução do EHO
 - Cálculo do Fitness:

1	0	..	0	1
---	---	----	---	---

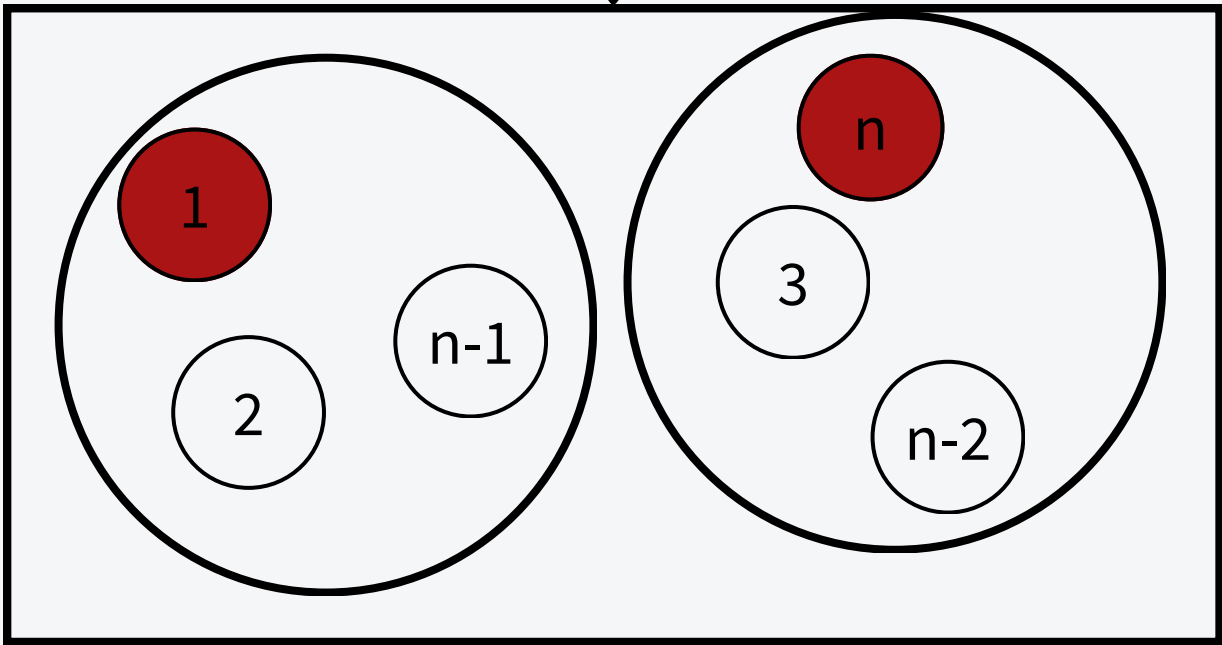
Realizando o mapeamento na base de dados

ID	Parâmetro 1	...	Parâmetro m
1	x_norm	..	y_norm
...
n	w_norm	..	h_norm

Projetando no espaço vetorial de dimensão m



Clusterização a partir da distância Euclidiana

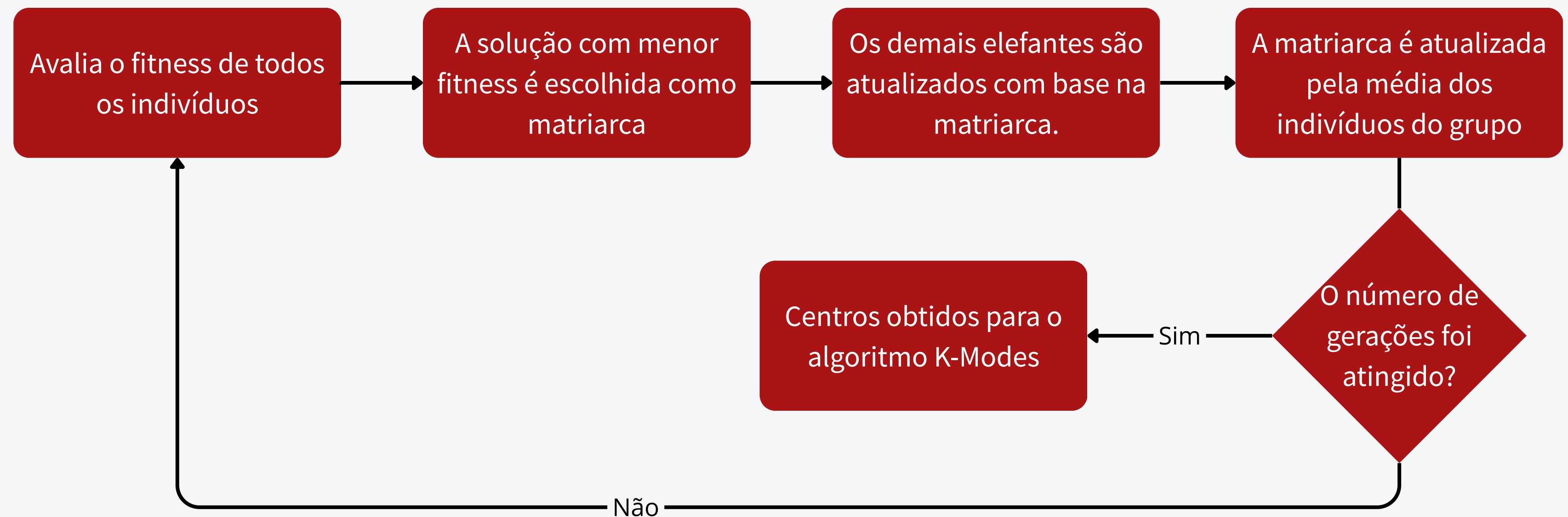


distância intra-cluster

Fitness

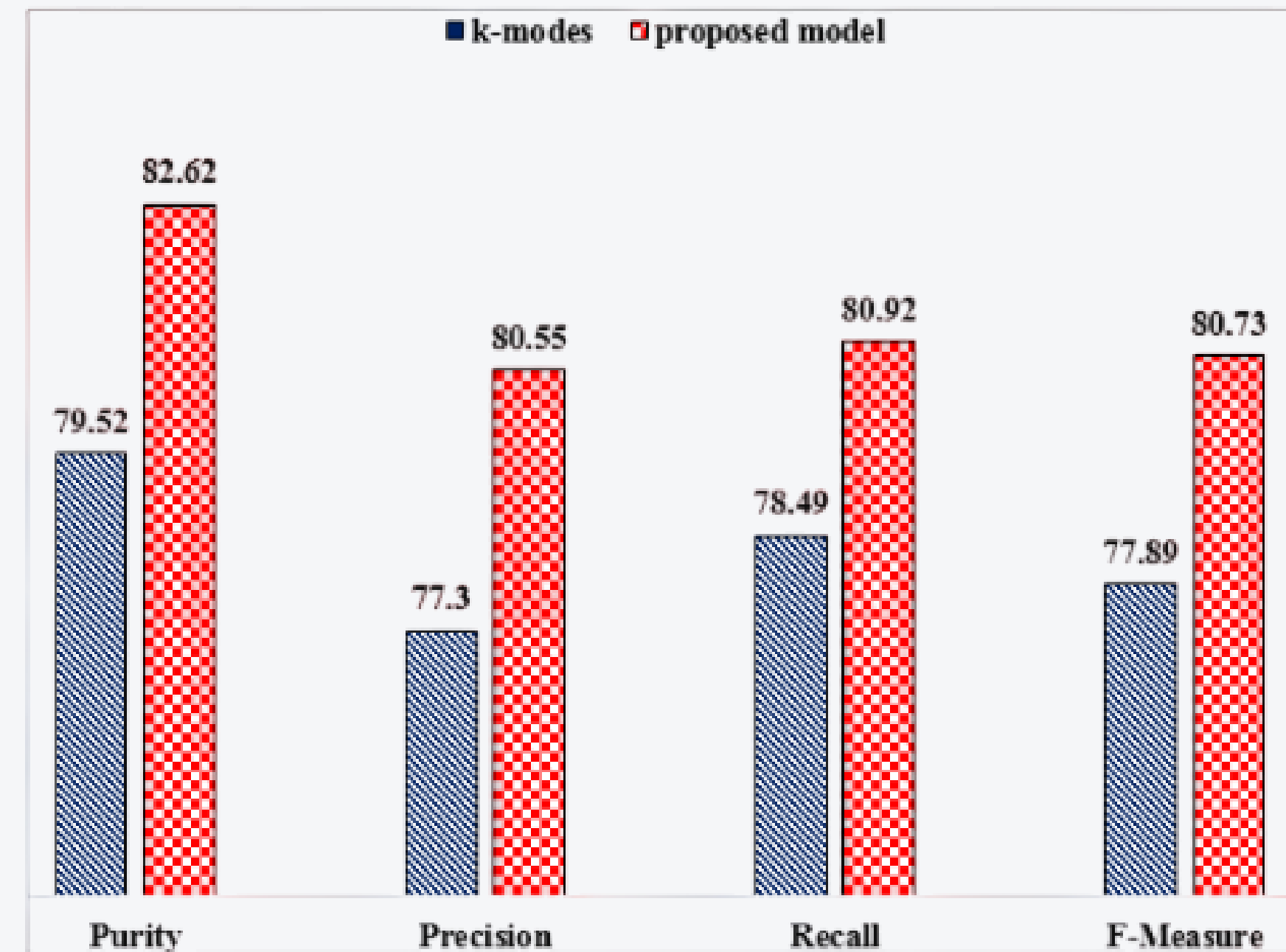
6.1. Modelo Proposto

- Passo 2 - Execução do EHO
 - Fluxograma Processamento:



6.2. Resultados

- Diagrama de comparação do modelo proposto com o K-Modes (seleção inicial aleatória dos centros) baseado em 200 iterações.
- Métricas:
 - Purity: quanto os elementos de cada cluster pertencem a uma única classe real.
 - Precision: dos elementos que eu agrupei juntos, quantos realmente são da mesma classe?
 - Recall: quantos dos elementos de uma classe eu consegui agrupar juntos?
 - F-Measure: um equilíbrio entre agrupar bem e não deixar itens da classe espalhados.



Referências

- [1] Wang, G. G., Deb, S., & Coelho, L. D. S. (2015, December). Elephant herding optimization. In 2015 3rd international symposium on computational and business intelligence (ISCBI) (pp. 1-5). IEEE.
- [2] Soleimanian Gharehchopogh, F., & Haggi, S. (2020). An Optimization K-modes clustering algorithm with elephant herding optimization algorithm for crime clustering. Journal of Advances in Computer Engineering and Technology, 6(2), 79-90.

Obrigado!
Dúvidas?

